Università degli Studi di Pavia
Dipartimento di Fisica

DOTTORATO DI RICERCA IN FISICA - XXXVIII CICLO

Statistical Physics for Economic and Social Systems: From Models to Simulations

Federica De Domenico

Submitted to the Graduate School of Physics in partial fulfillment of the requirements for the degree of
DOTTORE DI RICERCA IN FISICA
DOCTOR OF PHILOSOPHY IN PHYSICS
at the University of Pavia

Supervisors: Giacomo Livan, Guido Montagna

Contents

Research Context and Contributions

Compared to a century ago, physicists now work on topics well beyond the traditional boundaries of physics. This broadening reflects not only their curiosity and adaptability, but also their ability to stylise complex phenomena, capturing simultaneously micro-level mechanisms and macro-level patterns. In recent decades, this approach has played a pivotal role in advancing the study of complex systems, an interdisciplinary field that attracts researchers from physics, economics, sociology, biology, and many other domains.

My research falls within this field and spans two main areas that, while distinct in application, share a common methodological foundation rooted in statistical physics.

The first line of research focuses on opinion dynamics and social influence, specifically examining how peer interactions shape individual decisions and, ultimately, collective societal outcomes. Indeed, imitation plays a central role in human society: it functions both as a fundamental learning mechanism and as a social strategy, grounded in the need for belonging and acceptance. Individuals often adopt dominant behaviours not because they are better informed, but out of fear of exclusion. However, evidence indicates that conformity is not always beneficial, neither for individuals nor for society as a whole.

In our studies, we analyse the trade-off between personal information and the social pressure to conform, and how this interplay gives rise to emergent collective patterns. Using agent-based models, we investigate the conditions under which individuals conform, follow prevailing trends, or resist peer influence. We also examine the role of imitation versus randomness in decision-making. Key results include the identification of phase transitions in collective behaviour and insights into the emergence of meritocratic outcomes. Our analyses indicate that relying on individual information (or, in some cases, random choice) can produce more desirable results than imitation-based strategies. These models apply to a wide range of real-world phe-

nomena, from opinion polarisation and cultural diffusion to viral trends and financial behaviour.

The second research direction lies in the area of econophysics, where tools from statistical physics are applied to financial systems. Specifically, we present a systematic and original comparative study of several non-Gaussian distributions commonly proposed for modelling asset returns (i.e., relative price changes), including the Student's t, q-Gaussian, Truncated Lévy and Modified Weibull distributions. While each model has been individually proposed as suitable for fitting empirical data, we adopt a unified calibration framework that allows for a consistent comparison of their statistical properties and pricing implications. Through both qualitative and quantitative analyses, we show that, despite their functional differences, these models provide coherent representations of heavy-tailed return behaviour. This finding clarifies the relationship between alternative modelling approaches and contributes to a more flexible and empirically grounded understanding of financial dynamics.

From a methodological perspective, this work is based on agent-based modelling, Monte Carlo simulations, and analytical techniques from statistical physics, including stochastic processes and mean-field approximations. All simulations are implemented in Python, chosen for its flexibility, clarity, and extensive scientific libraries.

The research presented in this thesis is based on the following publications:

- [1] **De Domenico, F.**, Caccioli, F., & Livan, G. (2025). Peer influence breaks ergodicity in an opinion dynamics model with external information. *Journal of Physics: Complexity, 6*(4), 045007;

- [2] **De Domenico, F.**, Caccioli, F., Livan, G., Montagna, G., & Nicrosini, O. (2024). Imitation versus serendipity in ranking dynamics. *Royal Society Open Science, 11*(7), 240177;

- [3] **De Domenico, F.**, Livan, G., Montagna, G., & Nicrosini, O. (2023). Modeling and simulation of financial returns under non-Gaussian distributions. *Physica A: Statistical Mechanics and its Applications, 622*, 128886.

Thesis Outline

The dissertation is organised as follows. Chapter 1 provides an overview of complex systems, covering core concepts spanning statistical physics, stochastic processes and computational techniques. Emphasis is on instrumental

notions for the original work next presented, without aim of completeness. Chapter 2 examines two agent-based models we developed to study how social influence shapes individual decision-making and leads to emergent collective behaviours. Chapter 3 compares non-Gaussian models proposed for financial data and examines their implications for statistical modelling and option pricing. Finally, Chapter 4 summarises the main findings, discusses the limitations of the models and outlines directions for future research.

Acknowledgments

The original research presented in this thesis, along with the manuscript itself, would not have been possible without the careful mentoring of Professor Giacomo Livan and Professor Guido Montagna. I am deeply grateful for your supervision over these past years, both from a scientific and, even more so, a personal perspective. I will always cherish our conversations about physics and about all the other things that happen in life. Thank you sincerely for your humanity and your patience, especially during the times when I felt lost. I believe I have become a far better scientist and my perspective on the world has greatly improved thanks to your guidance.

I would like to sincerely thank the Department of Computer Science at University College London for its hospitality and for the wonderful people I had the chance to meet during my time there. A special thank goes to Professor Fabio Caccioli for the opportunity and for his help.

I also gratefully acknowledge the Erasmus Traineeship Program for the financial support during my stay abroad and INFN for providing the computational resources that enabled work far beyond what my personal laptop could have managed.

Finally, heartfelt thanks to Professor Daniela Rebuzzi for her invaluable help and support over these three years as Ph.D. coordinator.

To my family, friends and all the people I have met over the years: you have been the best part of this journey. Thanks for all your love and for supporting me, believing in me and caring for me. I carry with me so many memories of our time together (years, even decades! We are growing older). My heart and my life are full of love and I feel deeply grateful.

Chapter 1

Complex Systems: an Overview

Complex systems represent a broad and interdisciplinary field of research concerned with understanding how collective phenomena emerge from the interactions of many components. Such systems often exhibit common structural and dynamical features despite spanning seemingly disparate domains, from biological networks and ecological systems to financial markets and social institutions.

This chapter outlines the theoretical framework underlying the study of complex systems, with particular emphasis on socio-economic contexts. Rather than aiming for completeness, the goal is to provide a coherent foundation for the original research presented in Chapters 2 and 3.

We begin with the fundamental notions of interaction and emergence in Section 1.1, emphasizing how macroscopic patterns and structures can arise from simple local interactions. This leads naturally into a discussion of key features shared by many complex systems (including self-organisation, hierarchical structure, scale invariance, and adaptability) which recur across physical, biological, and socio-economic domains. Statistical physics is introduced in Section 1.2 as a central theoretical framework for connecting micro-level dynamics with macro-level outcomes and, in order to address the role of uncertainty and fluctuation, the chapter introduces stochastic processes, including Markov processes and stochastic differential equations. The discussion progresses to stable (Section 1.4) and heavy-tailed (Section 1.5) distributions, which are instrumental in modelling systems characterised by large fluctuations, extreme events, or scale-free behaviour. These include Lévy distributions, power laws, and Pareto-type inequalities, all of which are commonly observed in socio-economic data. In Section 1.6, critical phenomena, phase transitions, and mean-field approximations are then explored as frameworks for understanding abrupt systemic changes and collective behaviour. In this context, the Ising model serves as a paradigmatic exam-

ple. Computational methods are discussed in Section 1.7, focusing on Monte Carlo techniques and agent-based models. Finally, Section 1.8 introduces key concepts from network theory.

1.1 Interaction and Emergence

Complex systems are generally defined as systems composed of a large number of non-linearly interacting elements whose interactions lead to the emergence of non-trivial properties which characterise the system as a whole. This is, of course, just one of several possible definitions of complexity [4, 5].

Complex systems are a broad and profoundly interdisciplinary field, encompassing a diverse range of phenomena from both the natural and human-made worlds. Examples include the brain, bird flocks, social structures, financial markets, traffic flows, and the Internet. Accordingly, its study benefits from the contributions of professionals across multiple disciplines, including physicists, economists, physiologists, psychologists, and computer scientists, to name just a few.

One of key concepts of complex systems is emergence, standing in contrast to the reductionist approach. Both emergence and reductionism assume that reality can be organised into distinct levels of structure, yet the latter holds that a system can be fully understood by breaking it down into its constituent parts (each governed by fundamental laws) and that higher-level phenomena can, in principle, be entirely explained by lower-level mechanisms. In contrast, the emergentist view asserts that novel properties can arise from non-linear interactions among these components, which cannot be predicted solely by analysing the system's individual elements [6]. This idea is often encapsulated in the expressions "more is different" formulated by P.W. Anderson and "the whole is something besides the parts" [7–9]. A canonical example of emergence in statistical physics is provided by the Ising model (Sec. 1.6.1), which exhibits phase transitions that cannot be observed by examining individual spins in isolation but arise only through their coordinated interactions. Such behaviour exemplifies the core principles of emergence and underscores the relevance of statistical physics as the theoretical foundation for studying complex systems, as further discussed in Sec. 1.2.

In recent decades, scientific interest in complex systems has grown substantially, accompanied by an increasing demand for tools and methodologies capable of analysing and managing their inherent intricacies. Since the 1980s, significant progress has been made in the quantitative study of this field, driven by the development of theoretical frameworks and advances in com-

putational modelling. A notable demonstration of the relevance of such approaches emerged during the COVID-19 pandemic, when epidemic spreading models proved that quantitative and algorithmic frameworks can effectively anticipate the outbreaks of infectious disease [10]. These models are thus particularly valuable in contexts where empirical data are scarce or unreliable, providing essential support for, among the others, public health decision-making [11].

The centrality of complex systems to modern science was further underscored by the awarding of the 2021 Nobel Prize in Physics, which recognised "groundbreaking contributions to our understanding of complex physical systems." Half the prize was shared by Syukuro Manabe and Klaus Hasselmann for their pioneering work in the physical modelling of Earth's climate, including the quantification of variability and the reliable prediction of global warming. The other half was awarded to Giorgio Parisi for his discovery of the interplay between disorder and fluctuations in physical systems across scales, from the atomic to the planetary [12].

It is fair to say that, despite such recognition, a degree of scepticism persists among those unfamiliar with the role of physicists in addressing problems that lie beyond the traditional realms of physics [13]. One of the aims of this work is to challenge that view by demonstrating that physicists, equipped with robust theoretical tools and computational methodologies, are well-positioned to contribute meaningfully to the study of complex systems.

1.1.1 Key Properties

We now turn to a set of characteristic properties that consistently appear across a wide range of complex systems, following [4, 8, 14].

a) Self-organisation: Complex systems often exhibit the spontaneous emergence of structured, ordered patterns through internal reconfiguration. This phenomenon is particularly evident under conditions far from equilibrium, when a system may reach a critical point at which small perturbations are amplified, driving the system through bifurcations into new configurations. Such behaviour is typical of biological and social systems, as well as of physical settings such as crystallisation, chemical oscillations, and dune formation.

b) Adaptation: Adaptation refers to a system's ability to modify its internal structure or behaviour in response to changes in the external environment, enabling it to sustain performance under varying conditions. Mechanisms such as learning, evolution, or optimisation are often involved.

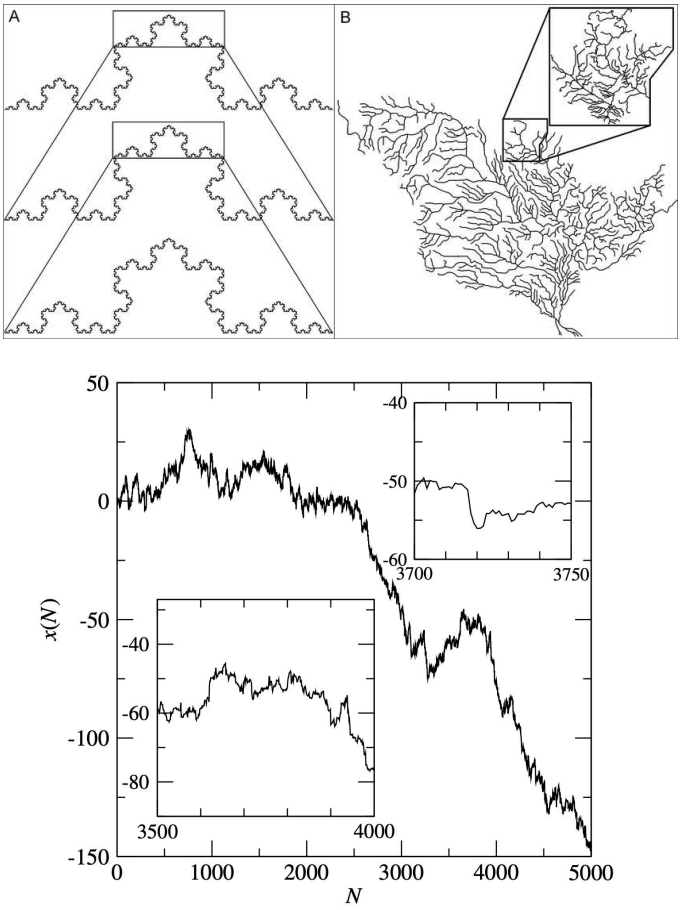

Figure 1.1: Illustrations of the property of self-similarity.
Upper panel: adapted from [15]. Left: An artificially created self-similar curve, where magnification of a tiny portion of the line results in exactly the same image as the previous view. Right: The naturally occurring Mississippi River drainage basin with a portion of the river drainage expanded to show its similarity to the whole. A scale bar is irrelevant in this figure because of the scale-invariant property.
Lower panel: adapted from [16]. Self-similar function obtained by summing random variables from a Gaussian distribution. Several successive 'zooms' reveal the self-similar nature of the function. Note however that, for high resolutions, one observes the breakdown of self-similarity, related to the existence of a discretisation scale.

c) Hierarchical structure: Many complex systems are organised across multiple scales, with elements at one level forming components of more extensive subsystems at the next. A typical example is the physical hierarchy of matter: atoms form molecules, which in turn constitute crystals.
Hierarchies can also emerge spontaneously as ranked structures that shape the flow of information, resources, and influence across different levels of a system. For instance, urban environments naturally develop stratified population distributions, with large metropolitan centres coexisting alongside smaller towns and villages. A further example is provided by brain activity [17–21]. Brain dynamics are organised hierarchically across multiple spatial and temporal scales: microscopic neuronal interactions give rise to mesoscopic population activity, which in turn supports macroscopic dynamics involving large-scale brain networks. This hierarchy is dynamical as well as structural, with fast, local processes embedded within and modulated by slower, global brain states through cross-scale interactions such as cross-frequency coupling. Such organisation enables functional specialisation at lower levels while supporting large-scale integration, flexibility, and robust cognitive function.
A further familiar example can be found in social systems, where hierarchies often arise on the basis of factors such as wealth, power, or achievement [22].

d) Scale invariance: Closely related to hierarchical organisation is the absence of a characteristic scale, whereby systems display similar structural or dynamical patterns across different magnitudes of observation. A hallmark of such scale-free behaviour is the emergence of power-law distributions, in which the probability $p(x)$ of observing an event of size x follows the relation $p(x) \sim x^{-\alpha}$, where α is a positive scaling exponent. Unlike exponential or Gaussian laws, power-law relationships lack a typical scale (Sec. 1.5.1). This property is observed in diverse natural and social contexts, such as the branching architecture of circulatory and neural networks, the fractal geometry of coastlines, or the recursive patterns in crystal growth (Fig. 1.1).

1.2 The Foundation: Statistical Physics

One of the most intriguing aspects of statistical physics lies in its broad applicability across domains that appear unrelated but can nonetheless be anal-

ysed within a unified framework. At first glance, condensed matter physics and socio-economic systems may seem to belong to fundamentally distinct modelling domains. Upon closer inspection, however, both share a common objective: explaining how microscopic interactions give rise to macroscopic behaviour, an insight enabled by the methods typical of statistical physics.

Statistical physics studies the laws governing the behaviour of macroscopic systems composed of a large number of individual constituents [23]. The macroscopic scale refers to phenomena accessible to direct observation, while the microscopic level involves interactions among the system's fundamental components. Due to the overwhelming number of degrees of freedom, exact analytical solutions are typically unfeasible. Statistical physics addresses this challenge by adopting a probabilistic framework that focuses on the collective behaviour of macroscopic observables [24, 25]. Statistical mechanics provides the mathematical structure to connect microscopic dynamics with macroscopic thermodynamic quantities [26], revealing regularities that emerge only examining the system as a whole, emphasising the central principle of emergence.

Notably, the connection between physics and the social sciences was recognized nearly a century ago. In his essay "Il valore delle leggi statistiche nella fisica e nelle scienze sociali" [27], written around 1930, Ettore Majorana emphasized the "formal analogy between the statistical laws of physics and those of the social sciences" and proposed an "identity in value and method" between the two. Interestingly, the use of statistical framework to analyse social behaviour anticipates its formal application in physics. In fact, the discovery of regularities in demographic data (e.g., birth rates, mortality, or crime statistics) played a central role in the development of modern statistical methods. These patterns, which emerged despite the apparent unpredictability of individual actions, inspired early scholars to seek quantitative, physics-like explanations for social phenomena [28].

A paradigmatic illustration of this cross-domain applicability is provided by the Ising model. In the physical context, magnetisation arises when individual spins, each capable of adopting one of two orientations, tend to align in the same way. If spins behave independently, magnetisation is unlikely to emerge. However, when a spin's orientation is influenced by that of its neighbours, a collective order can form, resulting in a macroscopic magnetic phase. Analogous mechanisms appear in social systems, as exemplified by Schelling's model of residential segregation [29]. In such models, individuals may occupy one of two possible states (e.g. binary voting setting), and their decisions are influenced by the preferences and opinions of the community around them. From this perspective, the core models used to describe physical and social systems reveal a striking structural similarity.

At first, it may seem counterintuitive to apply stylised, physics-based models to human behaviour, given the multifaceted and nuanced nature of social interactions. Indeed, human decisions are shaped by beliefs, emotions, history, and expectations, which are features that have no counterpart in inanimate systems. Nevertheless, statistical physics provides an effective framework for capturing the essential features of social phenomena: by focusing on large-scale patterns rather than individual peculiarities, it becomes possible to uncover emergent regularities shaped by the underlying network of interactions among individuals. Although individuals are far more complex than atomic spins, their aggregate behaviour can still exhibit predictable features, allowing for meaningful analysis through simplified models. Of course, notable differences remain between the two systems. Unlike atoms, individuals often act strategically to maximise personal utility, particularly in economic contexts (although this self-interested behaviour can lead to inefficient collective outcomes [30, 31]). Furthermore, while magnetic systems evolve based solely on prior configurations, human actions often incorporate expectations about the future. The presence of foresight, emotion, and irrationality marks a clear departure from the mechanical determinism of physical systems. Yet, paradoxically, it is precisely this complexity that motivates simplification: by abstracting from individual detail, we can detect emergent traits shared by otherwise unrelated systems.

Given the inherent randomness in these systems, we are naturally led to introduce the concept of stochasticity, a defining feature of systems in which it is impossible to determine all relevant variables through deterministic laws, with probability theory in this context capturing both variability and uncertainty [32].

1.3 Stochastic Processes

Stochastic processes provide a powerful framework for describing systems whose evolution in time is governed by probabilistic rules [33–37]. Formally, a stochastic process is a family of random variables $(X_t)_{t\in I}$ indexed by time, where I may be discrete or continuous, leading to discrete- or continuous-time dynamics. The system's behaviour is characterised by its joint probability distributions

$$p(x_1, t_1; x_2, t_2; \ldots; x_n, t_n), \tag{1.1}$$

which specify the likelihood of observing values $x_1, \ldots, x_n$ at times $t_1, \ldots, t_n$. Our primary focus is on discrete-time processes; a supplementary treatment of continuous-time processes can be found in [38].

A process is said to be stationary if these distributions are invariant under time translation:

$$p(x_1, t_1 + s; \ldots; x_n, t_n + s) = p(x_1, t_1; \ldots; x_n, t_n) \quad \forall s \in I, \tag{1.2}$$

implying that its statistical properties remain constant over time, a common assumption in physics and economics.

Historically, stochastic modelling emerged from 17th-century gambling theory and matured through the contributions of Bachelier, Einstein, Langevin, and Itô [32, 38, 39]. Brownian motion, first observed by Robert Brown in 1827, provided a unifying description of irregular yet statistically regular motion [40]. The theoretical explanation was then provided independently by Einstein in 1905 and by Smoluchowski [41] and later refined by Langevin [42]. Einstein explained this as the result of molecular collisions [43], while Langevin introduced a stochastic force (white noise) to capture its random impacts [42]. This formulation, later formalised by Itô, laid the foundations of stochastic calculus.

Remarkably, the earliest formal model of stochastic motion was not physical but financial. In 1900, Bachelier proposed a model for price fluctuations in the Paris market [44], laying the foundation for modern financial mathematics. Bachelier's discussion can be seen as a discrete model of Brownian motion, which is called random walk; this definition was introduced by Karl Pearson in 1905 to identify a process whose path is given by random steps.

1.3.1 Markov Process

A key simplification arises when a process satisfies the Markov property, namely when the conditional probability of the future depends only on the present state and not on the past history:

$$p\big(x_{t+s} \mid x_t,\, x_{t-1}, \ldots, x_0\big) = p\big(x_{t+s} \mid x_t\big), \tag{1.3}$$

where $p(x_{t+s} \mid x_t)$ denotes the transition probability of moving from state x_t to state x_{t+s}. Although no real-world process is perfectly Markovian, many exhibit short enough memory to be treated as such for practical purposes. Such memory-less processes include the symmetric random walk and, in continuous time, the Wiener process $W(t)$.
A Wiener process $W(t)$ is a continuous-time stochastic process with the properties listed below [34, 45]:

a) Initial Condition: $W_0 = 0$ almost surely; that is, with probability 1, though there may exist a set of outcomes of measure zero for which this does not hold.

b) Independent Increments: For any $0 \leq t_0 < t_1 < \cdots < t_n$, the increments $W_{t_1} - W_{t_0}, W_{t_2} - W_{t_1}, \ldots, W_{t_n} - W_{t_{n-1}}$ are independent random variables.

c) Stationary Increments: The distribution of the increment $W_{t+s} - W_s$ depends only on t, implying temporal homogeneity.

d) Gaussian Increments: For any $0 \leq s < t$, the increment $W_t - W_s$ is normally distributed with mean zero and variance $t - s$, i.e., proportional to the elapsed time.

e) Continuity of Paths: With probability 1, the function $t \mapsto W_t$ is continuous, though it is nowhere differentiable.

The Wiener process is central to stochastic calculus and quantitative finance (Sec. 3.6.1), providing a tractable model for memoryless noise. Notably, the term Brownian motion can refer both to the physical phenomenon observed in nature (e.g., the erratic motion of pollen grains in water) and to the stochastic process used to model it mathematically (i.e., the Wiener process).

1.3.2 Stochastic Differential Equations

Stochastic differential equations (SDEs) generalise ordinary differential equations by adding a noise term, usually driven by a Wiener process. Depending on how noise interacts with the system, it may be additive (which is independent of the system's state) or multiplicative (which scales with the state of the process). Three canonical examples illustrate their behaviour [46]:

a) Brownian motion, in the physical sense, refers to the random diffusive motion of a particle suspended in a fluid, whose probability density evolves according to the Einstein diffusion equation[1]. The Wiener process is the canonical mathematical model of this phenomenon: it has continuous paths and Gaussian, stationary increments. More generally, it admits a Langevin representation with additive white noise:

$$\mathrm{d}X(t) = \sigma \, \mathrm{d}W(t), \tag{1.4}$$

with $X(0) = 0$. Its solution is

$$X(t) = \sigma W(t),$$

[1]Einstein showed that the probability density $\rho(x,t)$ for the position of a diffusing particle satisfies $\partial\rho/\partial t = D\,\partial^2\rho/\partial x^2$, where D is the diffusion constant [43].

and its statistical properties are

$$\mathbb{E}[X(t)] = 0, \quad \mathrm{Var}[X(t)] = \sigma^2 t.$$

Hence, over time the displacement follows a normal distribution with variance growing linearly in t. This example highlights how a simple random mechanism at the microscopic level leads to well-defined macroscopic diffusion behaviour.

b) Arithmetic Brownian motion extends the previous model by introducing a constant deterministic drift term, modelling a system in which deterministic and random contributions accumulate linearly in time through additive noise:

$$\mathrm{d}X(t) = \mu\,\mathrm{d}t + \sigma\,\mathrm{d}W(t). \tag{1.5}$$

The solution is

$$X(t) = X(0) + \mu t + \sigma\,W(t),$$

yielding

$$\mathbb{E}[X(t)] = X(0) + \mu\,t, \quad \mathrm{Var}[X(t)] = \sigma^2\,t.$$

This process was adopted by Bachelier in his early model of asset prices. The parameter μ controls the deterministic trend: $\mu > 0$ yields an upward drift, $\mu < 0$ a downward drift, and $\mu = 0$ recovers pure Brownian motion. Since the process is not restricted to remain positive, negative values may occur, which historically led to the replacement of Bachelier's model by multiplicative (geometric) models of price dynamics.

c) Geometric Brownian motion is central in finance, especially in modelling stock prices, and involves multiplicative noise:

$$\mathrm{d}X(t) = \mu\,X(t)\,\mathrm{d}t + \sigma\,X(t)\,\mathrm{d}W(t). \tag{1.6}$$

Its solution is

$$X(t) = X(0)\exp\left[\left(\mu - \frac{1}{2}\sigma^2\right)\,t + \sigma\,W(t)\right].$$

The mean and variance are

$$\mathbb{E}[X(t)] = X(0)e^{\mu t}, \quad \mathrm{Var}[X(t)] = X(0)^2 e^{2\mu t}\left(e^{\sigma^2 t} - 1\right).$$

The exponential growth in variance reflects the compounding effect of multiplicative randomness over time.

These examples underscore how SDEs provide a flexible framework to represent both regular trends and irregular fluctuations.

Whether in physics, biology, or finance, stochastic processes serve as indispensable tools for interpreting the complexity of real-world phenomena and for understanding how randomness influences dynamic systems.

However, this formulation of SDEs is ambiguous and must be complemented by a proper mathematical definition of the corresponding integral, which was laid by Kiyosi Itô, who developed a rigorous theory of stochastic integration in the 1940s. An alternative formulation was later proposed by Stratonovich. The two interpretations are related, although different, and the choice between them depends on the application under consideration.

1.3.3 Itô's Lemma

Itô's lemma plays a central role in stochastic calculus, as it provides the means to compute the differential of a function of a stochastic process. In the following, we focus on the essential concepts relevant to our discussion; for a more comprehensive treatment of stochastic calculus, the reader is referred to standard texts such as [34].

Before presenting Itô's lemma, it is important to understand the nature of the stochastic integral that underlies the construction of stochastic differential equations [47, 48]. In the classical setting, the Riemann integral is built upon smooth functions and limits of sums evaluated at arbitrary points within subintervals. However, Brownian motion is nowhere differentiable and exhibits infinite variation on any time interval, rendering the classical definition inapplicable.

To address this, the Itô integral was introduced as a stochastic integral with respect to a Wiener process. Given a non-anticipative process $\varphi(t)$ (that is, one whose value at time t depends only on information available up to that time), the Itô integral over the range $[0, T]$ is formally defined as the L^2 limit (i.e., a limit in the mean-square sense[2]) of sums:

$$\int_0^T \varphi(t)\,\mathrm{d}W(t) := \lim_{n\to\infty} \sum_{i=0}^{n-1} \varphi(t_i)\left[W(t_{i+1}) - W(t_i)\right], \tag{1.7}$$

where $\{t_i\}$ is a partition of $[0, T]$ with step size tending to zero, and the integrand is evaluated at the left endpoint t_i of each subinterval. This choice

[2] A sequence of random variables $\{X_n\}$ converges in the L^2 sense to a random variable X if $\lim_{n\to\infty} \mathbb{E}[(X_n - X)^2] = 0$. This is also known as convergence in mean square.

ensures that the integral is non-anticipative, a key requirement for modelling systems in which future noise values are not observable.

A crucial property of the Itô integral is that it is a martingale (i.e., at any time the best predictor of its future value is simply its current value, given all the information available so far) and, in particular, has zero mean:

$$\mathbb{E}\left[\int_0^T \varphi(t)\,\mathrm{d}W(t)\right] = 0. \tag{1.8}$$

Additionally, it satisfies the Itô isometry:

$$\mathbb{E}\left[\left(\int_0^T \varphi(t)\,\mathrm{d}W(t)\right)^2\right] = \mathbb{E}\left[\int_0^T \varphi(t)^2\,dt\right], \tag{1.9}$$

which serves as a foundational result for proving existence and uniqueness of solutions to stochastic differential equations. The structure of the Itô integral also leads to the remarkable identity

$$(\mathrm{d}W(t))^2 = \mathrm{d}t,$$

which underpins the appearance of the second-order term in Itô's lemma.

We can now state Itô's lemma. Let $X(t)$ be a stochastic process satisfying the stochastic differential equation:

$$\mathrm{d}X(t) = \mu(t, X(t))\,\mathrm{d}t + \sigma(t, X(t))\,\mathrm{d}W(t), \tag{1.10}$$

where $W(t)$ denotes a standard Wiener process (Sec. 1.3.1), and μ and σ are functions that may depend on both time and the current value of $X(t)$. Then, for a function $f(t, X(t))$ that is once differentiable in t and twice in x, Itô's lemma provides the differential:

$$\mathrm{d}f(t, X(t)) = \left(\frac{\partial f}{\partial t} + \mu\,\frac{\partial f}{\partial x} + \frac{1}{2}\,\sigma^2\,\frac{\partial^2 f}{\partial x^2}\right)\mathrm{d}t + \sigma\,\frac{\partial f}{\partial x}\,\mathrm{d}W(t). \tag{1.11}$$

This formula encapsulates one of the fundamental departures from classical calculus: the presence of a second-order term, which arises from the non-vanishing quadratic variation of Brownian motion and ultimately reflects the nowhere-differentiable nature of its paths.

One of the most prominent applications of Itô's lemma is in the derivation of the Black–Scholes equation for option pricing (see Sec. 3.6).

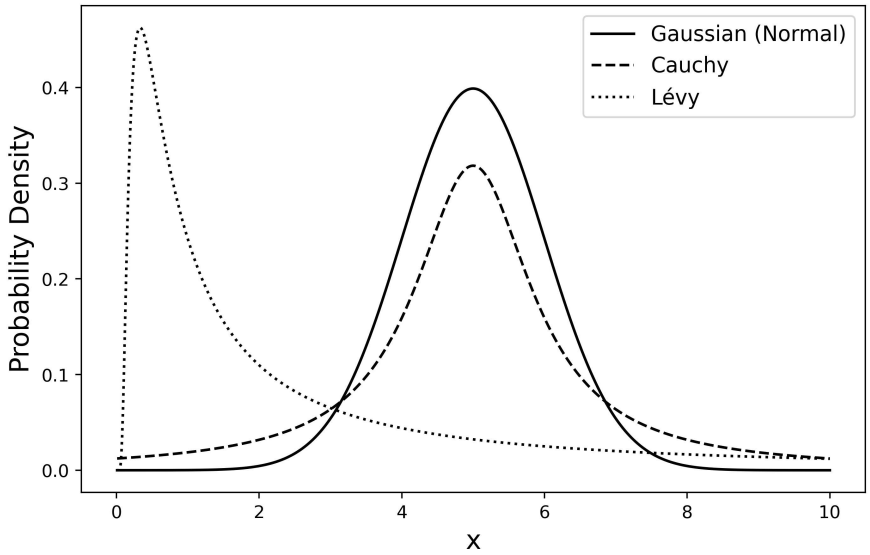

Figure 1.2: The Gaussian (solid), Cauchy (dashed), and Lévy (dotted) distributions, all members of the α-stable family, exhibit increasing tail heaviness and decreasing symmetry.

1.4 Stable Distributions

Among the many probability distributions used in modelling natural and social phenomena, stable distributions play a central role. First studied in depth by Paul Lévy in the early 20th century, they are defined by a distinctive property: the sum of two independent random variables drawn from a stable distribution, when appropriately rescaled and shifted, follows the same distributional form [49]. This invariance under linear combinations provides stable laws with structural robustness, particularly relevant for systems governed by the aggregation of many independent components [50]. Before turning to a detailed discussion of this property, we first highlight their key features.

1.4.1 Main Features

The behaviour of stable distributions is largely governed by a parameter $\alpha \in (0, 2]$, hence their alternative name "Lévy α-stable distributions". Mandelbrot referred to those with $1 < \alpha < 2$ and strong positive skewness as Paretian distributions, which he proposed as suitable models for financial data in 1960 [51].

Although Lévy's foundational work provided a full mathematical characterisation of this family, closed-form expressions for the probability density function $f(x)$ exist only in special cases. Recent developments in compu-

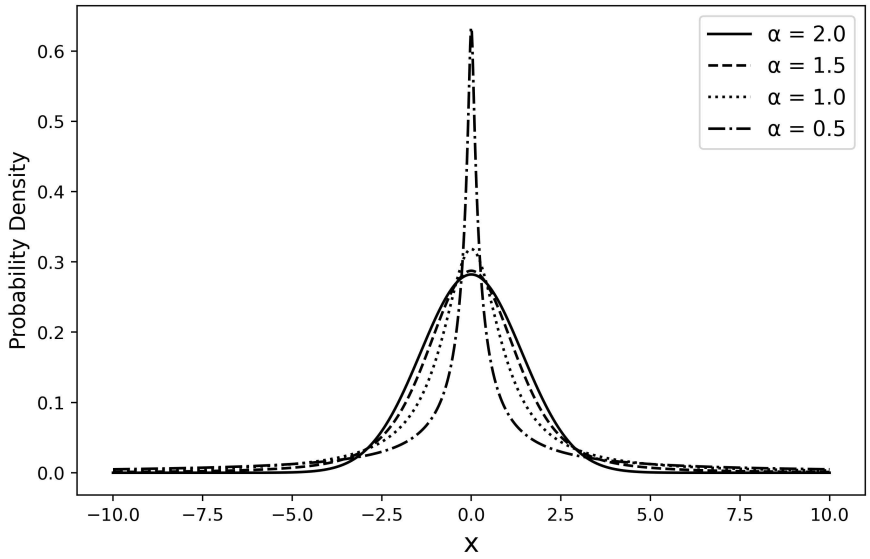

Figure 1.3: Probability density functions of symmetric stable distributions for different values of the stability parameter α. All curves are symmetric ($\beta = 0$) and normalised. As α decreases from 2 to 0.5, the tails become increasingly heavy and the distributions more sharply peaked.

tational algorithms have significantly improved our ability to handle these distributions, overcoming many earlier analytical limitations. For the general case, the most useful representation is the characteristic function, defined as the Fourier transform of $f(x)$:

$$\varphi(t) = \int_{-\infty}^{\infty} f(x)\, e^{ixt}\, \mathrm{d}x.$$

A random variable X follows a stable distribution if its characteristic function can be written as:

$$\varphi(t; \alpha, \beta, c, \mu) = \exp\left(it\mu - |ct|^{\alpha}\left(1 - i\beta\,\mathrm{sgn}(t)\,\Phi\right)\right), \tag{1.12}$$

where

$$\Phi = \begin{cases} \tan\left(\dfrac{\pi\alpha}{2}\right), & \text{if } \alpha \neq 1 \\ -\dfrac{2}{\pi}\log|t|, & \text{if } \alpha = 1 \end{cases}$$

and $\mathrm{sgn}(t)$ denotes the sign function. Stable distributions form a four-parameter family characterised by [52]:

a) Stability parameter $\alpha \in (0, 2]$, shaping peakedness and tail behaviour;

b) Skewness parameter $\beta \in [-1, 1]$, controlling asymmetry;

c) Scale parameter $c > 0$, determining the spread;

d) Location parameter μ, governing the central tendency.

The parameters α and β primarily define the distribution's shape: α controls tail heaviness, β determines symmetry ($\beta = 0$) or skewness ($\beta \neq 0$). It is worth noting that several parametrisations exist in the literature; in [49], Nolan provided a comprehensive overview of eleven forms and the transformations between them, which are particularly useful when fitting empirical data.

When $\alpha = 2$, the distribution reduces to the Gaussian; for $\alpha = 1$, $\beta = 0$, it becomes the Cauchy distribution. When $\alpha < 2$, the variance diverges and for $\alpha \leq 1$ even the mean is undefined. While this may seem problematic, it is precisely these features that make stable distributions well-suited for modelling extreme fluctuations, as they are heavy-tailed and can be strongly skewed, which are features often observed in empirical data but poorly captured by Gaussian models.

In Fig. 1.2, the few cases where the stable distribution admits a closed-form expression are depicted: the Gaussian, the Cauchy and the one-sided Lévy distribution ($\alpha = 0.5$, $\beta = 1$). To better visualise the effect of varying α while keeping symmetry fixed (i.e., $\beta = 0$), Fig. 1.3 shows several stable distributions on the same scale. As α decreases, the tails become progressively heavier and the peak sharper, highlighting the departure from Gaussianity as the variance becomes infinite.

A hallmark of stable distributions is their scale invariance:

$$p_N(x) = \frac{1}{N^{1/\alpha}} p\left(\frac{x}{N^{1/\alpha}}\right), \tag{1.13}$$

namely the system remains statistically self-similar when the observation scale changes. This fractal-like behaviour, as noted in [38], implies that "the whole looks like its parts." Such self-similarity (Sec. 1.1.1) has been observed in systems in financial markets, internet traffic and other complex systems. However, as discussed in the context of emergence, this property may break down in complex systems where "the whole becomes not only more than but very different from the sum of its parts" [7].

1.4.2 Central Limit Theorems

The significance of stable distributions lies primarily in their role as limiting distributions. According to the classical Central Limit Theorem (CLT), the

sum of independent and identically distributed (i.i.d.) random variables with finite variance converges to a Gaussian distribution as the number of terms increases. This underpins the widespread empirical appearance of Gaussian behaviour in both natural and social systems, especially when fluctuations are modest and additive in nature: this is why we speak of the "standard" distribution [53].

However, in many real-world systems the relevant quantity is the product of many small random factors, such as in models of compound interest, population growth, or asset prices evolving through proportional changes. In these cases, the multiplicative version of the CLT applies, which is a consequence of the additive version. Given two positive variables x_1 and x_2, we have:

$$x_1 \cdot x_2 = e^{\ln x_1} \cdot e^{\ln x_2} = e^{\ln x_1 + \ln x_2}.$$

For N variables, it follows that:

$$\prod_{i=1}^{N} x_i = \prod_{i=1}^{N} e^{\ln x_i} = \exp\{\sum_{i=1}^{N} \ln x_i\}.$$

Taking the limit as $N \to \infty$, we obtain:

$$\lim_{N\to\infty} \prod_{i=1}^{N} x_i = \exp\{\lim_{N\to\infty} \sum_{i=1}^{N} \ln x_i\}.$$

Defining $S_N = \sum_{i=1}^{N} \ln x_i$ and $S'_N = \prod_{i=1}^{N} x_i$, we obtain:

$$\lim_{N\to\infty} S'_N = \exp\{\lim_{N\to\infty} S_N\}. \tag{1.14}$$

In summary, if the variables have finite moments, the sum $\lim_{N\to\infty} S_N$ converges to a Gaussian random variable due to the additive CLT. Consequently, the product S'_N converges in distribution to a log-normal random variables as $N \to \infty$ [39].

When the assumption of finite variance is relaxed, the classical CLT no longer holds. In such cases, an extended version of the theorem applies: the sum of i.i.d. random variables with heavy-tailed distributions and infinite variance converges to a stable distribution with stability index $\alpha < 2$. This result considerably broadens the scope of convergence theorems [32, 54].

A key property underlying this generalisation is the stability of the family itself: if $X_1, \ldots, X_N$ are i.i.d. random variables drawn from a stable distribution with parameters (α, β, c, μ), then their sum $S = \sum_{i=1}^{N} X_i$ also follows a stable distribution, with updated parameters:

- Stability index: α (unchanged);

- Skewness: β (unchanged);

- Scale: $c \to N^{1/\alpha}c$;

- Location: $\mu \to N\mu$.

The stability of the distribution arises from the property that the characteristic function of the sum of two independent stable random variables is the product of their characteristic functions.

This generalises the Central Limit Theorem to distributions with infinite variance, thereby broadening its domain of applicability and reinforcing the idea that systems composed of many interacting random parts need not converge to Gaussian behaviour, especially when large fluctuations dominate.

A natural question arises: how can one adopt distributions whose moments may be infinite or undefined to model real-world systems? In fact, Lévy-stable distributions have proven highly effective in modelling diverse phenomena, such as foraging paths of animals (e.g., albatrosses), fluctuations in stock markets, and even physiological signals. For example, the inter-beat intervals of healthy human heart rhythms follow a Lévy distribution with $\alpha \approx 1.7$, whereas for patients with heart failure the distribution becomes more Gaussian [55].

Two main strategies are commonly employed to address the problem of infinite moments. First, one can restrict the domain of the distribution, imposing finite support so that all moments are finite, as discussed in [3, 39]. Second, one can use truncated versions of stable laws (such as Truncated Lévy Distributions), which maintain heavy-tailed behaviour over a finite range but decay more rapidly beyond a certain threshold. These approaches reconcile theoretical generality with empirical realism.

This topic will be further illustrated through numerical simulations and visualisations in Chapter 3, where we present concrete examples of the Central Limit Theorems.

1.5 Heavy-Tailed Distributions

In the previous section, we explored the Central Limit Theorem in both its classical form, leading to the Gaussian distribution under finite variance, and its generalised version, which incorporates stable distributions notable for their heavy tails and lack of finite variance.

These broader cases motivate the introduction of heavy-tailed distributions, so called because the probability of extreme events (namely outcomes far in one or both tails of the distribution) is higher than standard light-tailed models like the Gaussian [50, 56].

Formally, a distribution is considered heavy-tailed if its tail is not exponentially bounded. That is, for large x, the probability $p(x)$ decays slower than e^{-x}. Several distributions of interest fall into this category, including the log-normal, Lévy, Pareto, and power-law distributions. In this dissertation, particular attention will be given to four heavy-tailed families that are especially relevant to socio-economic systems: the Student's t, the q-Gaussian, the Modified Weibull, and the Truncated Lévy distributions. These will be analysed in detail in Chapter 3.

Notably, while all stable distributions with $\alpha < 2$ exhibit heavy tails, not all heavy-tailed distributions are stable. We now focus on power-law distributions, which play a central role in modelling socio-economic phenomena and various forms of collective behaviour.

1.5.1 Power Laws

A power-law distribution is characterised by the fact that the probability of observing a value x of a variable decreases as an inverse power of that value [53, 57]:

$$p(x) = C\, x^{-\alpha}, \tag{1.15}$$

where $\alpha > 0$ is the exponent and C is a normalisation constant. For the distribution to be normalisable, we must restrict the domain to $x \geq x_{\min} > 0$, and typically $\alpha > 1$ is required. Furthermore:

- The mean is finite only if $\alpha > 2$;
- The variance is finite only if $\alpha > 3$;
- In general, moments $\langle x^m \rangle$ exist only if $m < \alpha - 1$.

In practice, many empirical distributions follow a power-law behaviour only in their tails, motivating the use of $x_{\min}$ to exclude the non-asymptotic region. If plotted on a double logarithmic scale, power-law distributions appear as straight lines:

$$\ln p(x) = -\alpha \ln x + \ln C. \tag{1.16}$$

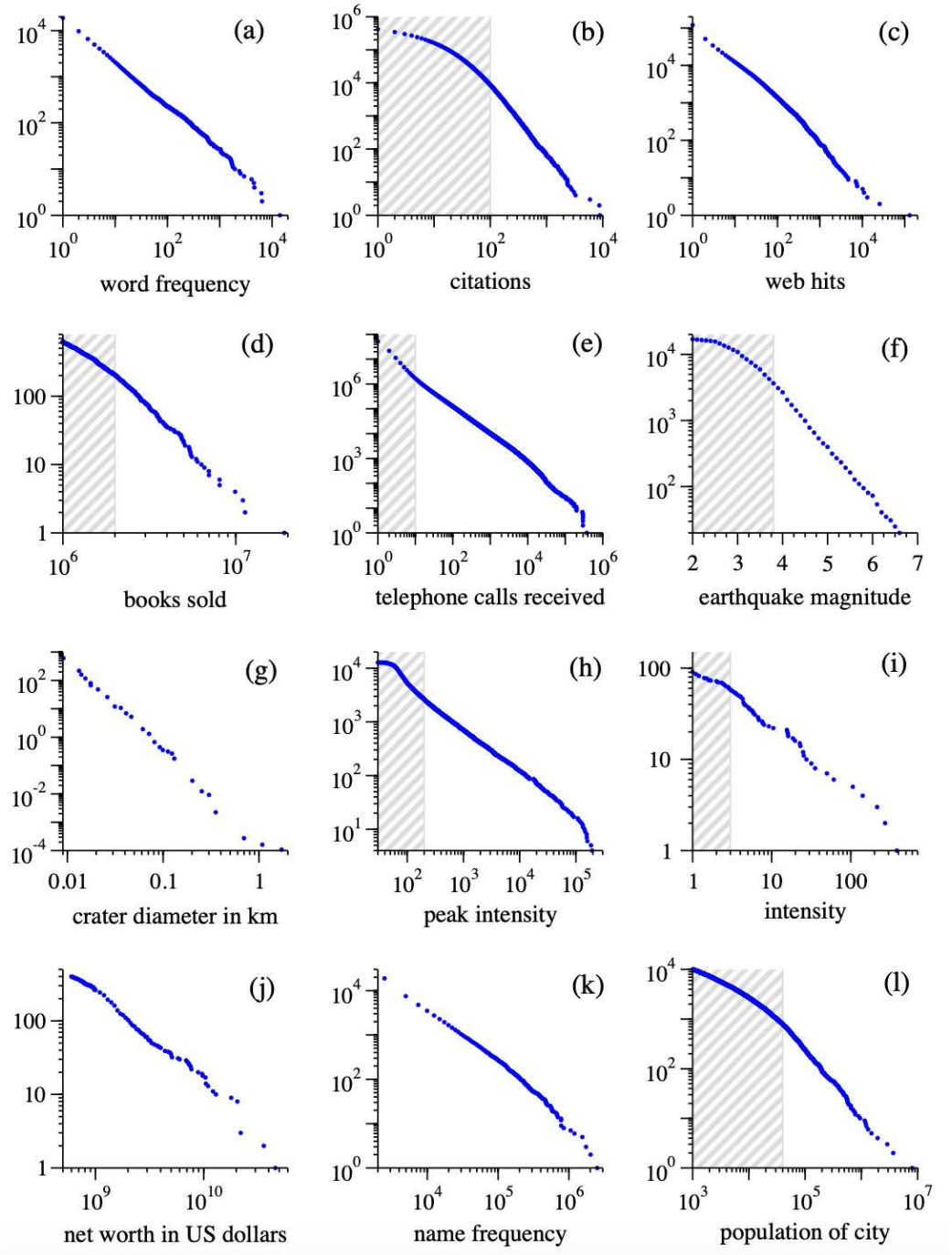

Figure 1.4: Cumulative distributions of twelve quantities which follow power laws, from [53]. Data in the shaded regions were excluded from the calculations of the exponents. (a) Numbers of occurrences of words in the novel Moby Dick by Hermann Melville. (b) Numbers of citations to scientific papers published in 1981, from time of publication until June 1997. (c) Numbers of hits on web sites by 60 000 users of the America Online Internet service for the day of 1 December 1997. (d) Numbers of copies of bestselling books sold in the US between 1895 and 1965. (e) Number of calls received by AT&T telephone customers in the US for a single day. (f) Magnitude of earthquakes in California between January 1910 and May 1992. Magnitude is proportional to the logarithm of the maximum amplitude of the earthquake, and hence the distribution obeys a power law even though the horizontal axis is linear. (g) Diameter of craters on the moon. Vertical axis is measured per square kilometre. (h) Peak gamma-ray intensity of solar flares in counts per second, measured from Earth orbit between February 1980 and November 1989. (i) Intensity of wars from 1816 to 1980, measured as battle deaths per 10 000 of the population of the participating countries. (j) Aggregate net worth in dollars of the richest individuals in the US in October 2003. (k) Frequency of occurrence of family names in the US in the year 1990. (l) Populations of US cities in the year 2000.

A common way to represent power-law distributions is through the cumulative distribution function (CDF) and the complementary cumulative distribution function (CCDF), both of which also exhibit power-law behaviour in the tails. The CDF is defined as:

$$P(x) = \int_{-\infty}^{x} p(x')\,\mathrm{d}x',$$

while the CCDF, which gives the probability of observing a value greater than x, is given by:

$$P_c(x) = 1 - P(x) = \int_{x}^{\infty} p(x')\,\mathrm{d}x'.$$

For power-law distributions of the form $p(x) = Cx^{-\alpha}$, defined for $x \geq x_{\min}$ with $\alpha > 1$, the CCDF takes the form:

$$P_c(x) = \int_{x}^{\infty} Cx'^{-\alpha},\mathrm{d}x' = \frac{C}{\alpha - 1}\, x^{-(\alpha-1)}. \tag{1.17}$$

This representation is particularly useful in empirical studies, as plotting the CDF or CCDF on a log-log scale yields a straight line, simplifying the identification and estimation of the exponent α.

Power-law distributions exhibit several noteworthy features. One of the most striking is that, unlike the Gaussian, they are not well-described by mean and variance due to potential divergence of moments. Secondly, they exhibit scale invariance (Eq. (1.13)).

In general, because of their heavy tails, power laws do not satisfy the assumptions of the classical Central Limit Theorem and hence they do not converge to a Gaussian distribution. However, when $\alpha \in (0, 2)$, they fall within the class of stable distributions discussed in Sec. 1.4.2, and the generalised version of the CLT still holds [38].

Power-law distributions appear across a wide array of disciplines, from demography and linguistics to physics, economics, and social sciences. Examples include the distribution of word frequencies in literary texts [58], scientific citations [59,60], city populations [53], earthquake magnitudes [61], wealth distributions and election results [62]. A selection of empirical distributions exhibiting power-law behaviour is shown in Fig. 1.4, adapted from [53]. These cases span several orders of magnitude and emphasise the ubiquity of power-law scaling.

1.5.2 Pareto Principle and Inequality Measure

The Pareto Principle, commonly known as the "80/20 rule", suggests that, in many systems, roughly 80% of outcomes result from 20% of causes. This empirical observation traces back to the pioneering work of Italian economist and sociologist Vilfredo Pareto, who in 1896 noted that approximately 80% of Italy's land was owned by just 20% of the population [63]. He formalised this insight using what is now known as the Pareto distribution, a mathematical expression of power-law behaviour. Later, engineer and quality management expert Joseph M. Juran generalised Pareto's observation to various industrial and economic settings, coining the term Pareto Principle to describe the tendency for a small proportion of inputs to generate the majority of outcomes [64].

The Pareto Principle has since become a strategic tool in diverse domains such as business management and operations research. For example, it is often observed that 80% of a company's profits come from 20% of its customers, or that a small fraction of products accounts for the majority of sales. Importantly, the percentages need not sum to 100%; rather, the key point is the disproportionate contribution of a small subset of elements to the total [65].

Mathematically, the Pareto distribution is a specific type of power-law distribution. Its CCDF follows:

$$P_c(x) = \left(\frac{x_{\min}}{x}\right)^{\alpha}, \tag{1.18}$$

for $x \geq x_{\min}$ and shape parameter $\alpha > 0$. The exponent α governs the tail heaviness: smaller values of α correspond to heavier tails and greater inequality. Pareto originally found $\alpha \approx 1.16$ when analysing income distributions in Italy.

Empirical evidence for Pareto-like behaviour has been found in a wide range of settings, most notably in the distribution of wealth. For instance, Klass et al. [66] analysed the Forbes 400 list (the ranking of the 400 wealthiest individuals in the United States) between 1988 and 2003. They found that the upper tail of the wealth distribution closely followed a Pareto law with an average exponent of approximately $\alpha = 1.49$. In this context, the relationship between individual wealth and rank is particularly significant: the data reveal that wealth w_r of the individual ranked r (with $r = 1$ for the richest) follows a scaling relation:

$$w_r = A\, r^{-\beta},$$

where A is a constant and $\beta = 1/\alpha$. Here, α directly quantifies inequality: smaller values indicate more concentrated wealth and thus greater inequal-

ity. Although the analysis focuses only on the wealthiest individuals, the findings provide compelling empirical support for Pareto's theory, especially when recognising that the Pareto distribution is scale-free, meaning the same relative pattern holds across scales.

This naturally leads to the need for formal tools to quantify inequality. One widely used measure of disparity is the Gini coefficient, introduced by Italian statistician Corrado Gini and Gaetano Pietra in the early 20th century [67, 68]. The Gini coefficient quantifies the deviation of a given income or wealth distribution from perfect equality among a group. It ranges from 0 (perfect equality) to 1 (maximum inequality), offering deeper insights into the structural disparities implied by Pareto-like distributions [69].

While several formulations of the Gini index exist, at this stage our focus is primarily conceptual. A more technical and operational definition will be further developed and quantitatively explored in Chapter 2, along with a focus on the role of ranking in the emergence and perception of inequality.

1.5.3 Preferential Attachment and the Matthew Effect

The pervasive presence of heavy-tailed distributions and systemic inequalities in socio-economic systems naturally leads to the investigation of the mechanisms that give rise to such patterns. One influential explanation is preferential attachment, which offers a generative model for how disproportionate outcomes can emerge from seemingly neutral processes. This mechanism helps explain the dynamics underlying many power-law distributions, including those observed in social, economic, and technological networks.

In his seminal 1955 paper, Simon [58] proposed a common underlying mechanism to explain the emergence of similar power-law distributions across diverse domains. He introduced a stochastic model that formalised what is now commonly referred to as the preferential attachment mechanism, often paraphrased as the "rich-get-richer" dynamic.

In this process, the probability that an entity (such as a word, a city, a scientific paper, or a node in a network[3]) attracts additional connections or instances is proportional to its current frequency or size. This principle was later extended to the context of growing networks, where new nodes preferentially attach to already well-connected nodes [70], thereby incorporating temporal dynamics. In such systems, the degree distribution (i.e., the prob-

[3]In this context, a *network* refers to a collection of *nodes* (or vertices) connected by *edges* (or links). Nodes typically represent entities such as individuals, cities, or webpages, while edges represent relationships or interactions between them, such as friendships, trade routes, or hyperlinks. The structure and dynamics of these connections often reveal important patterns underlying complex systems.

ability $P(k)$ that a node interacts with k other entities) follows a power-law form:

$$P(k) \sim k^{-\gamma},$$

with γ typically ranging between 2 and 3 in empirical observations [71].

Preferential attachment exemplifies a self-organising, multiplicative process that leads to scale-free distributions, a hallmark of many complex systems in nature and society. These dynamics highlight how older or more connected nodes continue to accumulate links over time, while newer or less connected ones struggle to gain visibility. The resulting network structure is highly unequal, with a small number of nodes dominating the connectivity landscape, reinforcing the "rich-get-richer"phenomenon observed in real-world systems.

Closely related to preferential attachment is the so-called Matthew Effect, introduced by sociologist Robert K. Merton [72]. He used this term to describe the presence of inequality and the lack of meritocracy in the scientific system, driven by a mechanism of cumulative advantage whereby established, high-ranking scientists receive disproportionate recognition and resources compared to their lower-ranked peers, regardless of the actual merit of their contributions. The name derives from a passage in the Gospel According to St. Matthew, which reads: "For to everyone who has, more will be given, and he will have an abundance; but from him who has not, even what he has will be taken away."

Merton noted that scientific credit is often assigned to individuals rather than to research teams, despite the inherently collaborative nature of scientific work. As a consequence, senior or already prominent scientists tend to accumulate increasing recognition over time, while younger or lesser-known researchers are frequently overlooked. This pattern becomes particularly evident in cases of simultaneous independent discoveries, where credit is usually granted to the more established figure, overshadowing equally deserving but less visible contributors. Moreover, the reception of a scientific contribution often depends on who expresses it: high-ranking scientists are more likely to receive attention and acknowledgment, even when presenting ideas of comparable merit. In essence, those who are already well positioned continue to attract recognition, while new voices struggle to emerge.

In addition, Merton noted that established scientists are often aware of this bias: some express hesitation about attaching their name to a paper, fearing it might draw attention away from the main contributors, while not doing so risks the work going unnoticed altogether.

Importantly, Merton emphasised that this logic of cumulative advantage

extends beyond the academic sphere. Similar dynamics are observed in domains such as media, economics, and social status, where "the rich get richer at a rate that makes the poor become relatively poorer" [72].

1.6 Critical Phenomena

The mechanisms explored so far (e.g., preferential attachment and cumulative advantage) exemplify how microscopic interactions can generate macroscopic inequality. To deepen our understanding of such emergent phenomena, we now turn to one of the most influential frameworks in statistical physics: the theory of critical phenomena. Here, simple local rules among agents can drive systems toward dramatic collective shifts, as it happens in the pivotal Ising model.

1.6.1 The Ising Model

Originally proposed by Wilhelm Lenz and analytically studied by his student Ernst Ising in 1925 [25,73], the Ising model has become a foundational framework for understanding cooperative behaviour in systems composed of many interacting elements [74–76]. Owing to its apparent simplicity, it effectively captures the essential features of phase transitions and has found widespread applications beyond magnetism, spanning natural and social systems from financial markets to earthquake modelling [29, 77, 78].

At its core, the Ising model consists of a lattice of sites (or nodes), each associated with a binary variable $s_i = \pm 1$. In the original physical interpretation, these represent atomic spins that interact with their nearest neighbours, tending to align. A schematic representation can be seen in Fig. 1.5. In a wider sense, each s_i can represent any binary state: yes/no decisions, molecules vs. holes, buyers vs. sellers.

This local tendency toward order is opposed by thermal fluctuations, modelled by the system's temperature. The competition between these effects leads to rich macroscopic behaviour, including phase transitions in dimensions greater than one [79]. For a system of n spins, the total number of configurations is 2^n. A central quantity is the magnetisation:

$$M(\mathbf{s}) = \frac{1}{n} \sum_{i=1}^{n} s_i, \tag{1.19}$$

which measures the degree of order of a given configuration $\mathbf{s}$. It reaches its maximum when all sites share the same value, namely in the configuration

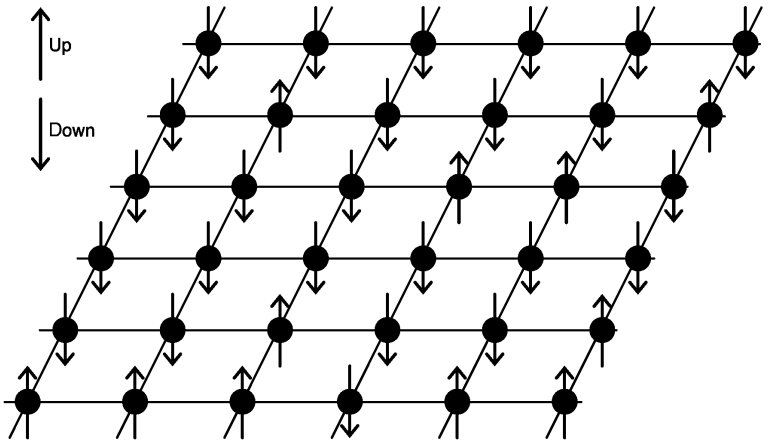

Figure 1.5: Schematic of a two-dimensional Ising model. Each node represents a spin that can point either up or down, with interactions favouring alignment between neighbours. From [80].

of maximum order.

The energy of a certain spin configuration (Hamiltonian) without an external field is:

$$H(\mathbf{s}) = -\sum_{\langle i,j \rangle} J_{ij} s_i s_j, \tag{1.20}$$

where $\langle i, j \rangle$ denotes summation over adjacent sites, and J_{ij} is the interaction strength (often set to a constant $J > 0$ for ferromagnetic interactions). This energy is minimized when neighbouring spins are aligned.

When the system is in contact with a thermal reservoir at temperature T, the probability of a configuration $\mathbf{s}$ is given by the Boltzmann distribution:

$$\mu(\mathbf{s}) = \frac{1}{Z} e^{-\beta H(\mathbf{s})}, \quad \text{with} \quad \beta = \frac{1}{k_B T}, \tag{1.21}$$

where $Z = \sum_{\mathbf{s}} e^{-\beta H(\mathbf{s})}$ is the partition function and k_B is Boltzmann's constant. Low-temperature regimes favour ordered configurations with low energy; at high temperatures, random configurations dominate. The central thermodynamic quantity associated with this probabilistic framework is the free energy, which characterises the macroscopic equilibrium behaviour of the system:

$$F(T) = -k_B T \ln Z. \tag{1.22}$$

The free energy plays a pivotal role in determining the dominant macrostates at a given temperature: through its minimisation, one can identify the most probable large-scale configurations of the system. Moreover, singularities or

non-analytic behaviour in the free energy or its derivatives signal the presence of phase transitions (Sec. 1.6.2), such as the onset of spontaneous magnetisation in the Ising model. This simple model displays a rich variety of behaviours: in two or more dimensions, it undergoes a continuous (second-order) phase transition at a critical temperature T_c. Below T_c, the system exhibits spontaneous magnetisation (global order); above T_c, disorder prevails. The presence of a critical point that causes the system to undergoing dramatic change is key to the discussion of Chap. 2.

Despite its simplicity, exact analytical solutions are available only in special cases, such as the one-dimensional chain, which was solved by Ising himself. He concluded that no phase transition was possible and, incorrectly, that the phenomenon was of no-interest. Analytical insights can be obtained also for the two-dimensional lattice. In more general cases, the model is explored using numerical simulations, especially Monte Carlo methods (Sec. 1.7.1) [74].

Beyond its original physical context, the Ising model has become a paradigmatic framework for social phenomena, due to the compelling analogy between magnetisation in condensed matter and polarisation in social systems. The same binary variables $s_i = \pm 1$ can be used to represent agents making dichotomous choice, such as supporting or opposing a political opinion, buying or selling a financial asset, or adopting or rejecting a social norm. Interactions between agents encode social influence, favouring alignment with peers, while external influences (analogous to temperature in the physical model) introduce randomness or individual autonomy. In econophysics, for instance, binary spins can represent asset returns ($s_i = +1$ for gains, $s_i = -1$ for losses) [81] or alternatively the decision to buy or sell [82]. This modelling framework has been successfully applied to understand financial bubbles, herding behaviour, market crashes, and even social segregation phenomena such as urban racial clustering [29].

Nearly a century after its introduction, the Ising model remains a cornerstone of complex systems theory for interacting systems, bridging microscopic rules and macroscopic phenomena, from magnets to markets and societies.

1.6.2 Phase Transition

Having introduced the Ising model as a foundational framework for interacting systems, we now turn to one of its most remarkable features: its ability to exhibit phase transitions. These transitions, driven by local interactions among components, provide a powerful paradigm for understanding abrupt, systemic changes across both physical and socio-economic domains.

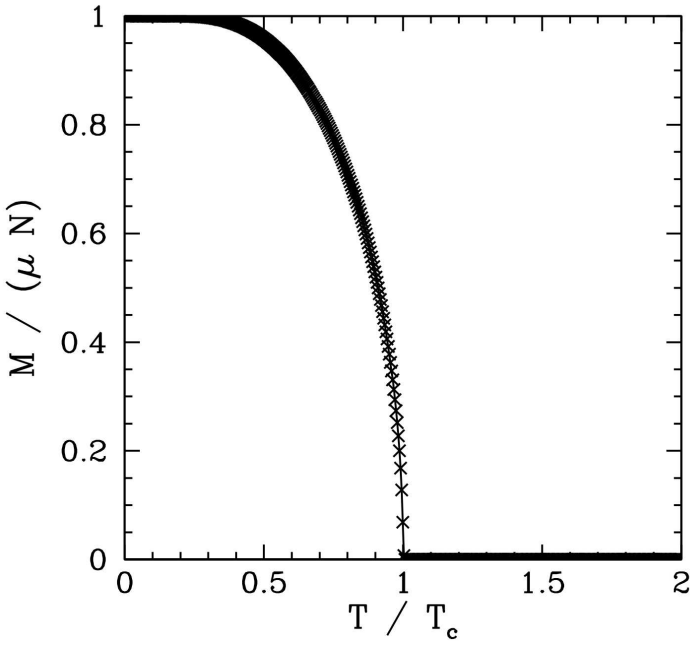

Figure 1.6: The net magnetisation M of a collection of N ferromagnetic atoms as a function of the temperature T in the absence of an external magnetic field, from [83]. Calculations are performed using the mean field approximation.

A phase refers to a state of matter in which macroscopic properties are uniform throughout the system [9]. A phase transition occurs when a small variation in an external parameter (e.g., temperature or pressure) induces a qualitative, dramatic change in these properties. Mathematically, phase transitions are associated with non-analytic behaviour in the free energy, defined in Eq. (1.22), or its derivatives [84]. In first-order transitions, the first derivative of the free energy is discontinuous at the critical point, while in second-order (or continuous) transitions, the discontinuity appears in the second derivative.

A canonical example of a this phenomenon is the liquid–gas transition, where abrupt changes in density occur at the coexistence curve. In magnetic systems, analogous behaviour is seen as magnetisation emerges below a critical temperature T_c, marking the onset of spontaneous magnetic order. Unlike first-order transitions, the magnetisation in models such as the Ising model vanishes continuously at T_c, characteristic of a second-order phase transition, as shown in Fig. 1.6. At the critical point, small variations in temperature or external field can induce large fluctuations in macroscopic observables, reflecting the system's increased sensitivity due to long-range correlations [23, 85].

The simplicity of the Ising model belies its depth: it captures how local interactions among constituents can generate global order. Statistical mechanics, in turn, offers a powerful lens for studying analogous transitions in socio-economic systems, where individual decisions are shaped by peer influence. This analogy is especially useful in modelling binary choices (such as whether to vote, smoke, or adopt a certain behaviour) where social interactions can result in abrupt shifts in collective outcomes. Such dynamics can produce polarisation, with the emergence of cohesive groups in the population with divergent behaviour [86]. In fact, while many social systems are continuous in nature (i.e., gradual changes in input yield gradual changes in outcome) under certain conditions they can experience sharp, systemic transitions without a substantial external trigger. These are referred to as social phase transitions [87–91], and they closely resemble second-order transitions in physical systems like the Ising model. Examples include spontaneous traffic jams [92], sudden outbreaks of cooperation [93], racial segregation [94], and financial crashes [95]. Two key ingredients are typically present: agents' propensity for social conformity and limited heterogeneity across the population [96]. Heterogeneity can include differences in preferences, thresholds for action, information processing, or susceptibility to peer influence [86, 97]. Low heterogeneity plays a role analogous to low temperature in statistical physics: it fosters large-scale coordination and increases the likelihood of abrupt change. For instance, the 1987 stock market crash has been linked to excessive homogeneity in traders' expectations [98]. Brock and Durlauf [99] apply tools from statistical physics to formalise such transitions in economics: agents play the role of particles and their choices are influenced by personal utility and choices of their neighbours, in order to explain phenomena such as sudden shifts in asset prices or volatility. Multiple equilibria can emerge and large-scale shifts in returns, volatility or demand can be due to small changes in expectations or policies [86]. In this context, a particularly elegant framework is offered by Moshe Levy [96], who introduced a deterministic model for endogenous collective shifts in binary decisions.

Together, these insights provide a rigorous foundation for understanding abrupt shifts in social systems, linking micro-level preferences to macro-level outcomes using the language and tools of statistical physics.

1.6.3 Mean-Field Approximation

As discussed in Sec. 1.6.1, the Ising model provides a compelling framework for describing how macroscopic order can emerge from microscopic interactions. However, the number of possible configurations for a system with n

binary variables grows exponentially as 2^n, precluding the possibility of exact analytical solutions in most cases. A few special cases admit exact results (e.g., the one-dimensional chain or the two-dimensional square lattice without an external field); yet, in general, approximate methods must be adopted to analyse the system.

Among these, the mean-field approximation occupies a central role. The key idea is to replace the explicit interactions between individual components with an average or "mean" effect, thereby reducing the complexity of the system while preserving its essential collective behaviour. Rather than considering the influence of each spin's neighbours explicitly, one assumes that each spin interacts with an effective field generated by the average state of the system. This approach reduces the exponential computational burden to a more tractable polynomial scaling, typically of the form $\mathcal{O}(n^\alpha)$ with $\alpha \geq 0$ [9]. Mean-field approaches offer significant qualitative insights into the behaviour of complex systems and, in certain regimes (particularly in high-dimensional settings or fully connected models) can produce quantitatively accurate predictions despite their simplifying assumptions.

Notably, the applicability of mean-field methods extends well beyond traditional physical systems and has proven particularly effective in modelling socio-economic phenomena. Just as spins tend to align under a mean field, agents' decisions are influenced by the average behaviour of the population. This analogy enables the analytical treatment of interacting-agent systems that would otherwise be intractable. In this context, particularly successful are the models developed within the framework of mean-field game theory [100–102], where collective patterns can be understood as emergent outcomes of decentralised, local interactions among heterogeneous agents. Examples are crowd dynamics, financial herding, or opinion alignment [103].

For these reasons, the mean-field approximation continues to serve as a foundational tool in the study of critical phenomena, offering a balance between analytical tractability and physical interpretability.

1.6.4 Ergodicity Breaking

The previous explorations on phase transitions and collective phenomena highlighted how microscopic interactions can give rise to emergent macroscopic order. In this context, a particularly significant concept in statistical physics is ergodicity.

The ergodic hypothesis, initially proposed by Boltzmann in the context of thermodynamics, states that a system's trajectory in phase space will, over time, traverse all accessible microstates consistent with its conserved quantities, spending equal time in each. A more rigorous formulation by Ehrenfest

consorts later specified that, in the limit of infinite time, all microstates associated with a given macrostate are visited with equal probability [26]. When valid, this assumption provides a powerful simplification: it allows the replacement of time averages of observables with ensemble averages, thus eliminating explicit time dependence in many dynamical calculations. Formally,

$$\lim_{T\to\infty} \frac{1}{T} \int_0^T f(\omega(t))\, \mathrm{d}t = \int f(\omega)\, P(\omega)\, \mathrm{d}\omega, \tag{1.23}$$

where $\omega(t)$ denotes the system's state at time t, $f(\omega)$ is a function of the system's state, and $P(\omega)$ is the invariant distribution of the stochastic process.

Nevertheless, many complex systems operate under conditions where ergodicity does not hold, a phenomenon known as ergodicity breaking [9]. In general, this is due to high sensitivity to initial conditions or perturbations, resulting in a smaller portion of the phase space to be explored.

Although originally developed in the context of physical systems, the concept of ergodicity breaking naturally extends to settings involving interacting agents. In economic modelling, for instance, ergodicity is often implicitly assumed; that is, the average behaviour across a population is presumed to be representative of the temporal evolution of a single agent. However, in systems characterised by nonlinearity and interaction, this equivalence may no longer be valid. Under such conditions, the long-run dynamics can exhibit multiple equilibria, the selection of which depends crucially on initial conditions or random fluctuations.

Empirical evidence increasingly highlights the divergence between ensemble and time averages in financial and decision-making contexts. Human behaviour is shaped not only by rational expectations but also by bounded rationality, heuristics, emotional responses, and subjective beliefs [104]. These behavioural factors introduce complexities that cannot be adequately captured within an ergodic framework. As a result, deviations from ergodicity can significantly affect both theoretical predictions and the reliability of policy-relevant forecasts [105, 106]. Empirical analyses further support this view, revealing systematic discrepancies between expected values and experienced outcomes [107]. For example, recent studies have shown that individual preferences under uncertainty often diverge from those predicted by models based on ergodic assumptions [108].

In summary, ergodicity breaking constitutes a critical concept in the study of complex systems, particularly those involving interaction, heterogeneity, and feedback. Its recognition enables a deeper understanding of the limitations of conventional modelling approaches and motivates the development of alternative frameworks better suited to describe multifaceted behaviour.

1.7 Computational Methods

Computational simulation has emerged as a crucial methodological tool in the study of complex systems. Computer-based methods help overcome analytical intractability and allow researchers to explore a broader spectrum of dynamics, including non-linearities, feedback loops, and heterogeneity, delivering both qualitative and quantitative insights. They function as a virtual laboratory, enabling the systematic exploration of hypotheses and the testing of scenarios in ways that are often impractical, costly, or impossible to conduct through real-world experimentation, helping identify potential inconsistencies or oversights [109]. Today, simulations are widely employed across academic disciplines and applied domains, from economics and sociology to epidemiology and political science.

In this section, we focus on two widely used computational approaches: Monte Carlo methods and agent-based models.

1.7.1 Monte Carlo Techniques

From a computational perspective, Monte Carlo methods play a central role in the practical implementation of statistical physics. These techniques encompass a broad class of stochastic algorithms designed to tackle mathematical and physical problems that are analytically intractable or computationally intensive [32]. At the heart of every Monte Carlo simulation lies a seemingly paradoxical requirement: the generation of random sequences using deterministic algorithms [110]. Once this requirement is satisfied, these methods can be applied to both deterministic and probabilistic systems. In the latter, randomness reflects intrinsic uncertainty.

Monte Carlo simulations generate large ensembles of independent samples to estimate statistical observables such as mean values, variances, and correlation functions. A classical example is found in financial modelling and risk assessment (see Ch. 3). In the former, randomness is introduced not to reflect uncertainty but as a computational strategy to approximate complex quantities. For example, numerical integration over high-dimensional spaces benefits greatly from Monte Carlo sampling, whereas traditional methods fail due to the exponential growth of computational complexity with dimensionality.

The historical development of Monte Carlo methods is closely associated with the pioneering contributions of Ulam, von Neumann, Metropolis and Fermi. Two key historical events underpin the method's conceptual devel-

opment: the construction of the first general-purpose electronic computer (ENIAC) and the Manhattan Project, the code name for the atomic weapons program [111, 112]. Fermi had experimented with statistical sampling techniques as early as the 1930s, though he did not publish his results, whereas Ulam independently formalised the method in 1946 while recovering from an illness and playing solitaire. The foundational paper, *The Monte Carlo Method* by Metropolis and Ulam (1949), publicly introduced the method for the first time. According to Metropolis, the name "Monte Carlo" was coined around that time and was inspired by Ulam's uncle, who used to borrow money from his relatives to gamble in the casinos of Monte Carlo, in the Principality of Monaco. [113].

A fundamental requirement of Monte Carlo simulations is the ability to generate sequences of uniformly distributed random (or, more accurately, pseudorandom) numbers in the interval $[0, 1]$. From this base distribution, generalisations to other ranges and probability distributions can be easily implemented, for instance by inverting the cumulative distribution function, which itself ranges over $[0, 1]$.

True random numbers are generated by inherently unpredictable physical processes. They are non-reproducible: the shortest possible description of a truly random sequence is the sequence itself. Examples of true randomness can be found in the natural world, such as the timing of atomic decays. By contrast, pseudorandom numbers are generated by deterministic algorithms that produce sequences which, while statistically indistinguishable from true randomness in most practical applications, are completely reproducible when initialised with the same seed. This reproducibility is particularly valuable in scientific computing, where debugging, testing, and validation require controlled experimentation. The earliest known pseudorandom number generator was the middle-square method, developed by von Neumann. In this method, an n-digit number is squared, and the central n digits of the result are extracted to produce the next number in the sequence.

Modern implementations have replaced early algorithms. One of the most widely used pseudorandom number generators is the Mersenne Twister, developed by Makoto Matsumoto and Takuji Nishimura in 1997 [114]. Implemented in Python, R, and MATLAB, it boasts a long period (i.e., the minimum sequence length before repetition) of $2^{19937} - 1$ and excellent statistical properties, including high computational efficiency.

Monte Carlo methods offer several practical advantages [115]. They can replace physical experiments, thus saving time and cost, enhancing safety, and eliminating the need for physical prototypes. Moreover, their flexibil-

ity allows them to handle nonlinear systems, stochastic processes, and high-dimensional integrals with relative ease. Since their inception, these methods have found application in a broad spectrum of domains: from nuclear and particle physics to quantitative finance [116], thermodynamics [117,118], meteorology, and project management [119]. In the latter, Monte Carlo help evaluate uncertainties in budgeting and scheduling. As Metropolis prophetically wrote in 1987: *"Where will all this lead? If one were to wax enthusiastic, perhaps—just perhaps—a simplified model of the brain might be studied... These studies, in turn, might provide feedback to computer architects designing the new parallel structures"* [113].

Despite their versatility, Monte Carlo methods are not without limitations. One significant drawback is their computational inefficiency: the convergence of estimators typically improves as $1/\sqrt{N}$, where N is the number of samples. Thus, reducing error requires large-scale simulations that consume significant computational resources, leading to substantial energy consumption and operational costs, particularly in high-performance computing environments. Additionally, challenges such as poor convergence for rare event estimation, statistical noise due to finite sampling, rounding and discretisation errors can impact result reliability [118]. Indeed numerical precision must be carefully managed. Truncation errors arise from approximating real numbers with a finite number of digits, while round-off errors accumulate from repeated finite-precision arithmetic operations [120]. To address these issues, various enhancements have been developed. For example, in computationally demanding scenarios, Monte Carlo methods benefit greatly from parallel computing architectures. Modern implementations often leverage multi-core processors or GPUs to accelerate sampling and analysis. Additional techniques such as variance reduction, importance sampling, and multilevel Monte Carlo strategies can improve convergence without incurring proportional increases in computational cost [84, 121]. For further discussion on Monte Carlo methods, including statistical testing of pseudorandom generators and numerical implementation guidelines, the reader is referred to [120].

In the following, Monte Carlo methods will be introduced and discussed in detail wherever their application is pertinent. We now turn our attention to agent-based models, a further computational tool which is widely adopted in a variety of fields and particularly central to the present research.

1.7.2 Agent-Based Models

Agent-Based Models (ABMs) are computational frameworks designed to simulate complex systems through the interaction of autonomous entities known as agents [122]. These agents can represent individuals, groups, firms, or institutions, each characterised by a set of properties and behavioural rules that govern their actions within a defined environment. Through these interactions, agents can evolve over time, producing emergent system-level outcomes. The ABM approach is inspired by principles of statistical physics: start with a minimal set of simplifying assumptions, explore the implications of these assumptions across multiple simulations, and possibly compare the emergent results with empirical data [123, 124].

The appeal of ABMs lies in their high degree of customisability. Agents may be rational or irrational, homogeneous or heterogeneous, and may interact globally (with the entire population) or locally (with nearest neighbours only). Behavioural rules can range from deterministic equations to heuristic, rule-based logic. Importantly, ABMs naturally accommodate heterogeneity (i.e., individual variation in behaviour), and stochasticity (i.e., random influences on decision-making) [125]. Factors such as emotions, beliefs, and subjective judgement are often difficult to formalise, yet are essential for capturing many phenomena regarding individuals and can be introduced in the simulation through specific parameters. This flexibility allows ABMs to replicate a wide range of realistic system behaviours that are often analytically intractable and to simulate complex dynamics [123].

As a bottom-up modelling approach, they facilitate the exploration of how micro-level behaviours and interactions give rise to macro-level patterns [109], exemplifying the principle that a system is "more than the sum of its parts" (Sec. 1.1). One example in this direction is the study of abrupt transitions arising from interactions among agents. This often involves identifying the type of phase transition (Sec. 1.6.2) the system undergoes, whether it is a continuous or a discontinuous one. These dynamics are typically plotted using phase diagrams, which delineate the parameter regions associated with distinct system states. A particularly intriguing scenario is multi-stability, where multiple stable equilibria exist, and the outcome depends heavily on initial conditions or system history [125, 126].

From a computational point of view, ABMs serve as laboratories for scenario testing, policy evaluation, and hypothesis generation [123, 124]. The dynamics is typically implemented through discrete-time simulations and, in this context, the choice of time step Δt is non-trivial: if not properly calibrated, it may introduce numerical artefacts that distort system dynamics.

More generally, ABMs must be constructed with careful attention to the appropriate level of abstraction, as overly detailed models risk becoming computationally intractable, while oversimplified ones may fail to capture essential dynamics [127].

The origins of ABM in the social sciences can be traced to Schelling's seminal 1971 work on residential segregation, which demonstrated how simple individual preferences could result in large-scale patterns without central coordination [29,94]. This idea was further developed by Epstein and Axtell in their Sugarscape model, which explored the emergence of artificial societies [128]. The key innovation in these models was the incorporation of individual-level decision-making as a driver of macro-level dynamics. The broader adoption of ABMs was facilitated by the advent of accessible personal computing resources in the 1980s, analogously to the analysis and modelling of financial data.

Today, ABMs are widely applied across disciplines, from sociology to epidemiology and political science. For example, they are adopted to model pedestrian flows, analysing supply chains, forecasting consumer behaviour or epidemic spread [125, 127]. Their versatility stems from the ability to encode agents with diverse objectives and cognitive capacities, operating in structured environments such as grids, lattices, or networks (see Sec. 1.8). In economics, specifically, ABMs are used to model agents heterogeneous and boundedly rational, in contrast to classical economic models that assume agent homogeneity and perfect rationality [129] and have been instrumental in reproducing empirically observed phenomena such as market bubbles, crashes, volatility clustering, heavy-tailed return distributions, and systemic risk [130].

In summary, agent-based models provide a powerful framework for investigating complex systems by simulating the behaviour of individual components and observing the resulting emergent phenomena. The modelling process typically begins with simplifying assumptions to isolate the system's essential features, which are encoded into agent behaviours and interaction rules, allowing simulations to be run under varying parameters. The outcomes can then be compared against empirical data to assess the model's validity and refine its structure.

A key advantage of ABMs is their compatibility with Monte Carlo methods: by repeatedly simulating a system under stochastic conditions and averaging the results, researchers can obtain statistically robust insights into the system's dynamics. This iterative process (modelling, simulating, evaluating, and refining) enables the identification of critical mechanisms underpinning

Figure 1.7: Core network of the United States airport connections. From [139].

observed real-world phenomena, ultimately deepening our understanding of complex systems.

1.8 Network Concepts

In recent decades, one of the most powerful and versatile representations of complex systems has emerged from the formalism of networks [131–134]. Originally rooted in the mathematical domain of graph theory, networks provide a flexible yet rigorous language to describe diverse systems composed of interconnected components. Whether the focus is on electric power grids, food webs, neural circuits, social relations, or the architecture of the Internet, networks offer a unifying framework to capture the essence of structure, interaction, and emergence [135–138].

Formally, a network (or graph) is defined by a set of nodes (also called vertices), representing the fundamental units of the system, and a set of edges (or links), describing pairwise relationships or interactions between them. Connections can be unilateral (directed networks) or bilateral (undirected networks). Several basic concepts underpin the study of networks and are essential for interpreting their structure and dynamics:

- Degree: the number of connections a node has. It serves as a measure

of centrality and influence;

- Clustering coefficient: quantifies the tendency of a node's neighbours to be connected;

- Path length: the number of steps required to connect one node to another. Networks with short average path lengths exhibit the so-called small-world property;

- Centrality measures: various metrics can be adopted to assess the importance of nodes beyond simple degree;

- Percolation: how the structure and functionality of a network change as nodes or edges are randomly or selectively removed, with particular focus on the critical thresholds at which a giant connected component (i.e., a subgraph in which a positive fraction of the entire network's nodes are reachable from each other) emerges or collapses.

Given a limited set of elements (nodes and links), networks can nevertheless capture a wide range of real-world systems, which often exhibit recurrent structural features. A prominent example is the small-world property, whereby most nodes can be reached from any other through a small number of steps despite strong local clustering, enabling rapid diffusion of information or influence even when interactions are predominantly local [133]. Closely related to this phenomenon is the role of weak ties, broadly defined as connections of low intensity, weight, or persistence that nonetheless link otherwise weakly connected regions of the network and facilitate the circulation of novel information or resources; in social networks, these often correspond to relationships between acquaintances rather than close friends [140].

Networks may further display heterogeneous connectivity patterns, leading to scale-free degree distributions in which most nodes have few connections, while a small number of hubs accumulate many links. Such structures are typically associated with growth processes driven by cumulative advantage or preferential attachment (Sec. 1.5.3) [70]. At an intermediate, mesoscopic scale, networks often organise into communities, defined as groups of nodes that are more densely connected internally than externally. These structures reflect functional or organisational differentiation and play a key role in shaping collective behaviour, while weak ties can be crucial in linking distinct communities and ensuring efficient diffusion across the network.

The simple representation of networks reveals rich behaviours, such as the identification of hubs (i.e., nodes with significantly higher connections than average), which has been proven crucial in understanding phenomena such

as viral diffusion, systemic risk, and vulnerability to attacks. Conversely, peripheral nodes with low connectivity may be locally insignificant but can still play key roles in maintaining the global functioning. Networks also are key in the study of cascading failures, where a local disruption (such as a bank default or a transmission fault) can propagate through the system, potentially leading to global collapse. Such features have motivated the design of fault-tolerant systems in domains ranging from banking to epidemiology and air traffic management [141–146].

Several types of networks are now widely adopted. A turning point for the emergence of random graph theory was in the 1950s, particularly through the seminal work of Erdős and Rényi [147]. In the Erdős–Rényi (ER) model, a network is constructed by connecting each pair of nodes with a fixed probability p, independent of all other pairs. This simple construction gives rise to well-defined statistical properties, including a degree distribution that follows Poisson distribution (where the probability of a node having a certain number of connections is sharply centred around the average degree) and a critical threshold for the emergence of a giant connected component.

However, many real-world systems deviate markedly from the homogeneity of ER graphs. Empirical observations across disciplines (from the World Wide Web to protein–protein interaction networks) consistently show heterogeneous degree distributions, with most nodes having few links and a small number acting as highly connected hubs. This property is captured by the Barabási–Albert (BA) model, which introduced the generative mechanism known as preferential attachment [70]. In this model, networks grow incrementally: at each step, a new node is added and connects to existing nodes with a probability proportional to their current degree. This "rich-get-richer" mechanism (Sec. 1.5.3) reflects the intuitive idea that well-connected nodes are more likely to attract new links, reinforcing their centrality over time. Empirical networks often exhibit approximate power-law degree distributions, typically with exponents in the range $2 < \alpha < 3$ [148], although real-world deviations are common. Such scale-free networks lack a characteristic degree scale and exhibit ultra-small-world properties, where average path lengths grow logarithmically or even slower with network size.

This structural feature has deep implications. It enhances robustness against random failures, as most nodes are not critical for connectivity; yet it also introduces vulnerability to targeted attacks, where removing a small fraction of hubs can fragment the network. Moreover, preferential attachment embodies a form of path dependence, as early advantages in connectivity can lock nodes into dominant positions. This phenomenon is typical of economic competition, academic citations, and online influence. Beyond

structural characterisation, scale-free models offer insight into the underlying dynamics (such as competition, cumulative advantage, and feedback loops) that govern the evolution of complex systems.

Although the Barabási–Albert model captures key macroscopic features of real networks, it does not account for mesoscopic structures like communities or modules, where nodes are more densely connected internally than externally. Such structures are especially prominent in social, biological, and technological networks, where modularity often reflects functional differentiation. To model these intermediate scales, one can employ the Stochastic Block Model (SBM) [149, 150]. The SBM generates networks by assigning nodes to hidden groups and specifying connection probabilities based on the pairwise group affiliations. This allows for the generation of networks with clear modular structures, hierarchical organisation, or even assortative mixing patterns, where nodes are more likely to connect to others with similar attributes or group memberships.

The network paradigm represents a profound shift in the way complexity is studied and interpreted, blending mathematical elegance with empirical relevance and finding fertile ground across disciplines. Networks may even offer insights into historical phenomena, such as the spread of rumours during the Great Fear uprising in France in 1789 [151]. Overall, networks remain one of the most fruitful metaphors for thinking about complex systems in the modern scientific landscape.

Chapter 2

Models under Social Influence

Individuals continuously shape their behaviour and are influenced by peers in their opinions, actions, and decisions. Broadly speaking, people strive for social approval [152] and, when their opinions align with those of others, the resulting sense of social validation tends to reinforce existing beliefs. Conversely, exposure to opposing views can induce discomfort, weakening their attachment to prior convictions. As Keynes observed, we often prefer to fail by conforming to the crowd rather than succeed alone [153]. Imitation is deeply rooted in human behaviour and constitutes one of learning strategies, alongside trial-and-error scheme. As a result, we often tend to adopt the opinions or behaviours of those we interact with [154–156].

In recent decades, physicists have increasingly applied tools from statistical physics to the study of social phenomena, aiming to uncover macroscopic regularities that emerge from microscopic interactions. Since the pioneering work of Galam in 1982 [157], spin-spin interactions have been adopted as an effective abstraction of social influence, enabling a rigorous analysis of collective opinion formation. As a result, the field of opinion dynamics has become a central and growing area of research at the interface of physics, sociology, and computational social science [28, 30, 31, 154, 158–160]. Stylised models have yielded insights into real-world scenarios, such as how a small number of influential individuals can sway the public opinion [161] or how key pieces of information, strategically delivered, can shift collective beliefs [162].

Different social phenomena can often be described using the same underlying formalism and, in this framework, agent-based models (Sec. 1.7.2) play a pivotal methodological role. Opinions are typically encoded as discrete or continuous variables [30, 79, 123] and agents iteratively update their states through social interactions. These dynamics reveal how global patterns of opinion can emerge: consensus, when all agents align; polarisation, when the population splits into opposing camps; or fragmentation, where multiple

opinions coexist [163]. A crucial ingredient in these studies is the topology of interactions, namely the structure of the connections through which agents influence one another [164]. Depending on the modelling framework, agents may be arranged on regular lattices or grids, embedded in spatial environments, or connected through complex networks that capture varied patterns of interaction.

This modelling approach has proven essential for understanding a wide range of collective behaviours, from the spread of infectious diseases to political opinion shifts. Contemporary examples include public discourse around Brexit and electoral polarisation in the United States. In economic contexts, agent interactions help explain stylised facts such as scaling behaviour in financial returns [165, 166], market crashes [167], and sudden, system-wide transitions [168].

2.1 Research Motivation

Given the above premises, our original research works in [1] and [2] contribute to the growing body of literature in opinion dynamics, the study of how individual opinions evolve through social interactions and influence. We implement stylised agent-based models to explore how individuals (and society as a whole) are influenced by peers in their decision-making processes. Both models rely on a fixed number of available states for the agents, who interact globally (i.e., with all the individuals in a reference group [169]), within a discrete-time framework governed by payoffs associated with each agent's action, and on probabilistic dynamics. The resulting outcomes are analysed through a set of carefully defined measures, designed to capture key features such as consensus, polarisation, inequality, and the distribution of agents among states.

More specifically, in [1] we analyse how the interplay between social influence and external information (considered as a ground truth) leads agents to adopt one of two possible states, with the aim of understanding under which conditions agents are able to correctly track the external signal. The states may represent binary decisions such as voting or adopting/non-adopting certain habits, such as smoking [170].

The model reveals a sharp transition between two distinct regimes: an ergodic regime when social influence is weak and a non-ergodic regime when peer influence is stronger. Intuitively, the latter case means that the personal trajectory of an individual can diverge significantly from the population average, resulting in fragmented outcomes. At the societal level, this marks a

transition from a flexible, adaptive system to one characterised by persistent consensus or division. For example, in an ergodic regime, individuals might shift political or cultural stances over time, whereas in a non-ergodic regime they become locked into persistent identities, echo chambers (where exposure to opposing views no longer leads to belief revision), or social norms. We further test the robustness of these findings by implementing the model on different network structures, where interactions become local and agents are influenced only by a limited number of neighbours.

The second study, presented in [2], explore the common human tendency to rely on rankings and to assume that top-ranked options (e.g., popular fields or social trends) are intrinsically superior. While we tend to associate high rankings with greater skill or merit, strong evidence suggest that luck, more than talent, often determines success [171] and that following one's own strategy may yield better outcomes compared to imitating those of top-ranked individuals [172].

In this context, we examine whether society functions more virtuously when agents pursue the latest trends (i.e., they imitate others under strong social influence) or instead engage in exploration of the available states or opinions, both in well-mixed populations and in networked interaction structures. The results are striking: when imitation dominates, the majority of the societal payoff is concentrated in the hands of a minority of agents, who are not necessarily the most talented. In contrast, when agents rely on random exploration, the resulting society is more egalitarian and meritocratic.

Some codes developed to analyse these models are publicly available at:

- `github.com/fdedo/Ergodicity-breaking`;
- `github.com/fdedo/Imitation-vs-serendipity`.

This chapter is organised as follows. Section 2.2 introduces the agent-based model in which agents can occupy binary states, with the aim of studying the interplay between social influence and external information in shaping individual opinions. In Section 2.2.3, we investigate two different sensitivity regimes: one where agents pay close attention to payoffs and one where they do not, considering both memory-less and memory-based dynamics. We find that the latter regime provides more insightful results and, thus, we focus on it. In this regard, Section 2.2.4, presents a mean-field approximation applicable to memory-based scenarios and analytical insights, which helps characterise the phase transition observed in the dynamics. Section 2.2.5 discusses possible real-world validations of our models. A brief summary of

results is provided in Section 2.2.6.
We then turn to the second model in Section 2.3, which investigates the interaction between imitation, randomness and ranking. Our aim is to determine whether individuals contribute to a more or less virtuous society by following current trends rather than freely exploring the available options. Section 2.3.2 presents the model and Section 2.3.3 explores several outcomes, including inequality, total wealth, meritocracy, and the distribution of agents across available actions. Analytical insights obtained from a simplified version of the model are discussed in Section 2.3.4 and alternative scenarios are examined in Section 2.3.5. Section 2.3.6 extends the dynamics to network settings, further confirming the robustness of our findings. Results are then summarised in Section 2.3.7.

Finally, Section 2.4 concludes the chapter, providing an overall summary of results from both models and outlines potential future directions.

2.2 Ergodicity Breaking in Social Dynamics

Following our original work [1], we focus on the role of social influence in shaping individuals' opinion when agents are simultaneously exposed to peer interactions and an external signal. The latter represents a sort of ground truth and our aim is to understand under which intensities of peer influence agents are able to correctly align with it.

2.2.1 Opinion Formation in Society

Human societies are intrinsically social and the views people hold are typically influenced by a mixture of interpersonal dynamics, cultural context, internal beliefs, and external sources of information [173]. Individuals strongly rely on others' opinions and on others' perception of themselves. Aligning with others often brings social validation and strengthens one's current stance, whereas diverging from group norms can lead to feelings of discomfort or social friction. When more options are available, it is often preferred the one that is more widely shared [161, 174].

The model presented in this work draws inspiration from the classical voter model, namely a particle system introduced to describe interaction between two populations that are competing for territory [175,176]. In addition, the dynamics bear strong conceptual resemblances to established models in statistical physics, particularly those involving binary-state systems such as the Ising model (Sec. 1.6.1) [177]. In such models, individual units (e.g., magnetic spins) adopt one of two states and interact locally with their neighbours.

These simple interaction rules give rise to complex macroscopic behaviours, including phase transitions, spontaneous symmetry breaking, and emergent collective phenomena [84, 178].

Our model retains several core features: agents choose between two discrete states, are influenced by their peers, and update their decisions according to stochastic rules [159, 179–181]. These ingredients capture the core mechanism of opinion dynamics, where social interaction and randomness jointly shape collective outcomes. However, our model significantly extends the reference frameworks in multiple directions. It incorporates an external signal (representing exogenous sources such as media or expert information) and agent updates their states through a weighted combination of social conformity and external input, rather than merely copying neighbours. The observed transition from an oscillatory regime to an absorbing one can be interpreted as a form of social phase transition, where the system shifts from a flexible, responsive state to one of stable consensus or polarisation. Finally, we introduce a memory-based extension that allows agents to integrate past information over time.

This analogy with physical systems provides a valuable theoretical framework for interpreting the emergence of collective behaviour in opinion dynamics and related socio-behavioural processes [182].

2.2.2 Agent-Based Model Implementation

Given the above framework, we now delve into the exploration of the interplay between social influence and external information [13, 28, 183]. We implement a discrete-time stochastic dynamics, in which agents are guided by payoffs when adopting one of two binary states and the evolution of the system unfolds probabilistically [184]. The specific model setup is in line with previous studies [185, 186].

- There are N agents in the system.
- Each agent can occupy one of two states, denoted by $+$ or $-$, which represent binary decisions, such as voting in a bipartisan election or adopting/not adopting a specific social habit.
- Social influence is represented by the fraction of agents in each state at time t, denoted by $f_+(t)$ and $f_-(t)$, with the constraint that

$$f_+(t) + f_-(t) = 1. \tag{2.1}$$

  at any time t. Since $f_-(t) = 1 - f_+(t)$, we focus our analysis on $f_+(t)$. When $f_+(t) \to 1$ or $f_+(t) \to 0$, agreement among agents emerge toward the $+$ or $-$ state, as they all share the same opinion.

- The weight of social influence in agent's decision is controlled by a parameter $\gamma \in [0, 1]$.

- External information $X(t)$ takes the value $+1$ with probability p and -1 with probability $1 - p$. This signal acts as a representation of a ground truth or external reference (e.g., expert advice, media).

- Each state is associated with a payoff $\Pi_{\pm}(t)$, depending on both the collective behaviour of the population and the current external signal. The payoff structure depends on whether the agent adopts a memory-less or memory-based strategy. In the memory-less version, the payoff at time t is:

$$\Pi_{\pm}(t) = \gamma\, f_{\pm}(t) \pm (1 - \gamma)\, X(t). \tag{2.2}$$

  Here, the $+$ payoff include a positive contribution from $X(t)$ when $X(t) = +1$ and a negative one when $X(t) = -1$. The opposite applies for the $-$ payoff. This implementation is in line with literature where payoff is higher when other agents behave similarly [169].
  In the memory-based version, the external signal is averaged over the last m time steps:

$$\langle X_m(t) \rangle = \frac{1}{m} \sum_{t'=t-m+1}^{t} X(t') \qquad \text{if} \quad t \geq m, \tag{2.3}$$

  and the payoff becomes:

$$\Pi_{\pm}(t) = \gamma\, f_{\pm}(t) \pm (1 - \gamma)\, \langle X_m(t) \rangle. \tag{2.4}$$

- At initial time $t = 0$, all agents are in the same state.

- At each time step $t = 0, \ldots, T - 1$, agents decide which state to occupy next according to a probabilistic rule. Specifically, they follow a Boltzmann distribution:

$$p_+(t + 1) = \frac{e^{\beta\, \Pi_+(t)}}{e^{\beta\, \Pi_+(t)} + e^{\beta\, \Pi_-(t)}}, \tag{2.5}$$

  with $p_-(t+1) = 1 - p_+(t+1)$. This update mechanism is widely adopted in decision-making models across physics [187], game theory [188] and social systems [169, 189].

- The parameter $\beta \in [0, \infty)$ in Eq. (2.5) controls the randomness of agent decisions and admits a dual interpretation. In statistical physics terms, it is the inverse of temperature. A high β (low temperature) results

in more deterministic behaviour, with agents consistently selecting options associated with higher expected rewards. However, suboptimal decisions are possible. A low β corresponds to higher randomness, where decisions are less influenced by payoff differences. In behavioural terms, β captures decision sharpness, i.e., how acutely agents differentiate between alternative payoffs. A high β indicates that even small differences in payoff lead to very different probabilities of choosing one state over another. A low β implies agents are more indifferent and prone to random choices.

- The structure of the payoff in Eqs. (2.2) and (2.4) ensures that the external signal dynamically pushes agents toward alignment with its current value. The symmetry of the payoffs, with a sign inversion between Π_+ and Π_-, ensures that agents respond in opposite directions depending on the sign of the signal. When the external $X(t) = +1$, the payoff Π_+ is higher than Π_-, encouraging agents to adopt or remain in state $+$. Conversely, when $X(t) = -1$, the signal favours state $-$. This reflects the tendency of agents to favour actions with higher payoffs, modulated by the "temperature" parameter β. In addition, it is clear that, as $\gamma \to 1$, peers' behaviour strongly influences the payoff associated to each action and, as a consequence, agent's behaviour. On the contrary, when $\gamma \to 0$, it is stronger the weight of the external information $X(t)$.

Remark. If not otherwise specified, simulations are performed with the following prescriptions, which ensures a suitable compromise between computational efficiency and the ability to capture meaningful population-level trends when reaching a steady-state regime. The number of agents is set to $N = 500$, the number of time iterations is $T = 1000$, the probability of the external signal $X(t)$ to be $+1$ is $p = 0.6$ and a number S of independent simulation is run. We analyse the low-noise scenario, corresponding to a high value of β (which, for practical purposes, we set to $\beta = 100$) and the high-noise scenario, with $\beta \in [0, 1]$. For both regimes, memory-less and memory based implementations are considered. For the latter, a time window of $m = 200$ steps is set to compute the average value in Eq. (2.3).

A final note concerns the symmetry of our system. The signal bias parameter p is defined over the full interval $[0, 1]$, yet the system exhibits a symmetry under the transformation $p \to 1 - p$. Specifically, reversing the signal bias is equivalent to inverting the external signal ($X \to -X$), which corresponds to relabelling the agent states. Therefore, we restrict our analysis to the range $p \in [0.5, 1]$ without loss of generality.

Throughout, we distinguish between agent states (denoted + and −) and the external signal $X(t) \in \{-1, +1\}$. Despite using similar symbols, they represent conceptually different quantities: the former describes individual decisions, while the latter captures external information.

The central focus of our analysis lies in comparing two key quantities. The first is the temporal average, which represents the mean fraction of agents in state + over the duration of a single simulation. Denoting each individual temporal average as $\langle f_+ \rangle_T^{(i)}$, where $i \in \{1, \ldots, S\}$ indexes the simulations, the temporal average is defined as:

$$\langle f_+ \rangle_T^{(i)} = \frac{1}{T} \sum_{t=0}^{T-1} f_+^{(i)}(t) \tag{2.6}$$

This quantity captures the long-term behaviour of the system along a single realisation of the stochastic dynamics.

The second quantity of interest is the ensemble average, obtained by averaging the temporal averages over multiple independent simulations. The ensemble average is given by:

$$\langle f_+ \rangle_S = \frac{1}{S} \sum_{i=1}^{S} \langle f_+ \rangle_T^{(i)} \tag{2.7}$$

This ensemble measure reflects the expected average outcome when sampling over different realisations of the stochastic process. Comparing the temporal average and the ensemble average allows us to assess whether the system exhibits ergodic behaviour: in ergodic regimes, the two averages should converge in the long-time limit, while persistent discrepancies indicate ergodicity breaking.

For clarity and brevity, explicit time dependencies (e.g., $X(t)$) may be omitted in the following sections when it does not affect the interpretation or outcome of the expressions.

2.2.3 Sensitivity Regimes

The study of the system begins by examining two extremal regimes: one of high sensitivity to payoff differences, defined by $\beta \to \infty$, and one of low sensitivity, i.e., $\beta \to 0$.

High-Sensitivity Regime

We start with the asymptotic case $\beta \to \infty$. In this high-sensitivity scenario, agents exhibit near-deterministic responses to differences in payoffs and, as such, they operate under minimal decision noise. In numerical simulations, this regime is approximated by setting $\beta = 100$, which has been found sufficient to capture the essential features of this low-noise setting. Additionally, tests indicate that the qualitative dynamics are preserved for other substantially large values of β, such as $\beta = 50$ or $\beta = 20$, confirming the robustness of the observed behaviours across a broad high-sensitivity range.

As introduced earlier, our focus is on the average fraction of agents in state $+$, denoted as $\langle f_+ \rangle$, computed both as temporal averages (Eq.(2.6)) and ensemble averages (Eq. (2.7)). Fig. 2.1 shows these quantities for several values of the social influence weight $\gamma \in [0, 1]$. The top panel refers to the memory-less version of the model (with payoff provided by Eq. (2.2)), while the bottom one to the memory-based dynamics (payoff defined in Eq. (2.4)). Both settings display a similar trend. For small values of γ, the system is ergodic (Sec. 1.6.4): the long-term behaviour of a single simulation reflects the typical behaviour across the ensemble. In this regime, which we call the oscillating phase, the population exhibits fluctuations between the two states, remaining responsive to the external signal. As γ increases, however, a critical threshold is reached (denoted γ^*) and the system undergoes a transition to a non-ergodic regime. Here, agents persist in a single state, whether $+$ or $-$, and remain there. This regime is called the absorbing phase, characterised by bistability and sensitivity to early dynamics (in the memory-less case) or cumulative signal history (in the memory-based scenario).

In the memory-less model, the transition occurs around $\gamma^* = 0.69$. Before the transition, agents spend about $p = 60\%$ of the time in state $+$, consistent with the bias of the external signal [190]. After the transition, individual simulations tend to converge to all agents in state $+$ or all in state $-$, depending on early fluctuations. As a result, the ensemble average settles near 0.5, not due to ongoing oscillations, but because half the simulations end up in each absorbing state.

The memory-based dynamics (bottom panel) show a similar pattern. For small γ, memory amplifies the external signal and most agents are locked into the favoured state $+$, and both temporal and ensemble averages approach 1. As γ increases, the system becomes more sensitive to early randomness and simulation outcomes diverge: some go to $+$, some to $-$, resulting in a wide spread of temporal averages. This spread reflects path-dependent dynamics, where early cumulative exposure shapes long-term decisions. The ensemble average follows a smoother decline, diverging clearly from most individual

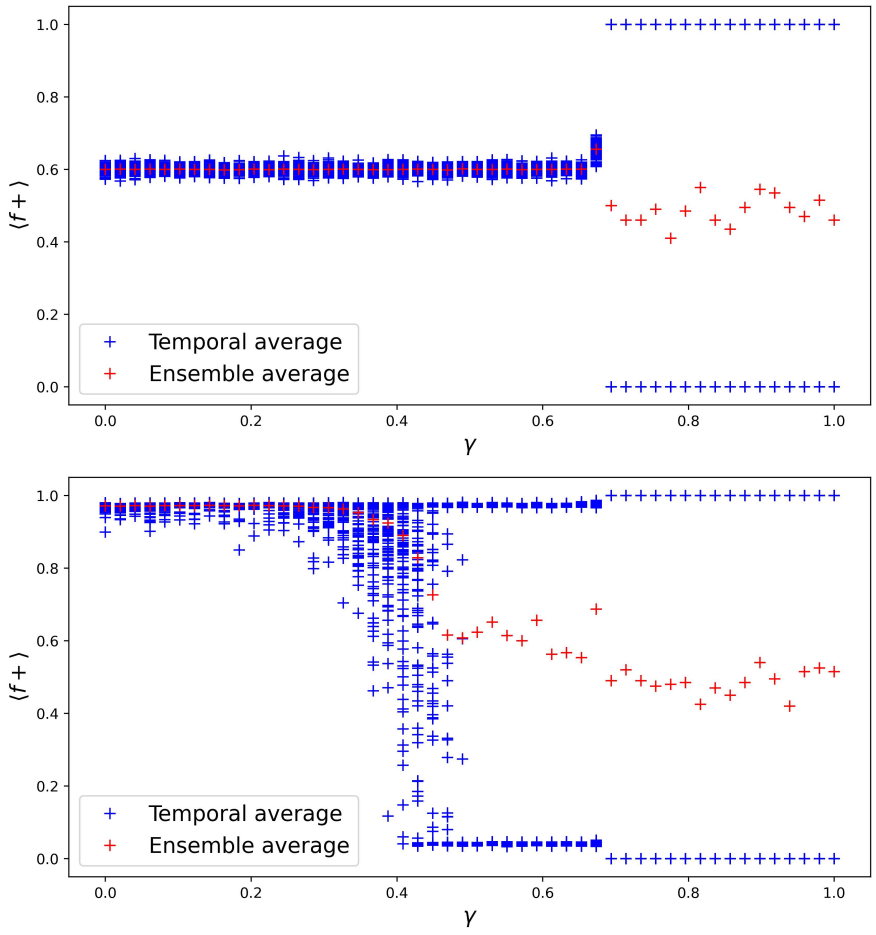

Figure 2.1: Comparison of temporal averages (blue markers) and ensemble averages (red markers) of the fraction of agents in state + (denoted $\langle f_+ \rangle$), as a function of the social influence parameter γ in the high-sensitivity regime, with $\beta = 100$. Top panel: memoryless dynamics, $\gamma^* = 0.69$. Bottom panel: memory-based dynamics ($m = 200$), $\gamma^* = 0.37$. We consider $T = 3000$ time steps and a sample size of $S = 200$; the other parameters are listed in Remark 2.2.2.

runs. In this regime, early fluctuations in the signal are averaged over the memory window and increasingly shape the agent's perception of the environment. Once a dominant trend emerges, memory reinforces it, locking the system into a consistent collective state. The coexistence of many stable outcomes (as shown by the wide spread of blue crosses) again indicates ergodicity breaking, but this time due to reinforced memory-induced path dependence, not just bistability. The transition happens at lower threshold $\gamma^* = 0.37$, as memory smooths the influence of the external signal, making agents less reactive to short-term noise. Notably, under this dynamic, it is harder to identify γ^*. To overcome this issue, we consistently define it as the first value at which a departure from ergodic behaviour appears, which is later confirmed through stability analysis in Sec. 2.2.4.

Phase-Space Exploration and Ergodicity Breaking

At this point, it is useful to clarify how ergodicity breaking can be interpreted in terms of phase-space exploration. At the microscopic level, the phase space consists of all possible configurations of the N agents, each of which can occupy one of two states, yielding 2^N distinct microstates. However, the analysis focuses on macroscopic observables. At the societal level, the relevant phase space is spanned by the fraction of agents in a given state, $f_+ \in [0, 1]$.

In the oscillatory regime, a single long trajectory explores a broad region of this macroscopic phase space, repeatedly visiting mixed configurations and yielding time averages that coincide with ensemble averages. By contrast, in the regime where ergodicity breaking occurs, the dynamics becomes effectively confined to one of two basins of attraction corresponding to near-consensus states ($f_+ \simeq 0$ or $f_+ \simeq 1$). Once the system enters one of these basins, transitions either to the complementary basin or back to the interior mixed region become extremely unlikely on accessible time scales.

As a consequence, a single realisation of the stochastic process explores only a subset of the macroscopic phase space, which is selected by early stochastic fluctuations and subsequently reinforced by memory effects. While both basins remain accessible at the ensemble level, their effective inaccessibility within individual trajectories gives rise to path dependence and to the breakdown of ergodicity in the practical sense relevant for the model. In this way, ergodicity breaking can be interpreted as the emergence of dynamically disconnected regions in the effective phase space of macroscopic observables.

This phase-space interpretation also clarifies the meaning of the wide spread of blue crosses observed in Fig. 2.1. Each blue cross represents the tem-

poral average obtained from a single realisation of the stochastic dynamics, corresponding to the mean fraction of agents in state + over the duration of one simulation. In an ergodic regime, all such averages would converge (up to finite-size fluctuations) to the same value. The observed dispersion instead indicates that different trajectories converge to distinct long-term macroscopic configurations.

Importantly, each of these configurations is dynamically stable on the time scales considered: once a trajectory approaches one of these values, it continues to fluctuate around it without departing significantly. This stability is directly reflected in the spread of blue crosses. If the macroscopic configurations were not dynamically stable, individual trajectories would repeatedly move between different regions of the phase space, and the corresponding temporal averages would converge to a common value.

To summarise, the divergence between temporal and ensemble averages beyond the critical threshold serves as clear evidence of ergodicity breaking, albeit through different underlying mechanisms: bistability in the memoryless case and memory-reinforced path dependence in the memory-based one.

It is worth noting that the mechanism of ergodicity breaking observed in the present model exhibits clear phenomenological similarities with the minority game discussed in [30], a stylised agent-based model in which a population of agents repeatedly chooses between two alternatives and those belonging to the minority side are rewarded. Despite this shared phenomenology, the origin of ergodicity breaking in the two settings is markedly different. In [30], the loss of ergodicity arises because the dynamics unfolds in a state space where many distinct long-lived configurations coexist, between which the system can transition only with difficulty. As the dynamics proceeds, transitions between these configurations become progressively rarer, since increasingly large collective fluctuations are required for the system to switch from one configuration to another. As a result, the characteristic time scales associated with these transitions grow rapidly and can become much longer than the duration of numerical simulations or empirical observations. This leads to ageing effects, meaning that the dynamical properties of the system depend on the time elapsed since its initialisation, and to an increasingly incomplete exploration of the available macroscopic states.

In contrast, ergodicity breaking in the model presented here originates from a reinforcement mechanism driven by social influence and memory. Early stochastic fluctuations bias the collective dynamics towards specific macroscopic configurations and, once such configurations are reached, the interplay between individual behaviour, collective outcomes, and memory effects reinforces them and makes transitions away from them increasingly

unlikely. In this sense, the system's past values play a central role: different realisations may settle into different long-lived configurations depending on early events, even though all configurations remain dynamically accessible in principle.

Thus, while both frameworks display an effective fragmentation of the macroscopic state space into regions that are rarely explored within a single realisation, the mechanisms responsible for this fragmentation are fundamentally different: in [30], the separation arises from progressively slower dynamics and long-lived trapping induced by the structure of the state space itself, whereas in the present model it emerges from endogenous feedback and reinforcement.

Low-Sensitivity Regime

For what concerns the low-sensitivity regime (i.e., high randomness), we explore two conditions: $\beta = 1$ and $\beta = 0$.

As shown in the top panel of Fig. 2.2, for $\beta = 1$ the averages show a smooth and monotonic decrease with γ. When social coupling is weak (small γ), agents remain moderately influenced by the external signal and the average behaviour reflects the signal's bias, yielding

$$\langle f_+ \rangle \sim p,$$

with $p = 0.6$. Due to the inherent randomness at this level of β, significant dispersion in temporal averages is observed across independent simulations. These fluctuations are expected in a regime where decisions are only weakly guided by payoff differences and agents randomly act (with full stochasticity reached in the degenerate scenario $\beta = 0$ [87, 180]).

As γ increases, social influence begins to outweigh the role of the external input and the population averages converge to

$$\langle f_+ \rangle \sim 0.5.$$

At the same time, the variation across simulation outcomes becomes smaller, indicating more coordinated and homogeneous collective dynamics.

No evidence of ergodicity breaking emerges: temporal and ensemble averages remain close. This behaviour is consistent with an ergodic regime where agents continuously adjust to both the external signal and the influence of peers. It is worth noting that, in this low-sensitivity regime, introducing memory into the decision rule does not significantly alter the results: the averaging over past inputs merely smooths short-term variations but does not meaningfully change the macroscopic patterns.

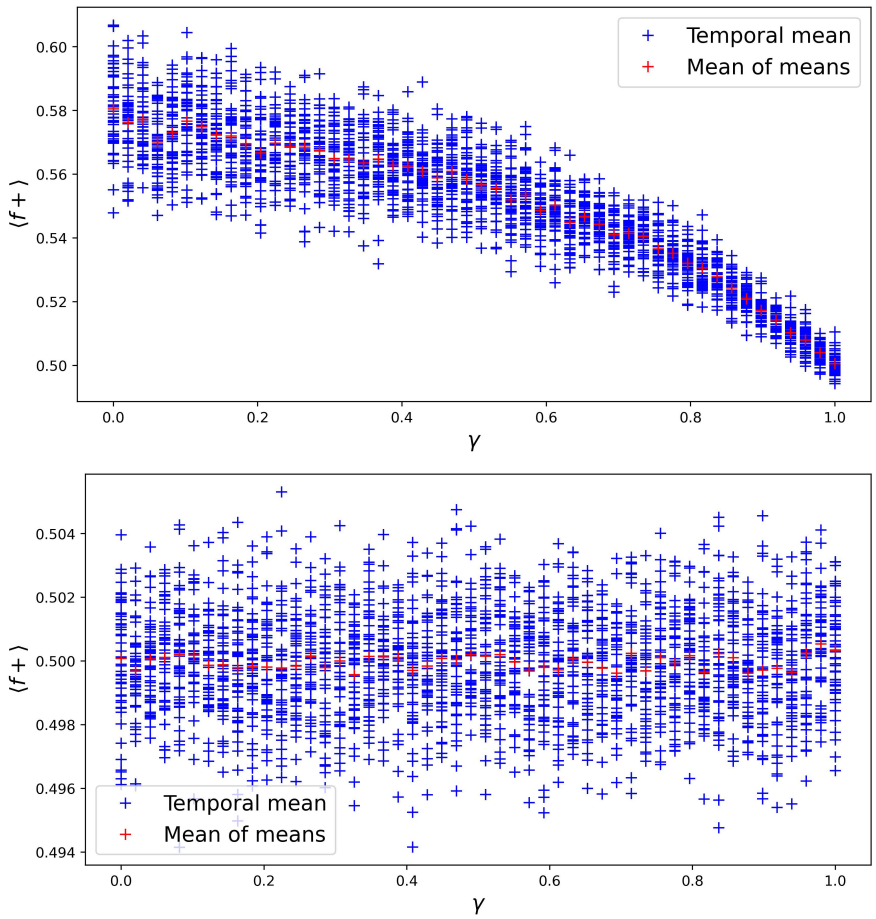

Figure 2.2: Comparison of temporal averages (blue markers) and ensemble averages (red markers) of the fraction of agents in state + (denoted $\langle f_+ \rangle$), as a function of the social influence parameter γ in the low-sensitivity and memory-less regime. Top panel: $\beta = 1$. Bottom panel: $\beta = 0$. The number of agents is $N = 100$ and sample size is $S = 50$; the other parameters are listed in Remark 2.2.2.

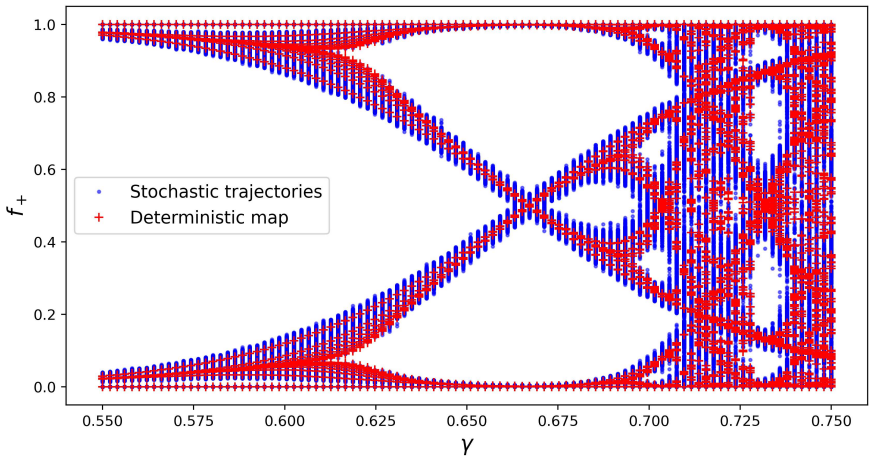

Figure 2.3: Comparison of the stochastic dynamics (blue markers) obtained from the agent-based simulations with their deterministic counterpart (red markers) derived from noise-free iterations of the model. Parameters are: $\beta = 10$, $p = 0.6$, $N = 1000$, $T = 1000$. Sample size is $S = 15$ for stochastic dynamics and 15 iterations are considered for the binomial tree.

In the degenerate case $\beta = 0$ (shown in Fig. 2.2, bottom panel), agents choose completely at random, with no response to either social or external inputs. This limit corresponds to a state of complete disorder, leading to

$$\langle f_+ \rangle = 0.5,$$

due to symmetry and randomness.

Trajectories

We continue our analysis by closely examining the state trajectories that agents follow over time in the high-sensitivity, memory-less regime, as shown in Fig. 2.3. To study these trajectories, we consider the points visited by agents under two types of dynamics:

a) The stochastic implementation, where agents probabilistically choose their state at each time step, as described in the agent-based model in Sec. 2.2.2;

b) A deterministic counterpart, in which randomness is entirely removed. Specifically, instead of relying on stochastic external signal or on probabilistic decision-making, we systematically simulate all possible state

transitions. We begin with the population entirely in state + or entirely in state −; then, at each time step, we explore both possible values for the external signal (+1 or −1) and agents update deterministically based on these inputs, without any randomness. This process creates a tree of possible outcomes, a branching structure that captures all potential state configurations the system can reach under idealised, noise-free conditions.

In the figure, the vertical spread of points exhibits the different states visited across simulations. The blue points represent the stochastic simulations and the red points show the corresponding deterministic outcomes. The two approaches show close structural agreement: both reveal that, for lower values of γ, both dynamics tend to converge toward extreme states (i.e., all agents choosing + or all choosing −), which is typical of the oscillating, ergodic phase. As γ increases, the branching structure becomes more complex, reflecting the growing influence of social reinforcement feedback.

Notably, this analysis only reflects the range of visited states and it does not convey how often each state is reached, how long the system remains in a particular state, or whether states are revisited, which are essential aspects for fully understanding the system's stability and whether it behaves ergodically.

2.2.4 Mean-Field Approximation

We continue to focus on the high-sensitivity regime, specifically on the mean-field approximation, which is a modelling approach that replaces the detailed interactions between individual agents with an average or "mean field" (Sec. 1.6.3), allowing the system's dynamics to be described in terms of aggregate quantities.

This approximation becomes appropriate when agents are both exposed to a smoothed version of the external signal (via memory) and exhibit highly deterministic responses (due to high sensitivity, i.e., large β) [191,192]. Memory alone reduces short-term fluctuations in the signal but does not suppress the stochasticity in agent decisions at low β (i.e., high randomness). Therefore, both memory and low noise (i.e., high β) are necessary to justify the assumption that the system can be described by its average behaviour. In fact, in many models mean-field approximation are often introduced in order to simplify complex interactions between individuals by replacing them with an average influence. In our scenario, when the memory window is sufficiently long, short-term fluctuations in the signal become negligible and, hence, the average of the input signal converges to its expected value over

time. Specifically, it holds that:

$$\langle X_m(t) \rangle \xrightarrow{m \to \infty} 2\,p - 1, \tag{2.8}$$

since $X(t)$ is a binary random variable taking values in $\{+1, -1\}$ with mean $2\,p - 1$. This simplifies the original model expression (Eq. (2.4)), which now becomes:

$$\Pi_{\pm}(t) = \gamma\, f_{\pm}(t) \pm (1 - \gamma)\,(2\,p - 1), \tag{2.9}$$

resulting in a deterministic approximation of the payoff in the long-memory limit.

We now proceed to further characterise this high-sensitivity regime by exploring the typical phase transition it undergoes. This is done through a combination of stability analysis, analytical insights, and an extension to networked populations to assess the robustness of our findings across different interaction structures.

Phase Transition Detection through Stability Analysis

The transition from the oscillatory phase to the absorbing one can be further investigated through a stability analysis within the mean-field approximation. Considering the payoff given by Eq. (2.9) and the fraction of agents in state $+$ at time t, denoted $f_+(t)$, we can derive a recursive map describing the evolution of the system:

$$f_+(t+1) = \frac{1}{1 + \exp\left(-\beta\,\left[2\,\gamma\, f_+(t) + 2\,(1-\gamma)\,(2\,p-1) - \gamma\right]\right)}. \tag{2.10}$$

Fixed points of this map are determined by the condition

$$f_+(t+1) = f_+(t), \tag{2.11}$$

which corresponds to equilibrium population states that remain unchanged under the dynamics. These fixed points are computed numerically and assessed for stability by examining the slope of the update function in their vicinity [96]. In particular, a fixed point is stable if small perturbations decay over time, i.e., if the absolute value of the derivative satisfies

$$|f'_+(t+1)| < 1. \tag{2.12}$$

As shown in the phase diagram of Fig. 2.4, we observe a qualitative change in the structure of the fixed points by systematically varying the parameter

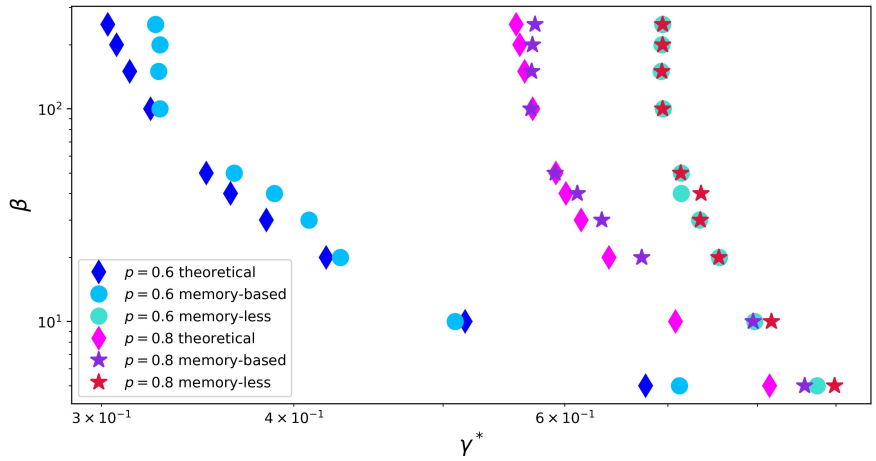

Figure 2.4: Critical social influence threshold γ^* as a function of sensitivity parameter β, for both memory-based and memory-less dynamics, compared with theoretical predictions (diamond markers) derived under mean-field approximation. The transition point γ^* marks the shift from oscillatory behaviour to absorbing dynamics. Results are shown for two signal biases, $p = 0.6$ and $p = 0.8$. We consider $N = 100$ agents and a sample size of $S = 50$. Other parameters are given in Remark 2.2.2.

γ. Below a critical threshold γ^*, the system admits a single, globally attractive fixed point, whereas above γ^*, multiple steady states exist, including bistable configurations separated by an unstable fixed point. This bifurcation marks a sharp transition from a dynamic, oscillatory regime to one in which the population becomes locked in one of two absorbing states. This behaviour is typical in such scenarios and confirms the findings presented earlier [96, 168, 193–195].

When compared with simulation results, the predictions from the mean-field approach (diamond markers) are in close agreement with results obtained under memory-based dynamics, discussed in Sec. 2.2.3. In fact, memory plays a critical role in suppressing stochastic fluctuations, thereby promoting convergence to deterministic behaviour. As expected, the agreement is less precise in the memory-less case.

It should be noted that rounding errors may slightly influence the estimation of critical values. These minor numerical discrepancies originating from the finite precision of floating-point representation may become significant near bifurcation points, particularly when evaluating derivatives close to the stability threshold $|f'_+(t+1)| = 1$. To mitigate such issues, we ensure that convergence criteria are sufficiently stringent and numerical derivatives are computed with care using small perturbations. However, while such round-

ing errors do not alter the qualitative conclusions of the analysis, they may explain minor deviations between theoretical predictions and simulation outcomes.

In summary, although some discrepancies arise due to finite-size effects and rounding errors, the mean-field framework remains effective in capturing the essential features of the transition and provides a powerful tool for the analysis of collective stability.

Analytical and Numerical Estimation of Flipping Probability

In the high-sensitivity and memory-based regime, sufficiently long simulations (e.g., $T > 100$) and moderate-to-large memory lengths (typically $m > 20$) reveal a sudden and irreversible transition from state $-$ to state $+$. Once this transition occurs, the agent generally remains permanently in state $+$. This phenomenon is illustrated in Fig. 2.5, which tracks the trajectory of a single agent over time.

To understand the conditions that enable such transitions, we first identify the critical threshold that must be exceeded for an agents to switch from state $-$ to state $+$. By setting $f_+(t) = 0$ in Eq. (2.4) and equating the two payoffs, we obtain the critical condition:

$$(1-\gamma)\langle X_m\rangle = \gamma - (1-\gamma)\langle X_m\rangle. \tag{2.13}$$

This leads to a threshold condition for the average signal:

$$X^* = \frac{\gamma}{2(1-\gamma)}. \tag{2.14}$$

Here, X^* denotes the minimum time-averaged signal required to trigger a state transition. Notably, this threshold becomes more stringent as γ increases and, beyond a certain point, the condition can no longer be satisfied.

In order to characterise these state transitions, we aim to compute the probability

$$P\left(\frac{1}{m}\sum_{i=1}^{m} X(i) > X^*\right), \tag{2.15}$$

both computationally and analytically. We begin with the theoretical derivation.

Let $S_m = \sum_{i=1}^{m} X(i)$ be the total signal over the memory window. Since each $X(i) \in \{-1, +1\}$ with probability $P(X(i) = +1) = p$, this sum can be

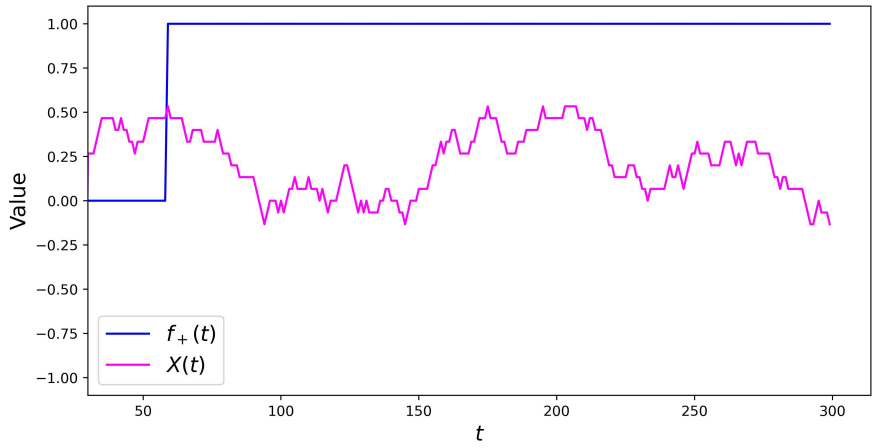

Figure 2.5: Illustration of a flipping event in the high-sensitivity, memory-based regime. The magenta line shows $\langle X_m \rangle$, with $m = 30$. The blue line represents the agent's state, which flips from state $-$ to state $+$ when $\langle X_m \rangle$ exceeds the critical threshold X^*. Parameter values are: $\beta = 100$, $p = 0.6$, $N = 1$, $T = 300$, $m = 30$, $\gamma = 0.5$.

written in terms of a binomial variable. Specifically, if $Y \sim \text{Bin}(m, p)$ counts the number of times $X(i) = +1$, we obtain:

$$S_m = 2\,Y - m. \tag{2.16}$$

Substituting this into the probability expression in Eq. (2.15) yields:

$$P\left(\frac{S_m}{m} > X^*\right) = P\left(Y > \frac{m\,(1 + X^*)}{2}\right). \tag{2.17}$$

Defining $k^* = \frac{m\,(1+X^*)}{2}$, we apply the normal approximation[1]:

$$Y \approx \mathcal{N}(m\,p, m\,p\,(1-p)),$$

and standardise:

$$P\,(Y > k^*) \approx P\left(Z > \frac{k^* - m\,p}{\sqrt{m\,p\,(1-p)}}\right), \tag{2.18}$$

where $Z \sim \mathcal{N}(0, 1)$. Substituting for k^*, we obtain:

$$\frac{k^* - m\,p}{\sqrt{m\,p\,(1-p)}} = \frac{m\,(1 + X^* - 2\,p)}{2\sqrt{m\,p\,(1-p)}} = \sqrt{m}\,\frac{1 + X^* - 2\,p}{2\sqrt{p\,(1-p)}}. \tag{2.19}$$

[1] $\mathcal{N}(\mu, \sigma^2)$ denotes a normal (Gaussian) distribution with mean μ and variance σ^2.

Thus, the flipping probability is approximated by:

$$P\left(\frac{1}{m}\sum_{i=1}^{m} X(i) > X^*\right) \approx \frac{1}{2}\left[1 - \operatorname{erf}\left(\frac{\sqrt{m}\,(1 + X^* - 2\,p)}{2\,\sqrt{2\,p\,(1-p)}}\right)\right], \tag{2.20}$$

where erf is the standard error function:

$$\operatorname{erf}(x) = \frac{2}{\sqrt{\pi}} \int_0^x e^{-t^2}\, dt. \tag{2.21}$$

This expression provides a compact analytical estimate of the flipping probability derived from the Gaussian approximation of the binomial distribution. Notably, the critical threshold X^* in Eq. (2.14) is independent of the signal bias p, whereas the probability itself in Eq. (2.20) is not.

In the limit of large memory ($m \to \infty$), the Law of Large Numbers ensures that the empirical average of the signal converges almost surely to its expected value:

$$\frac{1}{m}\sum_{i=1}^{m} X(i) \xrightarrow{\text{a.s.}} \mathbb{E}[X(i)] = 2p - 1.$$

As a consequence, fluctuations in the signal average become negligible and the system behaves as each agent perceives a constant input equal to its expectation. In this deterministic limit, the flipping probability becomes:

$$\lim_{m\to\infty} P\left(\frac{1}{m}\sum_{i=1}^{m} X(i) > X^*\right) = \begin{cases} 1, & \text{if } X^* < 2p - 1, \\ \frac{1}{2}, & \text{if } X^* = 2p - 1, \\ 0, & \text{if } X^* > 2p - 1. \end{cases}$$

This result implies that, in the limit of large memory ($m \to \infty$), the occurrence of flipping becomes entirely determined by the comparison between the threshold X^* and the expected value of the signal, $\mathbb{E}[X] = 2p - 1$. In other words, the system shifts from a stochastic regime to one governed by deterministic thresholds, which is a defining feature of mean-field behaviour. The empirical dynamics therefore reduce to a simplified effective description, where agent decisions depend on average quantities rather than on individual-level noise.

In simulations, we estimate the flipping probability by counting the frequency with which the condition $\langle X_m \rangle > X^*$ is satisfied in the memory-based setting. For each simulation run, we assess whether the agent's memory-averaged signal exceeds the threshold defined in Eq. (2.14) and record transitions accordingly. Repeating this procedure over an ensemble of S simulations yields empirical estimates of the transition probability.

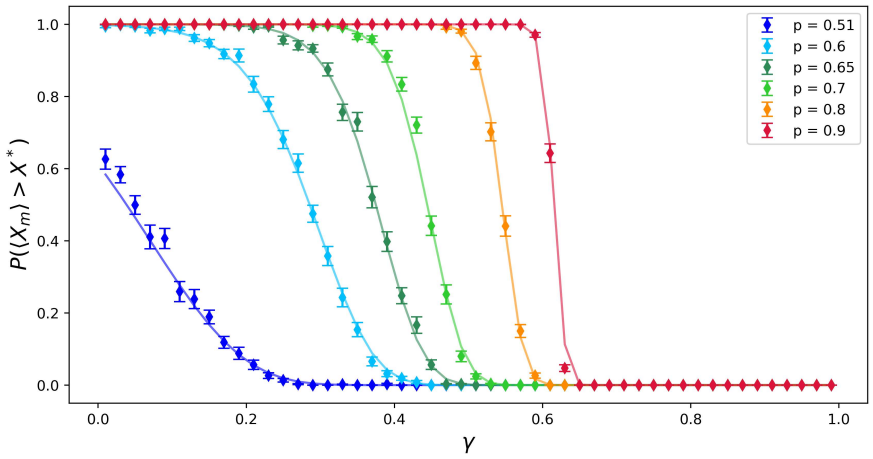

Figure 2.6: Comparison between analytical predictions (lines) and computational estimates (markers with error bars) of the probability that the memory-averaged signal $\langle X_m \rangle$ exceeds the critical threshold X^*, as a function of the social influence parameter γ. Each curve corresponds to a different value of the external bias p, ranging from 0.51 to 0.9. Other parameters are: $\beta = 100$, $N = 100$, $T = 1000$, $m = 200, S = 50$.

As shown in Fig. 2.6, there is strong agreement between analytical predictions and simulation results for the probability $P(\langle X_m \rangle > X^*)$. Several curves are considered, each for a different value of p, with theoretical predictions represented as continuous lines and simulation results as markers with associated error bars. The close agreement between the two across all regimes confirms the accuracy of the probabilistic approximation, validating the analytical treatment developed before.

The flipping probability decreases sharply with increasing γ, reflecting a transition from signal-driven to socially driven decision-making. This transition point depends on p: stronger biases (larger p) allow the signal to overcome social pressure for a broader range of γ. Eventually, all curves approach zero for sufficiently high γ, indicating that agents become locked into their initial state.

This analysis provides a quantitative characterisation of the conditions under which individual agents are able to respond to external information in the presence of social pressure. Moreover, it demonstrates that the stochastic dynamics of flipping in the memory-based regime can be effectively captured by a deterministic analytical model, which successfully predicts transition probabilities based on the relative magnitude of signal bias and social influence.

Network Implementation

The insights obtained from the mean-field analysis are not limited to fully connected populations, where each agent interacts uniformly with all others. On the contrary, we find that similar qualitative behaviours persist even in more complex and realistic settings, where agents are embedded within networks characterised by structured patterns of connectivity [28, 196]. Specifically, we examine several types of network topologies (Sec. 1.8), each differing in interactions among agents, including:

- Erdős–Rényi (ER) networks, where connections between agents are established independently with a fixed probability [147]. This setup mimics a well-mixed environment, making it a good reference point for comparison [197];

- Stochastic Block Models (SBM), which represent networks with community structures [149, 198];

- Barabási–Albert (BA) scale-free networks, with heterogeneous degree distributions, which mimic real-world systems by assigning more connections to highly connected nodes (hubs) [70, 148].

Despite the heterogeneity in node degrees, a consistent pattern emerges across all topologies: small values of the social influence parameter γ yield oscillatory dynamics, while, beyond a critical threshold γ^*, the system becomes trapped in absorbing states. Importantly, the critical thresholds are similar to those in the mean-field setting: $\gamma^* \simeq 0.71$ in the memory-less case and $\gamma^* \simeq 0.33$ in the memory-based one. Fig. 2.7 shows representative results for the ER network, which are qualitatively mirrored in SBM and BA networks.

This suggests that the phase transition is robust to changes in network topology, echoing prior results on voter models in complex networks [186]. In addition, by explicitly comparing regimes with and without memory, we provide a deeper understanding of how ergodicity breaks down in structured populations.

With the agent-based model established, we next address its empirical validation, focusing on abrupt shifts in system dynamics as a key point of comparison between theory and observation.

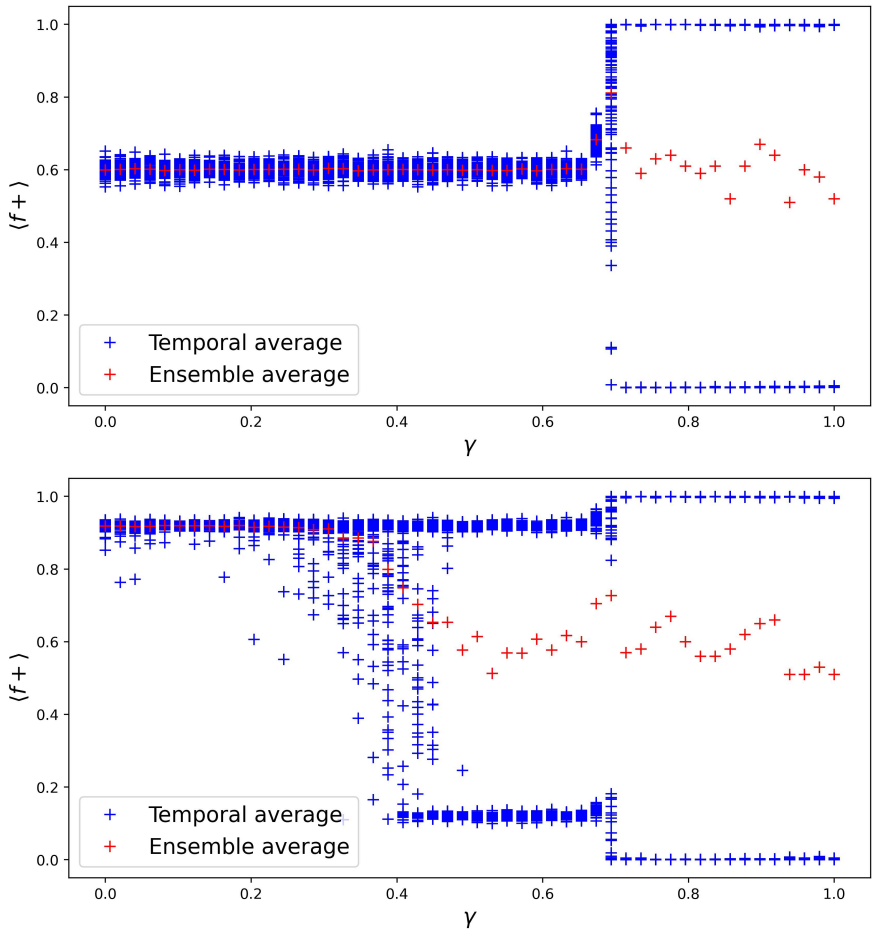

Figure 2.7: Temporal and ensemble averages for an Erdős–Rényi network. The ensemble size is $S = 100$, the average node degree is $\langle k \rangle = 10$, and the sensitivity parameter is $\beta = 100$. Other parameters are listed in Remark 2.2.2. Top panel: Memory-less scenario, with $\gamma^* = 0.71$. Bottom panel: Memory-based scenario, with $\gamma^* = 0.33$.

2.2.5 Model Validation

The stylised model presented in this study offers valuable potential for real-world applications, particularly in contexts where binary options are available and individuals are simultaneously exposed to both social influence and external information [185, 199–202]. Relevant examples range from voting patterns and collective mobilisation to consumer choices and public health behaviour. Translating the model into empirical settings requires estimating the critical social influence threshold, γ^*, which separates two fundamentally distinct behavioural regimes: an ergodic phase, in which individuals remain flexible and responsive to external cues, and a non-ergodic phase, in which collective behaviour stabilises into persistent consensus or polarised states [203, 204]. The conformity parameter γ is not merely a theoretical abstraction: it can be empirically estimated and calibrated to observed data, enabling the model to serve as a predictive tool. Once calibrated, identifying the critical threshold γ^* provides insight into societal rigidity, consensus lock-in, or the potential for sudden collective shifts

Several empirical strategies can support this estimation. One approach involves survey-based inference, adopted to assess the relative weight individuals assign to social conformity versus external information. A second strategy relies on controlled behavioural experiments, where participants make binary choices under varying levels of peer influence and external endorsements, allowing researchers to infer when system dynamics shift. A third strategy draws on large-scale datasets from digital interactions (such as social media platforms or online petitions) where individuals are continuously exposed to social validation mechanisms (e.g., likes, shares) and informational content (e.g., news, expert opinions). By observing when opinion cascades emerge or fail to emerge, researchers can infer whether the system is operating near a critical point. Previous studies have shown that online opinion dynamics are particularly susceptible to abrupt shifts driven by imitation, consistent with the transitions identified in our model [205].

This predictive capacity has clear implications for domains such as public health communication, electoral strategy, or policy design, where anticipating population-level responses to social and informational stimuli is of critical importance. The dynamics observed during the COVID-19 pandemic, for instance, provide a striking example of how rapidly social influence and information can interact to shift collective behaviour at scale. Another key example is the power of public endorsement, particularly evident during U.S. elections, where social influence can become so dominant that voters cease to rely on external information (i.e., data or expert sources) and instead base their decisions primarily on the opinions of others, ultimately leading to a

locked-in opinion state [202, 206]. Another example is the Brexit referendum, which highlighted how conformity pressures and polarised information environments can shape large-scale collective decisions [207]. Similarly, herding behaviour has long been documented among financial analysts producing earnings forecasts [208].

Concerning the feasibility of the outlined validation strategies, the validation of agent-based models cannot be approached through a single, standardised procedure, but must instead be adapted to the aims and empirical reference of the model under consideration. As discussed in [209], a fundamental distinction exists between the use of empirical data to specify and calibrate micro-level model components (such as individual decision rules and interaction mechanisms) and the use of empirical observations to validate macro-level simulation outcomes. Both forms of validation are often constrained by limited data availability, particularly in social systems where detailed information on individual behaviour, learning processes, and social interactions is rarely fully observable. Within this framework, case-based models (designed to reproduce specific, time- and space-bounded empirical phenomena) are generally more amenable to direct empirical validation, since relevant datasets can sometimes be collected or reconstructed. By contrast, stylised models and theoretical abstractions (which aim to isolate and analyse generic mechanisms rather than replicate particular empirical cases) face more substantial validation challenges. In this respect, the model presented here falls into the latter category, as it is intended to capture stylised features of collective dynamics and social influence rather than to provide a quantitative description of a specific empirical system.
These considerations are consistent with the broader methodological discussion in [210], where it is argued that empirical validation in agent-based modelling poses challenges that differ fundamentally from those encountered in traditional analytical or econometric approaches, due to the presence of heterogeneity, non-linearity, stochastic dynamics, feedback between micro and macro levels, and path dependence. In such settings, multiple models may reproduce similar aggregate patterns, and strong sensitivity to initial conditions can further complicate direct comparisons with empirical time series.
Against this background, empirical validation of agent-based models can be regarded as both desirable and feasible in principle, but its practical implementation must be carefully matched to the scope and purpose of the model, with theoretical abstraction and empirical realism viewed as complementary rather than competing objectives. Given the methodological effort and dedicated empirical work required to address these issues adequately,

the validation strategies outlined above are therefore left to future studies.

Taken together, these considerations underscore that the parameter γ is not only measurable, but also interpretable in substantive terms, with proper attention. Estimating and monitoring γ^* may thus offer valuable insight into the conditions under which societies transition between flexibility, consensus, and polarisation, with significant consequences for policy and governance. The examples presented further confirm the broad applicability of the model across domains, from political decision-making to market behaviour.

2.2.6 Overview of Results

Human behaviour is profoundly shaped by social influence. Individuals seek approval from peers and, when their opinions align with those of others, the resulting sense of social validation tends to reinforce existing beliefs. Conversely, exposure to opposing views can induce discomfort, weakening attachment to prior convictions. As a consequence, people may adopt collective behaviours (even when suboptimal) by conforming to prevailing norms or trends. This tendency can lead to the neglect of personal information or independent judgement, echoing well-documented phenomena such as informational cascades and herding behaviour [211, 212]. To characterise the trade-off between social conformity and external evidence (interpreted as an informational signal or ground truth), we develop a stylised agent-based model.

In our framework, agents occupy one of two discrete states, which can be interpreted as binary opinions or behaviours (e.g., voting preferences or lifestyle choices such as smoking or dropping out of school). Each agent receives a payoff associated with their decision and, at each time step, decides which state to adopt according to a Fermi rule, whereby transitions are more likely when the alternative state offers a higher payoff. The model captures realistic bounded rationality: while agents tend to favour strategies yielding higher payoffs, they may occasionally adopt less advantageous options due to random fluctuations.

Our main focus is on state occupancy, specifically the temporal and ensemble averages, to identify emerging phenomena in opinion dynamics. We examine how macroscopic outcomes depend on two core components: the sensitivity to payoff differences (controlled by parameter β) and the strength of social influence (governed by γ). These two parameters enter into the payoff definition and, consequently, in the dynamics.

Additionally, we investigate the presence or absence of memory in processing external signals.

When agents are highly sensitive to payoff differences (i.e., $\beta \gg 1$), they tend to adopt states aligned with higher expected rewards. We observe that outcomes in the memory-less and memory-based regime exhibit common features. Specifically, for low values of γ, agents predominantly follow the external signal and the system remains ergodic: time averages converge to ensemble averages and agents continuously switch between states, with $\langle f_+ \rangle \simeq p$, where p denotes the signal bias, correctly tracking the external information. Above a critical threshold γ^*, the system enters a non-ergodic regime, where it becomes trapped in one of the absorbing states. Agents cease to respond to the signal, breaking self-averaging properties. This transition is sharp, consistent with classical phase transition phenomena and ergodicity breaking.

Introducing memory modifies the dynamics: instead of reacting to instantaneous signals, agents respond to the average information over a time window of length m, which smooths out noise. The key difference from the memory-less dynamics lies in the lower critical value γ^* and the smoother nature of the transition. In addition, a different mechanism underlies the transition: in the memory-less case, ensemble averages reflect bistability across absorbing states. In contrast, the memory-based regime exhibits path dependence, where early fluctuations in the signal are reinforced over time, leading to divergent long-term outcomes.

The presence of memory enables the application of a mean-field approximation, where the time-averaged signal is approximated as a deterministic quantity drawn from a binomial distribution. Under the mean-field and high-sensitivity regime we confirm the critical transition analytically via stability analysis. To further characterise the dynamics, we compute the flipping probability of agent switching state, both analytically and numerically, with the critical threshold derived analytically. As γ increases, the likelihood of exceeding the threshold declines, reflecting the growing dominance of social influence. Theoretical predictions show excellent agreement with empirical measurements across various values of signal bias p, confirming the robustness of the analytical treatment. Notably, the presence of memory leads to irreversible flipping in the high-sensitivity regime, which means that, once the average signal exceeds the threshold, agents switch state and remain there. This phenomenon captures the way long-term exposure to biased signals can drive persistent behavioural change, a feature relevant to social phenomena such as political polarisation or cultural adoption [213].

Beyond fully mixed populations, we implement the model on various network topologies: Erdős–Rényi, Barabási–Albert and Stochastic Block Models. Despite differences in structural connectivity, the overall behaviour remains qualitatively consistent for both memory-less and memory-based dy-

namics: a critical γ^* separates ergodic and non-ergodic regimes, with similar value among the different network topologies. This suggests that the observed phase transition is a robust collective phenomenon, largely independent of the microscopic arrangement of interactions, in line with previous findings on noisy voter models [186].

In the low-sensitivity, high-randomness regime ($\beta \to 0$), no ergodicity breaking is observed, and state occupancy smoothly declines from alignment with the external signal to completely random behaviour. No dependence on memory is inferred.

Our model has potential applications across diverse empirical domains, ranging from electoral dynamics to social habits and collective action. It can help explain sudden shifts in collective behaviour that culminate in phenomena such as financial crises, viral trends, rapid mobilisation, or abrupt changes in public sentiment. Examples include social behaviours like school dropout, smoking, voting preferences, or adoption of birth control technologies [169,170], as well as community crime rates [214], and the stronger influence of peers compared to parents on students' academic performance [215]. The model also accounts for asset market volatility [99]. Electoral settings provide further illustration: opinion polling before major events (e.g., the Brexit referendum or U.S. presidential elections) often shows early volatility that eventually solidifies into entrenched positions, which is a process that can be interpreted as crossing the critical threshold γ^*. Similarly, in the adoption of social behaviours such as participation in climate strikes or vaccine uptake, reaching a critical mass of early adopters can lead to irreversible shifts, consistent with the model's transition from a fluctuating to a stable regime. The same mechanism applies to financial markets, where herding behaviour leads investors to follow others rather than rely on private information [211–213]. In this case, the ergodic phase reflects adaptive market behaviour, while the absorbing phase corresponds to persistent bubbles or crashes.

The model can be empirically validated by estimating the critical social influence threshold γ^* through surveys, behavioural experiments, and large-scale digital data. Identifying γ^* helps anticipate transitions between flexibility, consensus, and polarisation in real-world collective dynamics.

In summary, our model contributes to the growing literature on social imitation and collective decision-making. It captures how individual decisions emerge from the interplay between external evidence and peer influence, and how this interplay may drive abrupt, irreversible transitions. Importantly, it highlights the asymmetry of influence: in highly conformist populations, targeted interventions may be more effective than global information cam-

paigns. Future work could incorporate heterogeneous agents, dynamic conformity, or endogenous signal generation, further bridging the gap between stylised models and real-world complexity.

2.3 Imitation vs. Randomness in Society

The discussion now shifts to the original model developed in [2], which investigates the interplay between imitation and randomness in shaping societal dynamics. Specifically, it examines how individuals respond to trends (defined as the most popular actions at a given time, as determined by a ranking) and how such forms of social influence impact collective outcomes. Before delving into the details of the model, it is appropriate to provide an overview of the advantages and disadvantages of rankings.

2.3.1 The Role and Limitations of Rankings

Rankings serve as a tool for reducing complexity by ordering entities according to a common performance metric [216]. Over recent decades, nearly all aspects of society have become subject to rankings: people, countries, universities, businesses. This practice spans social, economic, and institutional domains. Since the 1970s, rankings have embodied the drive to quantify performance through global performance indicators, defined as "public, comparative and cross-national indicators that governmental, intergovernmental and/or private actors use regularly to attract attention to the relative performance of countries in a given policy area" [217]. In this way, rankings become instruments of governance as well as tools of evaluation.

When a particular metric is adopted to construct rankings, it gives rise to hierarchical structures [218, 219]. Once entities are ranked, the phenomenon of preferential attachment (Secs. 1.5.2 and 1.5.3) explains how power-law distributions emerge, based on the fact that resources attract further resources. For example, researchers who are highly cited are more likely to be cited again, and those who receive recognition (through funding, grants, or prizes) are more likely to continue receiving it. Co-authorship with already prominent scholars also plays a key role in reinforcing this dynamic [60, 220, 221]. This phenomenon is not limited to a specific area but applies across domains: from citations of scientific papers [59,222], to online social networks [223] and to wealth accumulation [63].

While rankings offer some benefits, such as improved resource allocation or increased transparency, they also introduce significant limitations. Chief among these is the reliance on predefined criteria, which may not capture

all dimensions of quality or value. As a result, there can be a disjunction between truly virtuous entities and those that merely align well with ranking methodologies, leading to significant discrepancies between actual talent and ranking position [171, 224].

Talent is undoubtedly a prerequisite for success, but it is often insufficient on its own. Factors such as randomness and cumulative advantage can play a decisive role. For instance, considering cultural trend, songs that become trendy (regardless of quality) achieve disproportionate success [225].

Furthermore, individuals and institutions tend to adjust their behaviour in response to being measured. This phenomenon is encapsulated by Goodhart's Law [226, 227]: "When a measure becomes a target, it ceases to be a good measure", i.e., any observed statistical regularity will tend to collapse once pressure is placed upon it for control purposes. In fact, when performance indicators are introduced, actors may align themselves more closely with the criteria being monitored. This phenomenon leads to a degree of standardisation and a loss of uniqueness [228]. In academic contexts, bibliometric criteria dominate university rankings, yet these metrics are limited, controversial, and prone to gaming (e.g., via self-citation) [229].

A further issue concerns the fact that rankings often serve promotional, not merely evaluative, purposes. For instance, university rankings function not only as comparative assessments but also as powerful tools for reputation-building, fund allocation, and policy formation [230]. The "rich-get-richer" mechanism (Sec. 1.5.3) operates here as well: the visibility conferred by high rankings further reinforces institutional positions, regardless of underlying quality. In this context, one concern is the potential for conflicts of interest in the compilation of such rankings, which can undermine their objectivity and make them poor indicators of actual talent or merit [231].

The pervasiveness (and the potential pitfalls) of ranking culture is also manifested by the fact that rankings themselves have become subject to evaluation: meta-rankings now exist to assess the robustness, fairness, and reliability of various ranking systems [230].

A further issue concerning rankings arises from the mistaken belief that replicating the strategies of top-ranked individuals or institutions will lead to similar success. Imitation is a deeply ingrained human behaviour and is central to social learning and cultural transmission [232]. However, empirical studies suggest that this influence and copy-mechanism is often misleading [233]. In fact, high-ranking entities may have benefited from unique, fortunate circumstances rather than reproducible strategies or they may also have adopted riskier approaches that paid off by randomness. Attempting to imitate these trajectories may not yield the same results and may in fact further reinforce existing inequalities. Thus, in general, it is better to develop

an own path to success rather than following someone else's steps [172, 234].

In summary, being ranked among the top brings prestige and, through self-reinforcing dynamics, increases the likelihood of remaining there. However, substantial evidence suggests that top-ranked individuals or institutions are not necessarily the most talented or deserving [171]. Social influence plays a critical role in shaping not only rankings themselves but also their perceived legitimacy.

Within this framework, our work aims to both confirm previous findings and provide new insights into the interplay between imitation and randomness when actions are ranked according to their popularity. Particular attention is given to identifying the conditions under which the resulting society is meritocratic, as assessed through appropriate quantitative measures. Notably, simulations prove to be extremely valuable in this context, as real-life data usually consist of a single time series with no possibility of repetition. Through simulations, artificial time series can be generated and multiple realisations explored, allowing us to examine a range of settings and scenarios [225].

2.3.2 Agent-Based Model Implementation

Given the framework outlining the interplay between randomness [171] and imitation [172], we implement an agent-based model (Sec. 1.7.2) to further investigate these phenomena, focusing in particular on the influence of trends, understood as the most popular actions.

We consider a probabilistic setting with an artificial society in which agents can play a finite set of actions. Specifically, at each time step of the discretised running time (not referencing physical time), each agent can either stick with the action currently played or switch to another one. Each action is selected with certain probabilities linked to fundamental quantities in our artificial setting.

More precisely, we implement the following setup in simulation, introducing an explicit analogy with the academic context in order to clarify the role of each item. Naturally, the model lends itself to broader interpretations, as explained in the following. Certain modifications can be introduced in alternative versions of the model, as outlined in Sec. 2.3.5; for now, we focus on the baseline scenario.

- There are N agents, labelled i. They can be thought of as researchers.
- Each agent i can play M actions, labelled j. For example, these may represent fields or topics in which the agents can conduct research and,

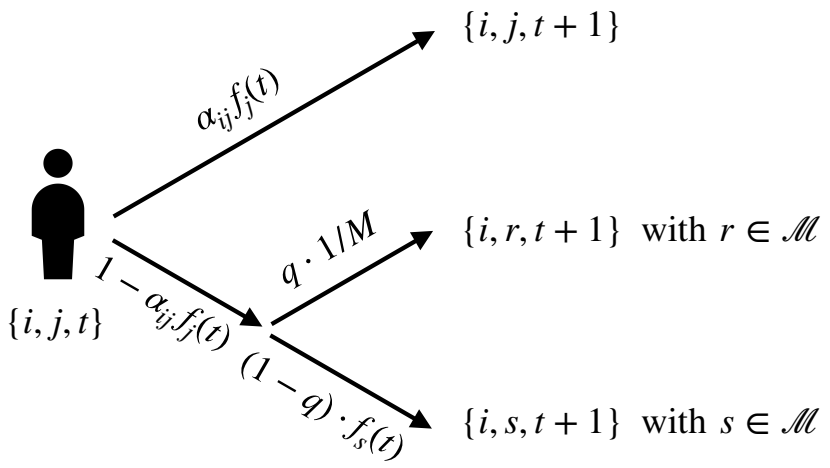

Figure 2.8: Stylised representation of the dynamics in our agent-based model in the baseline scenario. A generic agent i is in state j at time t and all the possible outcomes of their decision for the next time step are reported on the right. Choice probabilities are reported on the arrows. Parameters are as follows: α_{ij} account for the talent, $f_j(t)$ for the field popularity, q interpolates between imitation and randomness, M is the number of available fields.

consequently, publish. The action chosen by agent i at time t is denoted by $j(i,t)$; however, for simplicity, we often drop the dependence on i and t, and simply write j.

- A societal impact π_j is attributed to each action, with $\pi_j \in [0,1]$ for $j = 1, \ldots, M$, which quantifies its direct benefit to the society. Higher values denote a strong impact on society (e.g., cancer research) while lower values reflect less critical contributions. Each π_j is assigned at the beginning of the simulation and remains fixed.

- Agents can vary in talent, which we quantify using a random matrix whose elements $\alpha_{ij} \in [0,1]$ denote how skilled agent i is in field j.

- We define the attraction of each action j, i.e., its popularity, as the fraction of agents currently playing that action, denoted $f_j(t)$. Naturally, $f_j(t) \in [0,1]$. This formulation allows us to capture the influence of peer behaviour on individual decision-making.

- An import parameter is $q \in [0,1]$, which determines the role of randomness in our setting. Specifically, we observe imitation when $q \to 0$ and randomness when $q \to 1$. In our setting, q can be interpreted as reflecting the importance of the ranking of actions: high values imply

agents do not consider action rankings (i.e., trends), while low values imply agents imitate others and care about them. Another interpretation of q is the agent's propensity to explore available actions, with higher values denoting greater exploratory behaviour.

- The payoff gained by each agent i playing action j at time t is defined as:

$$\Pi_i(t) = \alpha_{ij}\, f_j(t). \tag{2.22}$$

Hence the reward (whether monetary, reputational, or success-related) increases with the agent's talent and the popularity of the field, further reflecting crowd influence [169]. The payoff accumulated over the time period $t = 0, \ldots, T$ by a single agent is then:

$$\Pi_i(T) = \sum_{t=0}^{T-1} \alpha_{ij}\, f_j(t) \tag{2.23}$$

- We are interested in the overall societal payoff, generated by all agents over time, defined as:

$$\Pi(T) = \sum_{i=1}^{N} \sum_{t=0}^{T-1} \alpha_{ij}\, f_j(t). \tag{2.24}$$

- Another quantity of interest is the overall societal benefit:

$$B(T) = \sum_{i=1}^{N} \sum_{t=0}^{T-1} \alpha_{ij(i,t)}\, \pi_{j(i,t)}. \tag{2.25}$$

This metric allows us to assess whether agents tend to engage with actions that are both beneficial to society and aligned with their individual talents. Dependencies are made explicit for the sake of clarity.

- The dynamics proceeds as follows and is illustrated in Fig. 2.8. At time $t = 0$, agents are assigned to fields proportionally to societal impact, i.e., to π_j. At each time step $t = 0, \ldots, T-1$, agents continue with their current action with probability

$$p(i, j, t+1) = \alpha_{ij}\, f_j(t), \tag{2.26}$$

i.e., proportional to the payoff in Eq. (2.22). That is, agents are more likely to stay with rewarding actions. With complementary probability they consider switching and, in that case, with probability proportional to q they select a different action randomly. With the remaining probability, they choose a new field based on popularity $f_j(t)$, thus accounting for current rankings and trends.

Remark. Regarding parameter choice in the simulations, the following settings ensure robustness and consistency of results. Time is run up to $T = 500$, by which point a steady-state is reached. A sample size of $S = 1000$ is considered to form a sufficiently large ensemble. Various values of $q \in [0, 1]$ are tested to explore all behavioural regimes. We consider $N = 250$ agents and $M = 100$ actions, for which we obtain stable results. We verify that outcomes remains consistent under changes in the N/M ratio; if not, exceptions are noted. We choose $N > M$ because in most real-world cases there are more agents than fields to occupy. We include sample standard deviations for all reported metrics.

The baseline implementation is conceived as a memory-less process, i.e., as a Markov chain (Sec. 1.3.1). Both $\alpha_{i,j}$ and π_j are pseudorandom deviates (Sec. 1.7.1) drawn from a uniform distribution in the range [0,1]. In the baseline scenario, both quantities are extracted at the beginning of the simulation and remain fixed for the entire duration. In the memory-based setting, agent skill becomes time-dependent, allowing skill to be gained through repetition.

With the agent-based model framework now established, we turn our attention to the analysis of key quantities.

2.3.3 Success, Talent, and Social Inequality

The first observable we examine is the overall payoff, as defined in Eq. (2.24). This represents the total payoff (interpreted as success, recognition, monetary income) generated by the entire society over the time evolution. Results are presented in Fig. 2.9 and suggest that, when imitation is the dominant mechanism (i.e., for low values of q), the overall payoff reaches its maximum. Conversely, when randomness dominates (i.e., for high q), the total payoff attains its minimum. A phase transition (Sec. 1.6.2) separates the two regimes and occurs at a critical point denoted q^*. The width of the error bars indicates significant variability in this transition region, which is typical of critical phenomena.
This initial analysis raises several important questions:

1) Is the society equal, i.e., is the total payoff evenly distributed among the agents?

2) Is the society meritocratic, i.e., is there a correlation between an agent's talent and their achieved payoff?

3) How are agents distributed among the available actions?

4) Do agents engage in actions that are beneficial to society?

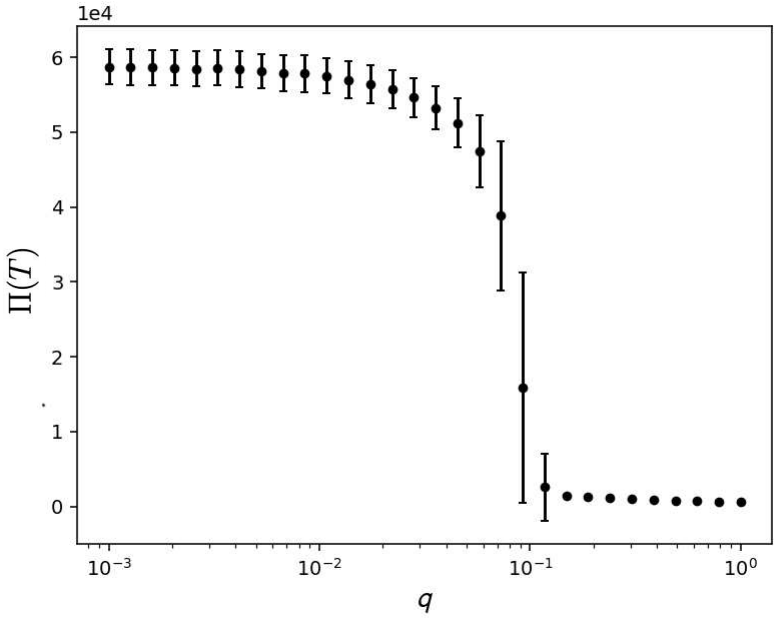

Figure 2.9: Overall payoff $\Pi(T)$ generated by the society along the time evolution for several values of parameter q, which accounts for imitation (low q) and randomness (high q) in the dynamics. Parameter choices are reported in Remark 2.3.2.

5) Are agents who engage in socially beneficial activities well-suited to them?

We proceed by examining each of these questions in detail.

Measuring Inequality with the Gini Coefficient

To address the question about equality, we analyse the Gini coefficient, a well-established measure of inequality introduced by Corrado Gini and Gaetano Pietra in the early 20th century [67, 68, 235]. The Gini coefficient quantifies the deviation of a distribution (such as income or wealth) from perfect equality among a group. It ranges from 0 (perfect equality) to 1 (maximum inequality).

In our model, it is defined by

$$G(T) = \frac{1}{N\,\Pi(T)} \sum_{i<k} |\Pi_i(T) - \Pi_k(T)|\,, \tag{2.27}$$

where the sum is computed over all unordered pairs of agents (i, k), by comparing their individual payoffs $\Pi_i(T)$ and $\Pi_k(T)$, defined in Eq. (2.23). The resulting total disparity, given by the sum of absolute differences, is then

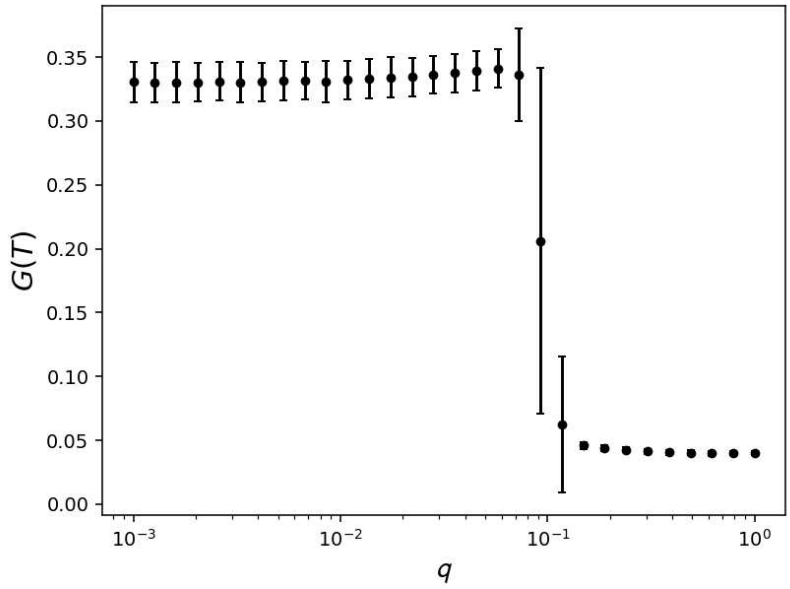

Figure 2.10: Gini coefficient $G(T)$ for the society, capturing the balance between imitation (low q) and randomness (high q) in the dynamics. Parameter choices are reported in Remark 2.3.2.

normalised to ensure that the Gini coefficient lies within the interval $[0, 1]$. A value of 0 indicates that all agents have received equal payoffs, while values approaching 1 suggest highly unequal outcomes.

Fig. 2.10 displays the results of our investigation. When imitation dominates the decision-making mechanism ($q \to 0$), the Gini coefficient is high, indicating that the total payoff is concentrated among a small number of agents and the society is highly unequal. In contrast, when randomness dominates ($q \to 1$), the payoff, despite being lower, is distributed more equally as reflected by a lower Gini coefficient. A phase transition is again observed between these regimes, reflecting the behaviour noted in the payoff curve.

Notably, the error bars shown in Figs. 2.9 and 2.10 represent the dispersion of the measured observables across independent realisations of the stochastic dynamics. Their increase at the transition reflects the fact that, near criticality, different runs of the model can settle into distinct long-term macroscopic states: small stochastic differences in early fluctuations or initial conditions can bias the subsequent evolution over long times, leading to strong run-to-run variability. From a physical perspective, away from criticality the dynamics is strongly attracted towards a well-defined macroscopic state, lead-

ing fluctuations to be quickly damped and ensemble averages to be sharply concentrated. Near criticality, by contrast, the system becomes weakly stable: relaxation times increase, correlations persist over long temporal scales, and stochastic perturbations are no longer efficiently suppressed but can instead propagate and influence the dynamics over extended periods. As a consequence, statistical averages converge more slowly and display larger uncertainties, which is reflected in the enhanced error bars observed close to the transition [236, 237].

It is also worth noting that the parameter q can be interpreted as the agent's propensity for exploration: when q is low, imitation dominates the decision-making process and agents primarily follow popular choices, behaving as imitators; conversely, when q is high, agents act as explorers, selecting actions independently of popularity and engaging in random exploration.

Meritocracy and Rank Correlation

The Gini analysis reveals that high total payoff in the imitation regime is not evenly distributed, whereas, when randomness dominates, the society is more equal as agents evenly contribute to it. This raises a further question: is this concentration of payoff meritocratic? Specifically, is it the most talented agents who are generating the greatest payoff?

To explore this, we compute the correlation between each agent's payoff and their average fitness, defined as

$$\langle \alpha_i \rangle = \frac{1}{M} \sum_{i=1}^{M} \alpha_{ij}. \tag{2.28}$$

i.e., the agent's average talent across all actions. To investigate meritocracy, we adopt Kendall rank correlation coefficient, denoted τ, which is a non-parametric measure of the strength and direction of association between two ranked variables. A value of $\tau = 1$ indicates perfect agreement (i.e., more talented agents receive higher payoffs), $\tau = 0$ indicates no correlation, and $\tau = -1$ indicates complete disagreement [238].

We analyse the correlation between payoff $\Pi_i(T)$ and $\langle \alpha_i \rangle$ for each agent:

$$\tau = \frac{C - D}{\frac{1}{2}N(N-1)}, \tag{2.29}$$

where C is the number of concordant pairs and D is the number of discordant pairs among all unordered pairs of agents (i, j). A pair is concordant if the ordering of $\Pi_i(T)$ and $\Pi_j(T)$ agrees with the ordering of $\langle \alpha_i \rangle$ and $\langle \alpha_j \rangle$,

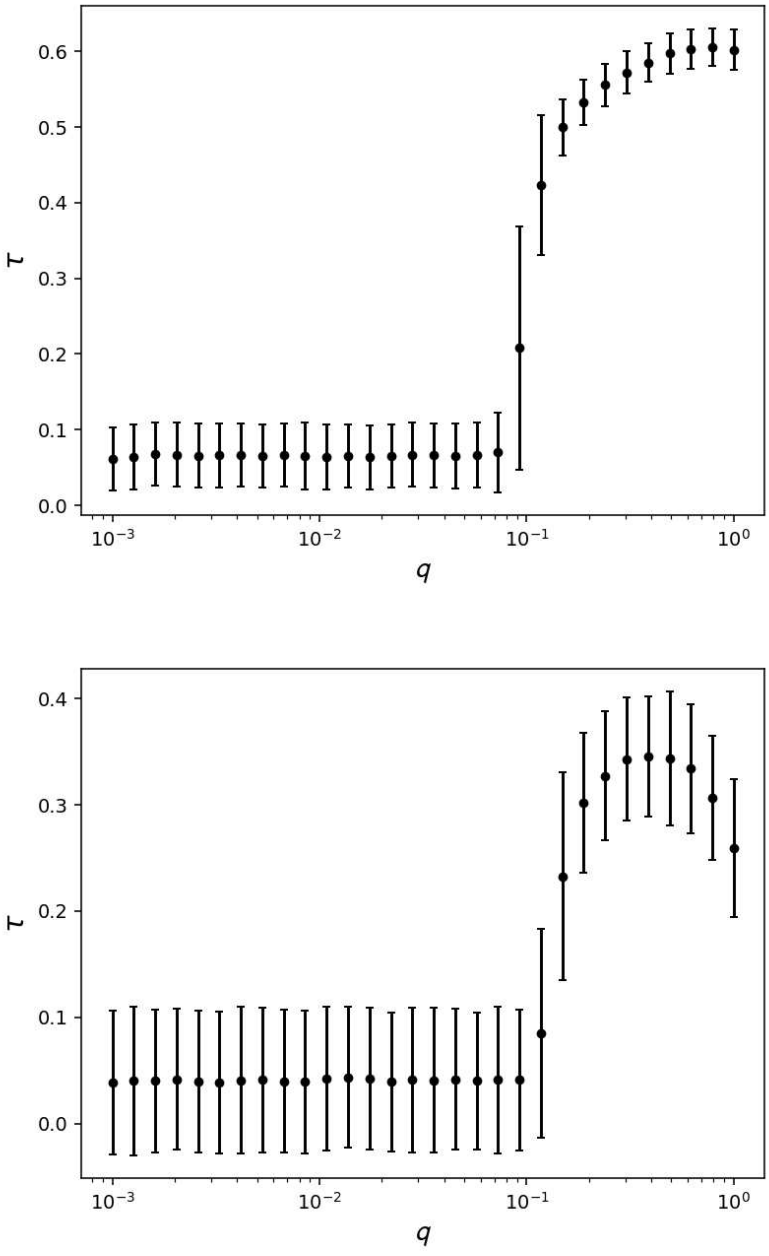

Figure 2.11: Average Kendall rank correlation coefficient τ between agent's payoff and fitness. Top panel: $N = 250$, $M = 100$. Bottom panel: $N = 100$, $M = 250$. The remaining parameters are the same as in other simulations and are reported in Remark 2.3.2.

and discordant otherwise. In our implementation, we compute Kendall's τ between the vectors of agents' average fitness and their corresponding payoffs. This yields a single scalar coefficient, which we interpret as an aggregate measure of the correlation between talent and outcome at the level of the entire society. In other words, τ summarises whether, on average, higher-skilled agents tend to achieve higher payoffs.

Results are shown in Fig. 2.11. When $q \to 0$, the rank correlation is near zero, implying a complete absence of meritocracy. After the phase transition at q^*, the value of τ increases, indicating that talent becomes increasingly aligned with payoff. Some variation in the shape of this transition is observed depending on the ratio between the number of agents N and the number of actions M: for $N > M$, the curve is monotonically increasing, while for $N < M$, a non-monotonic pattern may emerge. Nonetheless, the overall conclusion holds: higher q values correspond to more meritocratic societies. This insight aligns with findings in previous studies [171] and supports the conclusion that societies driven more by randomness than imitation may achieve a greater degree of equality and meritocracy. In this context, the emergence of desirable outcomes through randomness can be referred to as a form of serendipity, namely the occurrence and development of events by randomness in a happy or beneficial way [239, 240].

Action Occupancy and Collective Convergence

We next analyse how agents distribute themselves across the available actions. To quantify this, we examine the participation ratio (PR), a localisation measure that quantifies how evenly agents occupy the space of actions:

$$PR(t) = \frac{1}{M} \frac{1}{\sum_{j=1}^{M} \left(\tilde{f}_j(t) \right)^4}, \tag{2.30}$$

where $\tilde{f}_j(t)$ is the normalised vector of frequencies with respect to its L^2 norm:

$$\tilde{f}_j(t) = \frac{f_j(t)}{\sqrt{\sum_{l=1}^{M} f_l(t)^2}}. \tag{2.31}$$

This normalisation ensures that the vector $\tilde{f}(t) = \{\tilde{f}_j(t)\}_{j=1}^{M}$ has unit length in the L^2 sense, i.e.,

$$\sum_{j=1}^{M} \big(\tilde{f}_j(t)\big)^2 = 1. \tag{2.32}$$

As a consequence, $\tilde{f}_j(t)$ can be interpreted as the relative weight or "effective fraction" of agents associated with action j at time t. When most agents

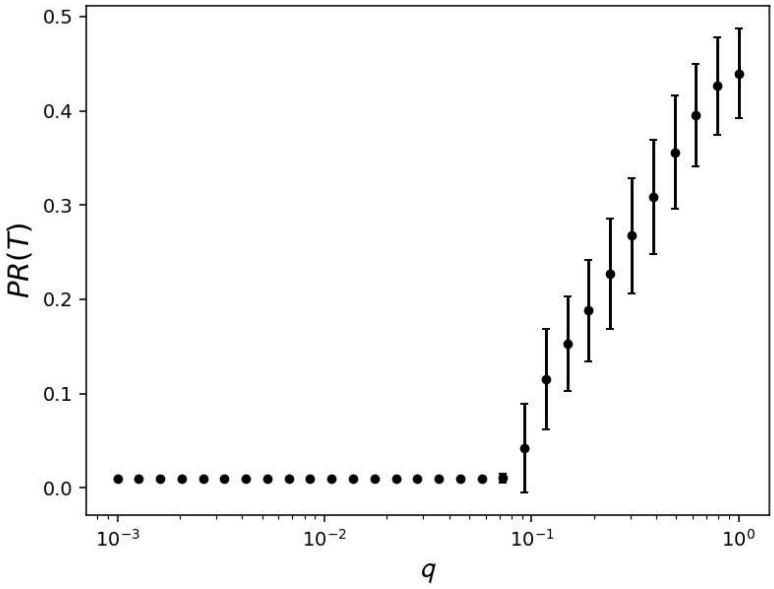

Figure 2.12: Participation ratio PR(T) of the entire community over the course of the simulation. Parameters are the same of other simulations and are reported in the text.

concentrate on a single action, one component $\tilde{f}_l(t)$ dominates, while the others are nearly zero (i.e., $\tilde{f}_l(t) = 1$ for some action l, $\tilde{f}_s(t) = 0$ for $s \neq l$). This leads the participation ratio to be $PR(t) = 1/M$. Conversely, if agents are evenly distributed across actions, all components of $\tilde{f}(t)$ contribute similarly (i.e., $\tilde{f}_j(t) = M^{-1/2}$ for each j), leading to $PR(t) = 1$.

We study the participation ratio at the end of the simulation, namely $PR(T)$. Results shown in Fig. 2.12 reveal that when imitation dominates ($q < q^*$) agents tend to converge on a single action, which represents an absorbing phase. Although agents may attempt to switch actions, they are likely to return to the most popular one due to its high occupancy, which reinforces its appeal. After the critical threshold q^*, agents begin to diversify and spread across multiple actions. However, the participation ratio remains below 1 in all cases, indicating that many possible actions remain unoccupied.

Socially Beneficial Actions

The previous analysis shows that agents tend to condense on a single action when imitation is strong. In order to scrutiny whether this dominant field is beneficial for society or not, we examine the distribution of societal impact

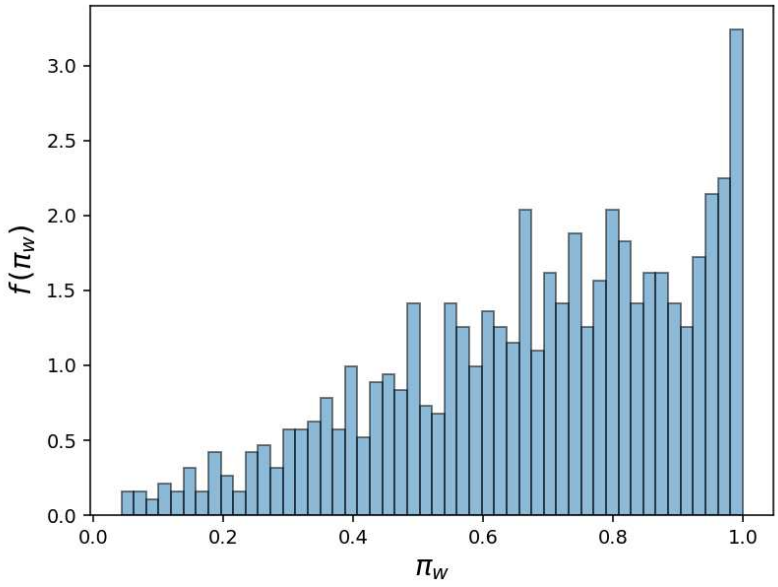

Figure 2.13: Normalised distribution of the beneficial impact π_j of the winning (i.e., most frequently played) action by agents at $q = 0$. The parameter settings are the same as in the other simulations and are reported in Remark 2.3.2.

value π_j associated with the most-played action (referred to as the "winning" action π_w) at $q = 0$. Results are shown in Fig. 2.13. While the winning action typically has an above-average benefit, in about 20% of simulations the system converges on an action with societal impact below 0.5. This suggests that imitation-driven convergence does not always favour high-value outcomes.

Allocation of Talent in High-Value Actions

Finally, we ask whether agents who engage in beneficial actions are also talented in those areas. To investigate this, we analyse the societal benefit metric defined in Eq. (2.25):

$$B(T) = \sum_{i=1}^{N} \sum_{t=0}^{T-1} \alpha_{ij(i,t)}\, \pi_{j(i,t)}.$$

which accounts for both the societal impact of the chosen actions and the agent's talent in those actions. Fig. 2.14 shows the results. For $q < q^*$, the average societal benefit $B(T)$ tends to be higher, yet it exhibits large

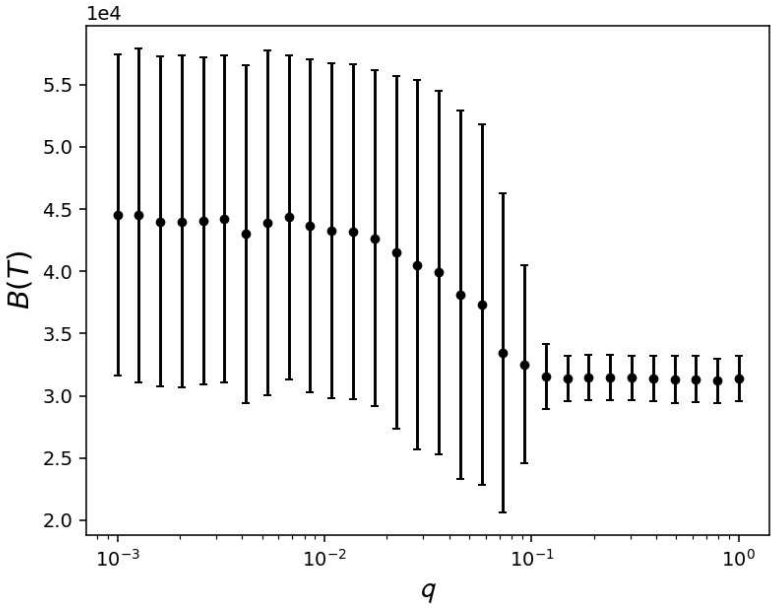

Figure 2.14: Overall benefit $B(T)$ of the entire community over the course of the simulation. The parameter settings are the same as in the other simulations and are reported in Remark 2.3.2.

variability across simulation runs. This suggests that, when imitation dominates, high-impact outcomes are possible but not guaranteed. As q increases beyond q^*, the system becomes more stable, with lower variability but also reduced average societal benefit. This produces more uniform (homogeneous) outcomes across simulations, as shown by the smaller standard deviations, but these outcomes are also less beneficial on average, reflecting a drop in collective efficiency.

Having addressed the key questions raised earlier through numerical simulations, we now turn to a simplified analytical framework to gain further insight into the underlying mechanisms of the model.

2.3.4 Analytical Insights

To gain analytical insight into the model's dynamics, we examine a minimal version of the system under the following assumptions:

a) Agents choose between two possible actions, i.e., $M = 2$;

b) The skill parameter is assumed to be homogeneous across all agents

and actions, such that $\alpha_{ij} = \alpha \ \forall i, j \in \{1, 2\}$;

c) The stochastic dynamics of the agent-based model are replaced by a deterministic approximation, where expected transitions between actions are described using transition rates;

d) Let $n_j(t)$ denote the number of agents performing action j at time t, subject to the constraint:

$$n_1(t) + n_2(t) = N, \tag{2.33}$$

where N is the total number of agents.

e) As in the baseline scenario, we define the popularity (relative frequency) of action 1 as

$$f(t) = \frac{n_1(t)}{N}. \tag{2.34}$$

Since the total population remains constant, the popularity of action 2 is simply $1 - f(t)$.

According to the dynamics described in Sec. 2.3.2, at each time step, agents decide whether to continue with their current action or switch to the other. The choice is based on three quantities: their skill α, the popularity of the action $f(t)$, and a control parameter $q \in [0, 1]$, that governs the balance between randomness and imitation.

We now compute the net change in the number of agents selecting action 1 from time t to $t + 1$. Starting with agents currently playing action 1, a fraction $\alpha f(t)$ continue playing it, due to a favourable match between skill and popularity. The remaining agents, a fraction $1-\alpha f(t)$, considers leaving. Among these, a portion $q/2$ will randomly choose to stay with action 1, while a portion $(1 - q)f(t)$ choose to stay through imitation. Therefore, the total number of agents in action 1 at time t who continue with it at time $t + 1$ (generally denoted n_{in}^{out}) is

$$n_1^1 = n_1(t) \left[\alpha f(t) + (1 - \alpha f(t)) \left(\frac{q}{2} + (1 - q)f(t)\right)\right]. \tag{2.35}$$

Next, consider agents currently playing action 2, whose number is $n_2(t) = N - n_1(t)$. Of these, a fraction $1-\alpha(1-f(t))$ are open to switching. Among the switchers, a fraction $q/2$ will randomly select action 1, and a fraction $(1-q)f(t)$ will imitate its popularity. Thus, the number of agents switching from action 2 to action 1 is

$$n_2^1 = (N - n_1(t)) \left(1 - \alpha(1 - f(t))\right) \left(\frac{q}{2} + (1 - q)f(t)\right). \tag{2.36}$$

The total number of agents choosing action 1 at time $t+1$ is therefore

$$n_1(t+1) = n_1^1 + n_2^1. \tag{2.37}$$

To express this in terms of $f(t)$, we divide both sides by N:

$$\begin{aligned} f(t+1) =& f(t)\left[\alpha f(t) + (1-\alpha f(t))\left(\frac{q}{2} + (1-q)f(t)\right)\right] \\ &+ (1-f(t))\left[1-\alpha(1-f(t))\right]\left(\frac{q}{2} + (1-q)f(t)\right). \end{aligned} \tag{2.38}$$

Now we compute the difference $f(t+1)-f(t)$, divide by Δt, and take the continuous-time limit $\Delta t \to 0$ to obtain the following differential equation:

$$\frac{df(t)}{dt} = (2\,f(t)-1)\;\alpha\,f(t)\,(1-f(t))\;\left((1-q)+\frac{q}{2}\right). \tag{2.39}$$

This equation governs the time evolution of the fraction of agents choosing action 1; at the same time, it recovers the fraction of agent in state 2 from Eqs. (2.33) and (2.34). The factor $(2f-1)$ drives the system toward reinforcing whichever option is more popular at time t, while the term $f(1-f)$ ensures zero net flow when all agents concentrate on a single action ($f=0$ or $f=1$). The coefficient modulates the influence of imitation and exploration, increasing with the skill level α and decreasing with the exploration parameter q.

To better understand the qualitative behaviour of the system, we analyse the steady-state solutions by solving

$$\frac{df}{dt} = 0.$$

This leads to a cubic equation with three fixed points: $f=0$, $f=1$, and $f=1/2$. The latter corresponds to a symmetric configuration where both actions are equally popular. Whether this symmetric solution is stable or not depends on the value of the exploration parameter q.

To investigate the stability of $f=1/2$, we define the constant:

$$C = \alpha\left((1-q)+\frac{q}{2}\right), \tag{2.40}$$

so that the differential equation simplifies to:

$$\frac{df}{dt} = C\,(2f-1)\,f\,(1-f). \tag{2.41}$$

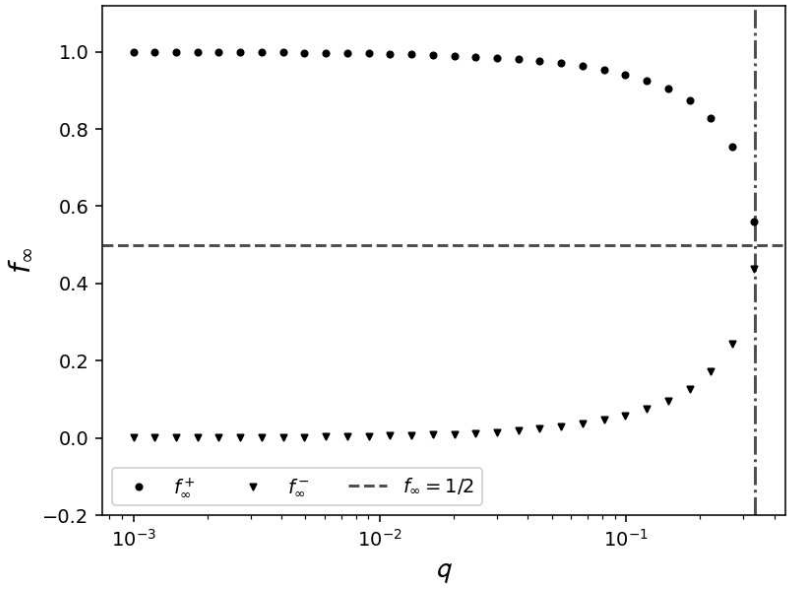

Figure 2.15: Bifurcation diagram showing the steady-state solutions f_∞ of the simplified model as a function of the exploration parameter q, for a fixed skill level $\alpha = 0.5$. Three branches of solutions emerge: a symmetric fixed point at $f = 0.5$ (shown in red) and two asymmetric solutions (shown in black) which become real and stable when q falls below the critical threshold $q^* = \frac{1}{3}$. For $q > q^*$, only the symmetric solution persists and becomes stable, indicating a transition to balanced coexistence.

Since $C > 0$ for all $\alpha > 0$ and $q < 1$, the fixed point at $f = 1/2$ is unstable, and the system tends to evolve toward either $f = 0$ or $f = 1$, depending on initial conditions. This behaviour reflects the dominance of imitation in low-exploration regimes, where popularity self-reinforces. To more precisely characterise the transition between symmetric and asymmetric behaviour, we return to the original cubic equation derived in the simplified analytical model. The two non-trivial stationary solutions (obtained numerically) are given by:

$$f_\infty = \frac{1}{2} \pm \frac{1}{2}\sqrt{\frac{q\,(\alpha - 2) + \alpha}{\alpha\,(1 - q)}}, \tag{2.42}$$

where f_∞ denotes the long-term popularity of action 1 (the occupation of state 2 can be easily recovered as they summed to 1). These solutions are real only when the factor under the square root is non-negative, which imposes the condition:

$$q \leq \frac{\alpha}{2 - \alpha}, \tag{2.43}$$

defining a critical threshold q^*.

Below this threshold, the system admits two asymmetric equilibria, and the dynamics are driven toward dominance of one action. Above the threshold, the symmetric state $f = 1/2$ becomes the only real and stable solution, representing balanced coexistence of both actions.

For example, when $\alpha = 0.5$ (see Fig. 2.15), the threshold is:

$$q^* = \frac{0.5}{2 - 0.5} = \frac{1}{3}. \tag{2.44}$$

Thus, for $q < 1/3$, the system exhibits spontaneous symmetry breaking, with one action dominating. For $q > 1/3$, exploration prevents condensation, and agents remain evenly distributed between the two actions in the long run.

Although the derivation of q^* is obtained in a simplified scenario, with two possible actions and under the assumption of constant and homogeneous skills (which, in the example provided, is equal to its average), it provides a clear illustration of how a critical point separates qualitatively distinct patterns, one where agents condensate on a single action and one characterised by evenly played fields. This analysis reveals the underlying mechanism behind the model's transition from imitation-driven consensus to exploration-induced diversity.

2.3.5 Alternative Setups

The baseline scenario described in Sec. 2.3.2 can be extended to better reflect empirical conditions. Without introducing excessive complexity, we consider the following three modifications:

a) A population divided into sub-groups, each characterised by a different value of parameter q;

b) A skill-learning mechanism, where agents improve their skills over time, i.e., $\alpha_{ij}(t)$;

c) A modified decision rule where the societal benefit π_j influences the dynamics.

Sub-Groups with Different q Values

The first modification is motivated by interpreting the parameter q as reflecting the agents' tendency toward exploration of the alternatives they can play. This approach is well-established in game theory and economics [241, 242].

Specifically, agents with low q behave as imitators, while individuals with high q are explorers. In the baseline model, all individuals share the same value of q, so the population is either fully imitative or fully explorative. When extended to sub-groups, this reduces to the baseline if one group numerically dominates the other.

To avoid trivial cases and promote heterogeneity, we consider two subgroups of comparable size: one group with a fixed $q_1 = 0.03$ and one with $q_2 \in [0, 1]$ which is varied across simulations. Fig. 2.16 displays the Gini coefficient and Kendall correlation coefficient for this mixed-population setup. To ease computation and focus on qualitative trends, smaller population sizes than in the baseline scenario are used. However, the results remain qualitatively consistent

Dynamic Skill Learning

In the second modification, agents improve their skills through repetition. If agent i continues to play action j, their skill at it improves. This is modelled by updating:

$$\alpha_{ij}(t) \to \alpha'_{ij}(t+1),$$

where $\alpha'_{ij}(t+1)$ is sampled from a uniform distribution in the interval $[\alpha_{ij}(t), 1]$. We track the average skill level for each action across the population over time:

$$\bar{\alpha}_j(t) = \frac{1}{N} \sum_{i=1}^{N} \alpha_{ij}(t). \tag{2.45}$$

This allows us to observe how well the population learns each action. Results for two regimes are shown in Fig. 2.17. In the top panel, we consider an imitation-dominated setting ($q = 0.1$), in which agents converge on a single action, increasing skill only in that domain while neglecting others. In the bottom panel, we examine an exploration-dominated setting ($q = 0.8$). In this scenario, different actions are played and agents improve their skills across many actions and eventually the average skill levels for all actions rise, approaching 1.

This dynamic highlights how exploration encourages broad-based skill development, while imitation leads to narrow specialisation. It reinforces our main finding: serendipity (interpreted here as high exploration) leads to more equitable, diverse, and ultimately positive outcomes, now extending to skills as well as payoffs.

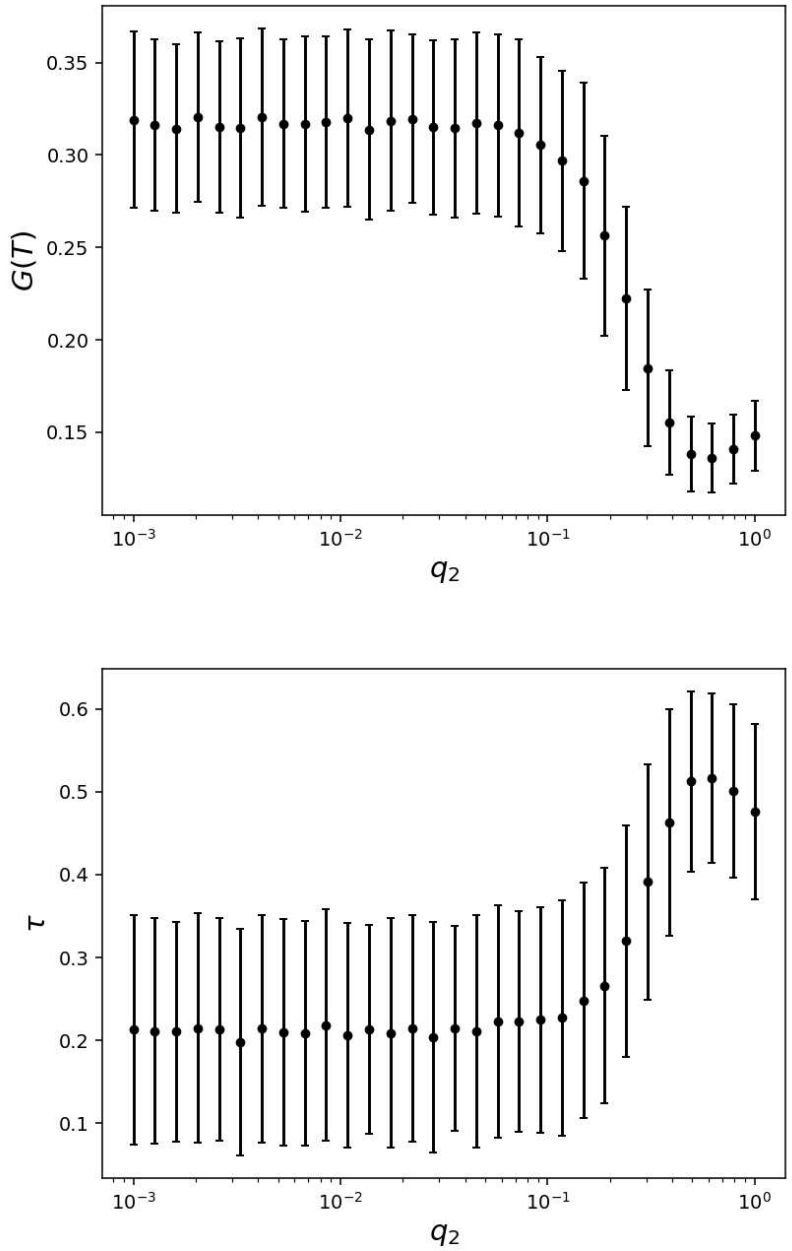

Figure 2.16: Gini coefficient (top panel) and Kendall correlation coefficient (bottom panel) for two sub-groups ($q_1 = 0.03$, $q_2 \in [0, 1]$ reported on the x-axis). Sub-groups sizes are $N_1 = 12$, $N_2 = 13$, with $M = 25$ actions, $T = 500$ time steps and sample size is $S = 500$.

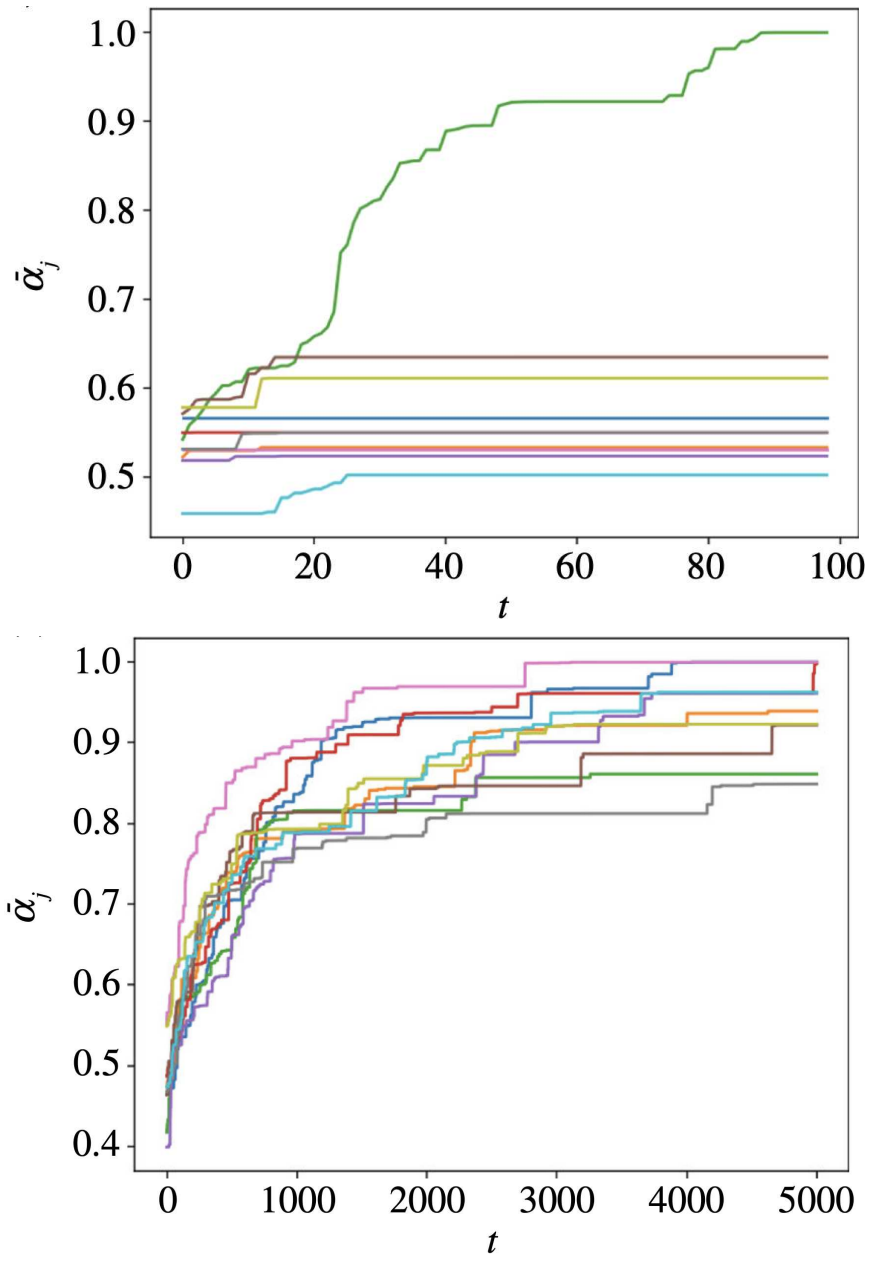

Figure 2.17: Temporal evolution of average skills $\bar{\alpha}_j(t)$ across all agents. Top panel: imitation regime ($q = 0.1$); bottom panel: exploration regime ($q = 0.8$). Parameters are: $N = 25$, $M = 10$, $T = 1000$, $S = 1000$. Note that different time horizons are used to account for the varying convergence rates.

Incorporating Societal Benefit into Dynamics

In the final modification, we alter the decision-making dynamics by incorporating the societal benefit π_j directly into the update rule, which becomes

$$\alpha_{ij}(t)\, f_j(t)\, \pi_j. \tag{2.46}$$

This modification means that an agent's decision is now also influenced by the societal impact of the action. However, simulations reveal that this change does not significantly affect the outcomes: neither the total payoff nor the Gini coefficient (see Fig. 2.18), nor the most played action at $q = 0$ (Fig. 2.19).

In this setting, we also revise the definition of fitness to incorporate societal impact, defining it as:

$$\langle \alpha_i \rangle = \frac{1}{M} \sum_{i=1}^{M} \alpha_{ij}\, \pi_j. \tag{2.47}$$

This new definition weights an agent's skill by the societal usefulness of the actions, providing a more meaningful measure for evaluating merit. A comparison between the original and the updated fitness-based Kendall correlation is shown in Fig. 2.20. No departure from baseline results can be perceived.

These alternative setups illustrate the model's flexibility in capturing more realistic dynamics and confirm the robustness of the core findings: while imitation can lead to higher payoff, it brings inequality and suboptimal collective choices. Exploration, on the other hand, fosters diversity, meritocracy, and more equitable outcomes. These results hold even when the decision rules are modified or heterogeneity is introduced.

2.3.6 Network Implementation

The robustness of results can be further assessed by implementing the dynamics on network structures (Sec. 1.8), where agents can directly observe and imitate each other. Specifically, we consider two different classes of network topologies: the first is a mean-field network, in which each agent is connected to all others; the second is an Erdős-Rényi random network, where links between agents are established independently with a fixed probability p [136, 147].

We begin by analysing the mean-field network, in which the system is fully

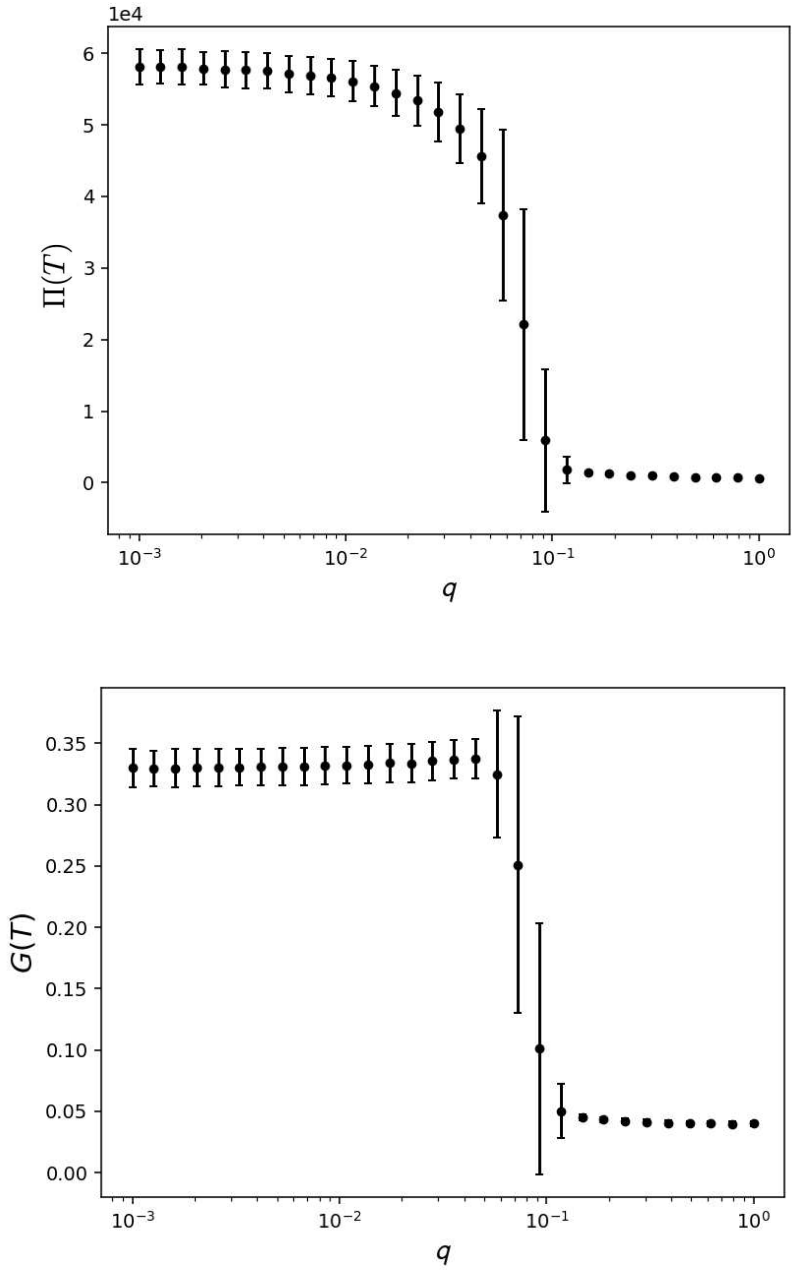

Figure 2.18: Top panel: Cumulative societal payoff $\Pi(T)$ at the end of the simulation for several values of the parameter q. Bottom panel: Gini coefficient $G(T)$ for the same values of q. The simulation involves $N = 25$ agents, $M = 10$ actions and $T = 1000$ time steps. Results are averaged over a sample of size $S = 1000$.

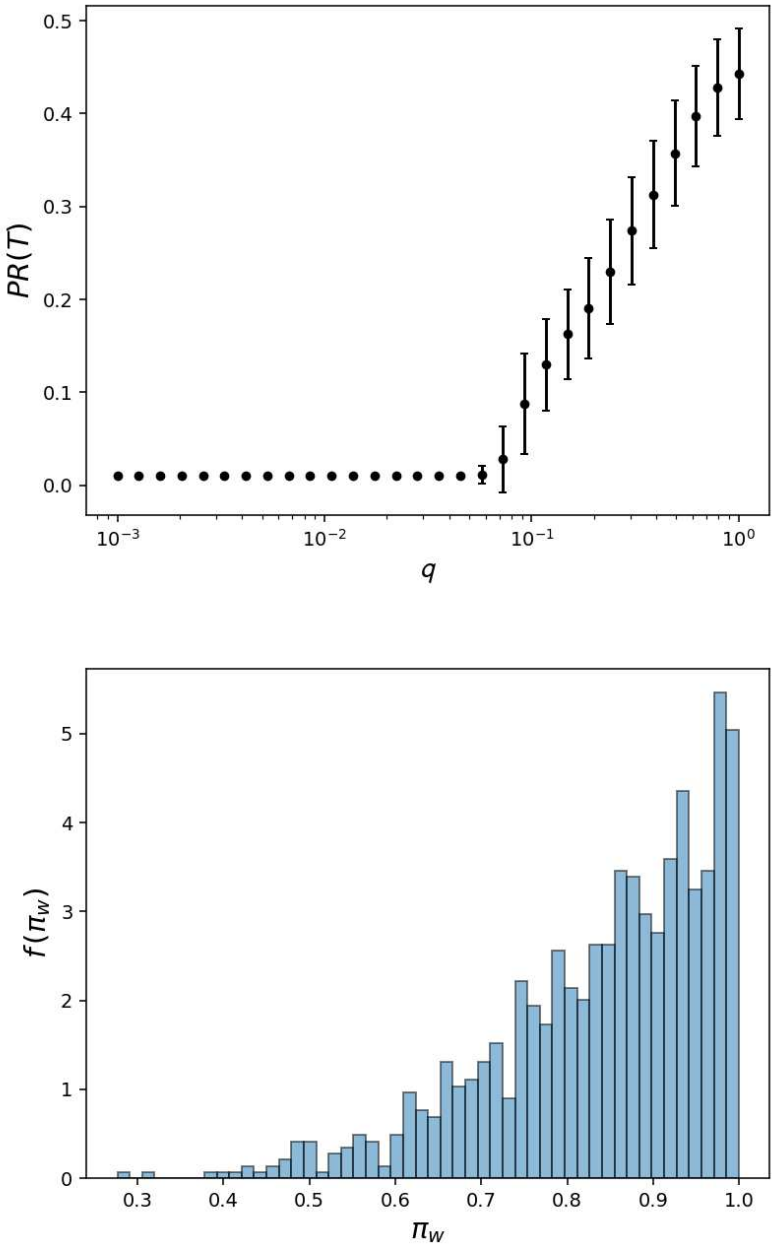

Figure 2.19: Top panel: Participation ratio $PR(T)$ at the end of the simulation. Bottom panel: normalised distribution of the payoffs of the most played actions at time T for $q = 0$. Parameters are given in Fig. 2.18.

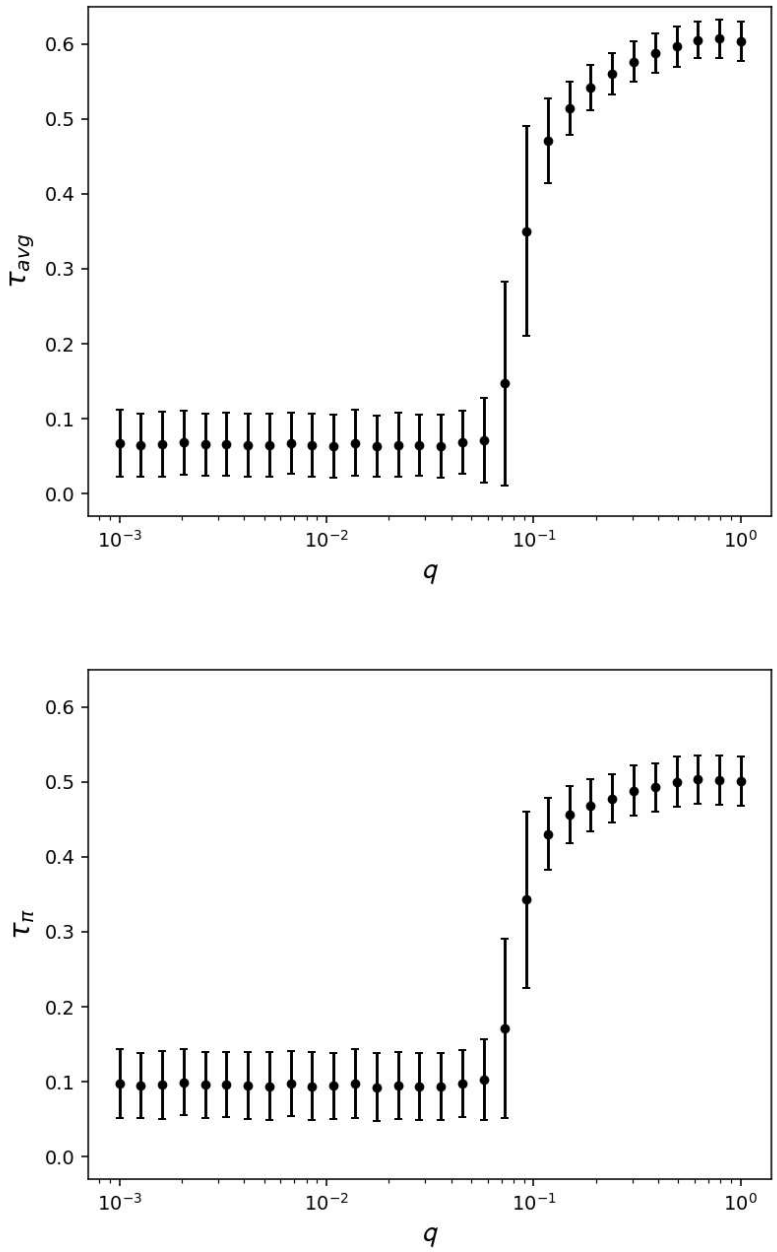

Figure 2.20: Kendall rank correlation coefficient τ between each agent's payoff $\Pi(T)$ and agent's fitness. Top panel: fitness is defined as $\Phi_i^{avg} = 1/M \sum_{j=1}^{M} \alpha_{ij}$. Bottom panel: fitness incorporates societal benefit, $\Phi_i^{\pi} = 1/M \sum_{j=1}^{M} \alpha_{ij}\, \pi_j$. Parameters are given in Fig. 2.18.

connected and each agent interacts with the entire population. As a consequence, the neighbourhood of any agent coincides with the full set of agents, and no differences in terms of interaction structure arise with respect to the standard, non-networked formulation of the model (Sec. 2.3.2). Numerical simulations confirm this analytical expectation. As shown in Fig. 2.21, the macroscopic observables coincide (within statistical fluctuations) with those obtained in the original dynamics. For reference, we report the total payoff and the Gini coefficient; the Kendall correlation coefficient and the participation ratio follow the same pattern as in the standard scenario and are therefore not shown.

The second type of network under examination is the Erdős-Rényi one. The average degree is linked to the probability of connections through the equation

$$\langle k \rangle = p\,(N-1), \tag{2.48}$$

for a system with N agents. In our simulations, we set $\langle k \rangle = 15$. While the overall behaviour remains consistent with the mean-field scenario, some minor differences emerge due to the local nature of interactions, as shown in Figs. 2.22 and 2.23. In particular, the total payoff $\Pi(T)$ decreases more gradually after the transition: agents may continue to exploit locally successful actions for longer times, mitigating the immediate loss of collective payoff observed in the mean-field case. Consistently, the Gini coefficient $G(T)$ exhibits a local maximum after the transition, suggesting that inequality can transiently persist even as imitation weakens. The Kendall correlation coefficient τ is systematically lower and displays a shallow minimum immediately after the transition, indicating a delayed recovery of meritocracy under locally reinforced dynamics. By contrast, the participation ratio $PR(T)$ closely follows the mean-field behaviour, indicating that the collective organisation of actions is unaffected.

The highlighted effects arise from finite and local information and do not signal the emergence of qualitatively new regimes, further supporting the robustness of the underlying mechanisms.

Overall, these results show that implementing the dynamics on a network structure does not alter the main conclusions, as the key results of the standard scenario are recovered.

2.3.7 Overview of Results

Rankings provide a means of reducing complexity by ordering entities (such as individuals, institutions, or countries) according to a specified performance

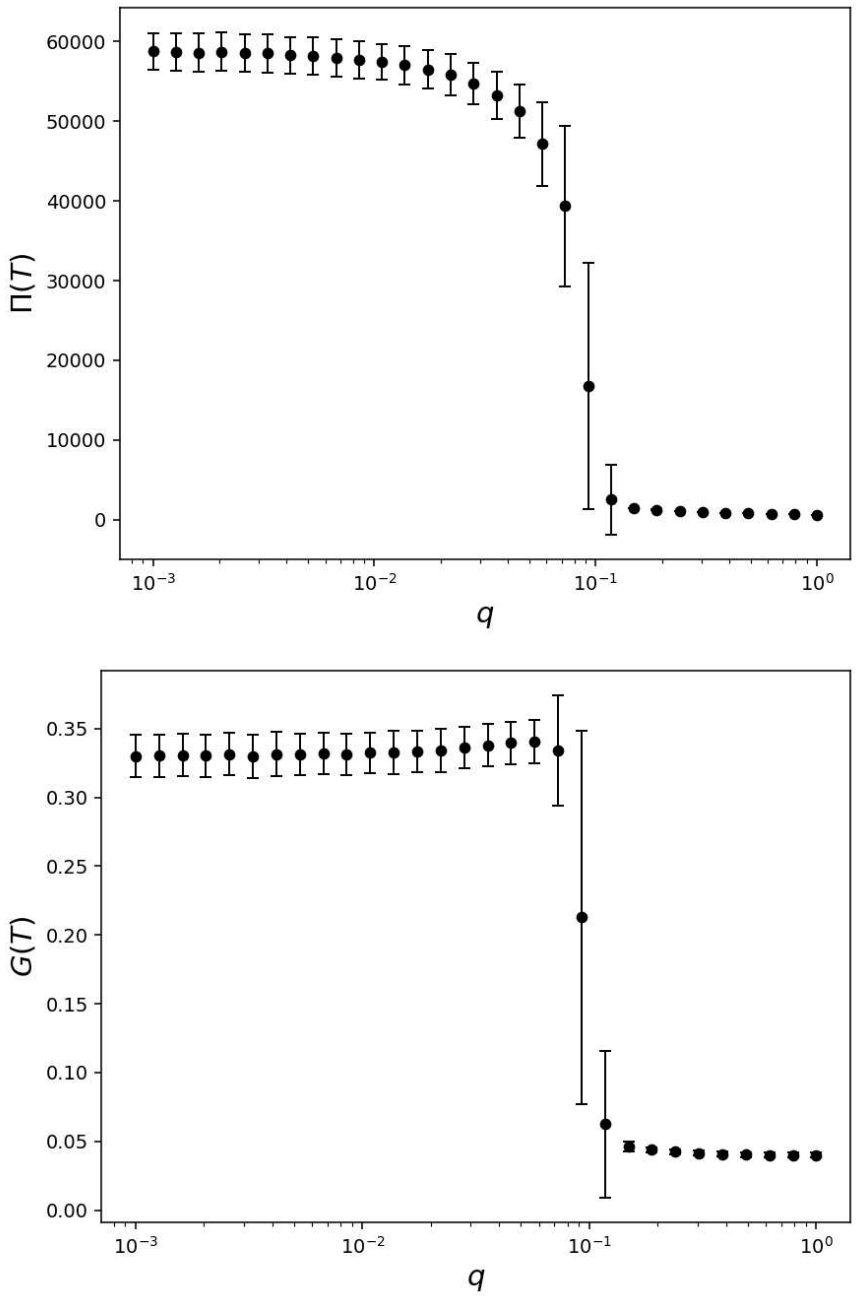

Figure 2.21: Overall payoff $\Pi(T)$ (top panel) and Gini coefficient $G(T)$ (bottom panel) for the dynamics implemented on a mean-field network. Parameter choices are reported in Remark 2.3.2.

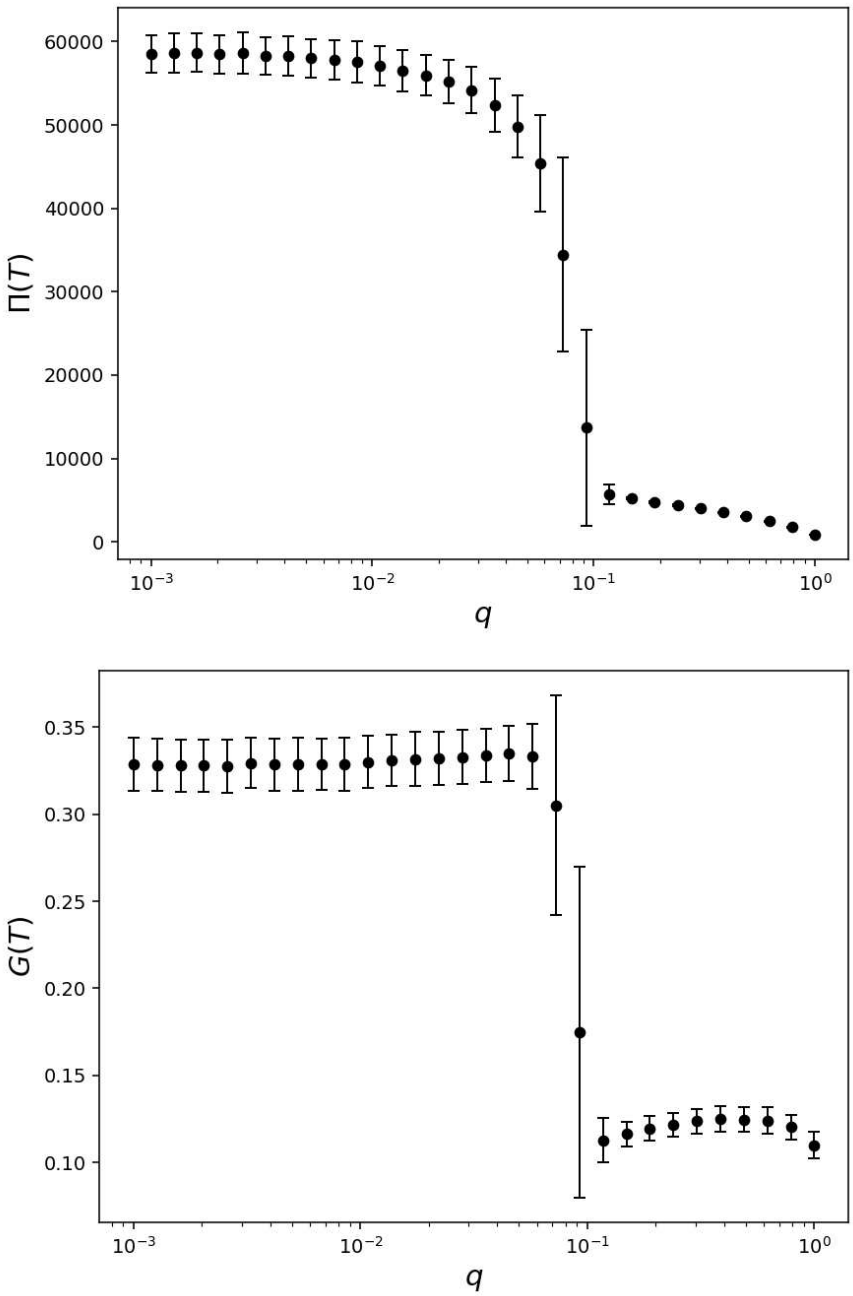

Figure 2.22: Overall payoff $\Pi(T)$ (top panel) and Gini coefficient $G(T)$ (bottom panel) for the dynamics implemented on an Erdős–Rényi network with average degree $\langle k \rangle = 15$. Parameter choices are reported in Remark 2.3.2.

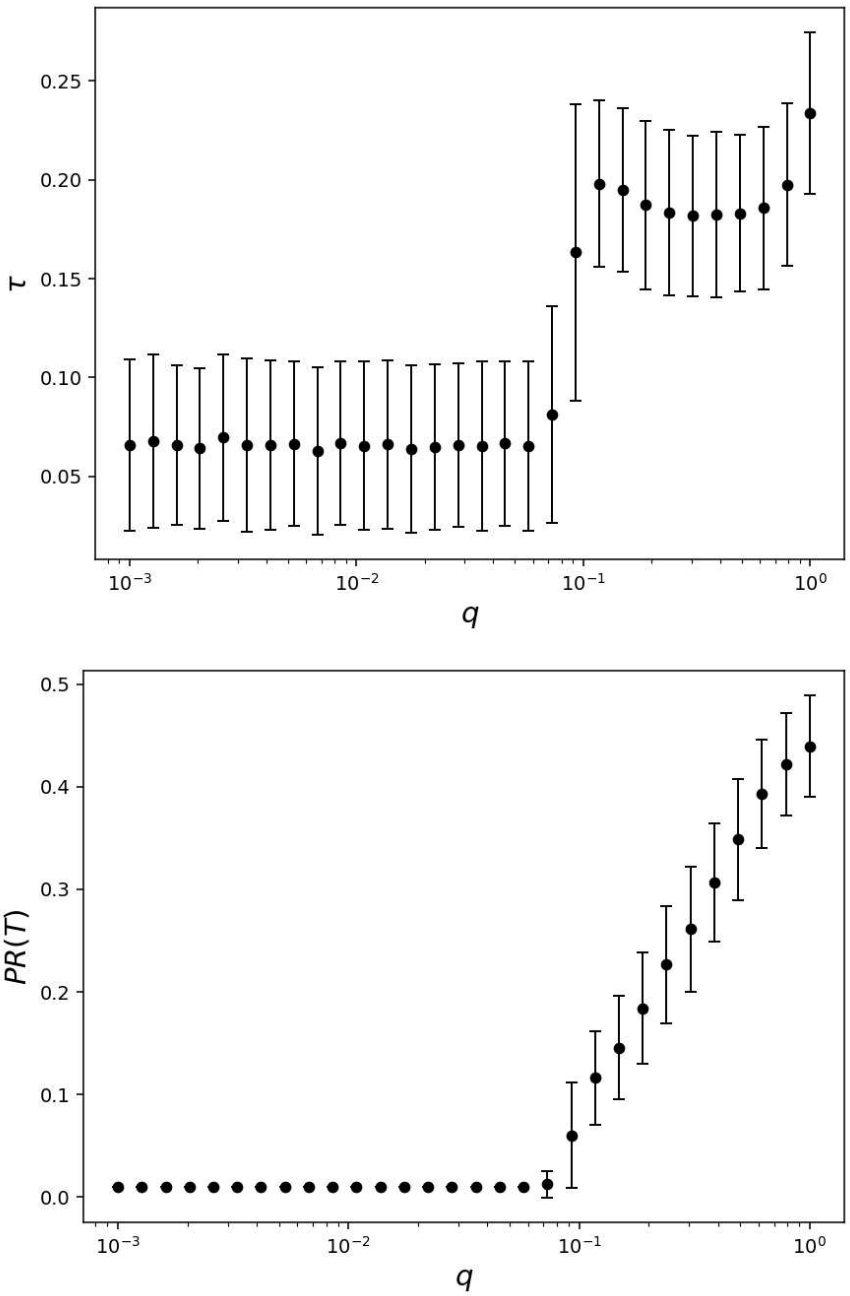

Figure 2.23: Kendall correlation coefficient τ (top panel) and participation ratio $PR(T)$ (bottom panel) for the dynamics implemented on an Erdős–Rényi network with average degree $\langle k \rangle = 15$. Parameter choices are reported in Remark 2.3.2.

metric. While such simplification can be useful (for example, enabling a better allocation of resources [216]), it carries several drawbacks. Firstly, evidence supports the common sentiment that the highest-ranked items are not the most talented [171,233] and that self-reinforcing mechanisms, such as the "rich-get-richer" dynamic (Sec. 1.5.3), often transform early success into long-lasting opportunities [243]. Another key feature is that individuals tend to imitate top performers in their pursuit of success; however, it has been shown that following one's own strategy is generally more beneficial [172].

To investigate the relationship between rankings, decision-making and social influence, we implement an artificial society. In our agent-based model, individuals can move across a fixed set of actions. At each time step, they receive a payoff corresponding to their chosen action and may remain in that field with a probability determined by their individual skill and the action's popularity, which encourages persistence in rewarding activities. Alternatively, they may consider switching. In that case, they can either imitate others by selecting an action based on its popularity (reflecting sensitivity to prestige, rankings, and a tendency to follow fashionable or trendy choices) or explore a new field at random.

The tendency to imitate or explore is governed by a parameter q, which defines the degree of imitation within the population and reflects the importance agents assign to rankings. Alternatively, q can be interpreted as the agent's propensity for exploration. When q is low, imitation dominates the decision-making process and agents primarily follow popular choices, behaving as imitators. Conversely, when q is high, agents act as explorers, selecting actions independently of popularity and engaging in random exploration.

To gain insight into the system's behaviour, we examine several key variables. First, we analyse the total payoff accumulated by agents over time. It reveals that, when imitation is the dominant mechanism and agents are influenced by rankings (i.e., action popularity), the aggregate payoff reaches its highest value. However, the Gini coefficient (a standard measure of inequality) shows that only a small number of agents contribute significantly to this total. Furthermore, the Kendall rank correlation coefficient indicates that these successful agents are not necessarily the most talented. The participation ratio (a metric of localisation) shows that, in this regime, all agents converge on a single action, which is not necessarily the most beneficial for society as a whole. Overall, when imitation prevails as the driving mechanism for switching actions, the society becomes unequal and non-meritocratic, with highly skilled agents failing to achieve corresponding success.

This phenomenon is attributable to the dominant role of imitation: during the early stages of the dynamics, a few agents may choose actions well-matched to their skill profiles. Their success attracts widespread imitation,

leading a disproportionate concentration of agents on the same actions. This collective behaviour generates a self-reinforcing loop that discourages individual exploration and prevents the majority of agents from identifying actions better aligned with their own abilities. These observations suggest that, while imitation may benefit a small subset of agents, it generally leads to suboptimal outcomes for the majority, including those who are, in principle, more talented.

Conversely, when randomness is the dominant mechanism, the overall accumulated payoff is lower, but it is more evenly distributed among agents, who populate the available fields more uniformly. Additionally, the correlation between payoff and talent becomes positive, indicating a more meritocratic system. In this context, we may speak of serendipity, namely a positive and unexpected outcome [239]. Thus, the system evolves towards greater diversity and fairness. Interestingly, randomness has also been shown to outperform more sophisticated strategies, for example in the context of financial trading [244].

Between these two regimes, a sharp transition is observed across all the measured quantities. This is a typical feature of complex systems, as discussed in Sec. 1.6.2. Notably, similar abrupt transitions have been documented in social systems where agents tend to conform to one another [193, 245]. Examples include social climbing phenomena [246], stock market dynamics [247] and the sudden emergence of traffic jams [248]. These analogies highlight the broader relevance of our findings.

Moreover, we explore different configurations of the model, including the incorporation of learning from repeated actions, the introduction of societal benefit into the dynamics, and the segmentation of the population into two subgroups. The results remain consistent with the baseline scenario, indicating that our model offers deep insights and displays robust behaviour under varying settings. Implementing the dynamics on networks further confirms the robustness of these results [249]. The abstraction from any specific domain ensures broad applicability of the model to systems governed by individual decision-making and collective dynamics, consistent with findings reported in [171, 243].

A clear real-world example of our model's setting is provided by the behaviour of the scientific community during the COVID-19 pandemic. As the crisis unfolded, researchers rapidly redirected their attention to pandemic-related topics. This thematic shift was largely driven by imitation: as initial COVID-19 studies gained visibility, increasing numbers of researchers followed the trend, leading to a dramatic rise in publications across diverse disciplines within just a few months [250]. However, this surge came with

drawbacks. Many studies published during this period were later criticised for methodological weaknesses or premature conclusions [251]. These outcomes underscore how the "publish or perish" culture prevalent in academia can amplify imitation dynamics [252, 253] and lead to condensation around specific topics, as described by our model. Moreover, such example underline the value of this type of modelling for informing institutional and policy decisions.

Beyond academia, similar dynamics are evident in entrepreneurial settings. Here, agents may be interpreted as startup founders choosing between imitating successful business models or pursuing innovative, riskier ventures. Evidence from recent startup trends indicates that imitation-based strategies (e.g., following popular sectors) often yield mixed outcomes [254]. This finding is consistent with our model's predictions: while imitation can lead to the dominance of a particular choice, it does not necessarily yield optimal results.

An important potential refinement of the model would be to allow the societal impact of each action to vary over time. In its current formulation, these values are fixed. However, in real-world contexts, the perceived value of different actions often changes due to external events or gradual societal transformations. For example, the COVID-19 pandemic significantly increased the relevance of research related to public health, rendering those topics more impactful.

In addition, the model could be empirically validated using the approaches discussed in Sec. 2.2.5, further grounding its relevance in real-world observations.

The results discussed in this model raise a number of ethical and policy-relevant considerations. In particular, the finding that imitation-driven dynamics can amplify inequality and weaken the connection between individual merit and collective outcomes challenges the fairness of social and institutional systems. This is especially relevant for settings that strongly rely on visibility, rankings, or mechanisms of cumulative advantage. Our model further confirms that such structures may systematically privilege early success, chance events, or social reinforcement over individual competence or effort [72, 225].

From an ethical perspective, our findings undermine meritocratic narratives by highlighting how unequal outcomes can arise even in the absence of intrinsic differences between individuals, and how responsibility for success or failure may consequently be misattributed at the personal level. From a policy standpoint, these results stress the importance of environments that limit excessive herding and reduce path dependence. From a practical per-

spective, introducing variability in ranking criteria over time, or avoiding rigid and uniform evaluation rules, may help prevent agents from merely adapting to fixed ranking mechanisms rather than competing on substantive performance. Promoting diversity of exposure and more decentralised decision-making processes may also help. More broadly, the analysis suggests caution when adopting performance metrics as proxies for individual talent, as institutional evaluation choices can translate into persistent and self-reinforcing inequalities.

While the models considered here are intentionally stylised, they indicate that addressing inequality may require deep structural interventions in the ways performance is evaluated and resources are allocated.

Having introduced and characterised the agent-based models proposed in [1] and [2], we now turn to a comparative summary of their main features. Our aim is to highlight the key similarities and differences in their mechanisms, as well as the principal contributions each study makes to the broader field of opinion dynamics. By examining the models side by side, we clarify how both frameworks capture critical transitions in collective behaviour and offer insights into the role of social influence in shaping individual decisions.

2.4 Summary and Discussion

The original contributions proposed in [1] and [2] and discussed in greater depth in this chapter provide new insights into how social influence shapes individual decision-making and emergent collective behaviour. These works extend the field of opinion dynamics by introducing two novel agent-based models that capture distinct yet complementary aspects of imitation and conformity within social systems.

Imitation plays a central role in human society: it is an innate learning mechanism and also satisfies the social need for belonging and acceptance. Often, individuals conform to the prevailing opinions or behaviours of the majority (not necessarily due to superior information) but out of fear of social exclusion. Such tendencies are evident across a wide range of domains, from financial decision-making [208] to dietary habits [255]. Our models provide a formal framework to understand the limitations and breakdown of this phenomenon under varying social pressures.

The two models presented, though distinct in their purposes, share common features: both are agent-based systems with discrete-time dynamics, where agents probabilistically decide whether to retain or alter their current state based on specific payoffs. The first model focuses on binary-state dy-

namics and incorporates both memory-dependent and memory-less settings. It explores how the interplay between social conformity and an external signal (interpreted as a ground truth) can lead to ergodicity breaking, which emerges as a phase transition. Specifically, as social influence increases, the system shifts from an ergodic regime (where agents remain responsive and able to track the external signal) to a non-ergodic regime (in which they become trapped in a single state and resistant to change). This transition highlights how strong social influence can inhibit adaptability and dampen responsiveness to new information. A key theoretical contribution of the model is the derivation of a closed-form expression for the critical threshold at which this shift occurs, paving the way for empirical validation of the transition threshold through data from surveys, controlled experiments, or digital analyses of collective behaviour.

The second model examines the interplay between imitation and randomness in shaping agents' behaviour, and is formulated as a memory-less system. Agents choose among multiple available states and may incorporate the relative popularity (i.e., rankings) of these options into their decision-making process. The model reveals how imitation leads to the emergence of a non-meritocratic society, with agents converging on a single action that is not necessarily beneficial for the group. Conversely, a more randomness-driven system (while seemingly less rational) leads to more meritocratic outcomes, where talent more fairly corresponds to success and where states are more equally explored.

Both models exhibit abrupt phase transitions, a hallmark of many complex systems. In one, the transition is from adaptability to inertia; in the other, from a non-meritocratic to a meritocratic regime. Such critical thresholds are not merely mathematical artefacts but they offer meaningful interpretations for real-world phenomena. For example, our models help explain how a small, strategically placed minority can shift majority opinion [161]. In the imitation-versus-randomness model, this occurs when a few agents take actions that are individually rewarding, prompting others to imitate them. In the ergodicity-breaking model, by contrast, strong conformist tendencies mean that interventions targeting a small number of well-positioned individuals may be more effective than broad, undifferentiated information campaigns. Conversely, when social influence is weaker, the clarity and perceived reliability of external signals play a more decisive role in shaping collective behaviour.

These insights have direct implications for policymaking. Social actions such as climate protest participation, vaccine uptake, or online activism are known to be highly sensitive to peer influence. The first framework captures this via a shift from oscillatory to absorbing states, consistent with empirical

findings that show how reaching a threshold of peer adoption often triggers self-sustaining cascades. The second model further illustrates that simply following popular actions does not necessarily lead to success, reinforcing the notion that an overreliance on rankings does not inherently promote societal merit or virtue.

Despite their stylised formulation, both models consistently reproduce key results across diverse scenarios. Variants incorporating memory effects (rather than memoryless dynamics) confirm the stability of the main results. In the ergodicity model, introducing memory smooths the impact of noise and strengthens alignment with longer-term trends. In the imitation-versus-randomness model, memory influences agents' skills through the introduction of a learning scheme; the resulting outcomes remain qualitatively unchanged.

Future work could enhance these models by calibrating them against real-world data through various methodologies. The increasing availability of large-scale datasets from social media and digital platforms provides a valuable resource for such efforts [164]. Additionally, incorporating continuous opinion states [163] could enable more nuanced representations of belief dynamics.

In summary, situated at the intersection of opinion dynamics and statistical physics, these works contribute to the growing body of research investigating how individual decision-making (shaped by stochasticity and social interactions) gives rise to emergent collective phenomena. Within this interdisciplinary framework, the models not only reproduce established patterns of collective behaviour, such as consensus formation, but also identify novel conditions under which societies converge toward either accurate or distorted shared beliefs. This modelling approach allows for a deeper understanding of the mechanisms behind large-scale opinion shifts and highlights the structural parameters that govern their dynamics.

Importantly, our findings indicate that imitation does not necessarily constitute the optimal strategy neither for individuals, who may remain unrecognised despite possessing genuine merit, nor for society, which may lose its capacity to align with objective truth. Relying exclusively on the opinions of others, driven by the fear of social judgement, may ultimately undermine both talent and success. In contrast, decisions based on individual information or even random exploration can sometimes lead to more fair and equal outcomes.

Chapter 3

Statistical Models for Finance

In recent decades, the intersection between physics and economics has grown into a vibrant research area, largely inspired by the tools and methods developed within statistical physics. This cross-disciplinary approach emerged as a field of interest in the 1980s, driven by the digitalisation of stock exchanges and the unprecedented availability of large-scale financial datasets. These technological advances enabled, for the first time, a systematic comparison between theoretical models and empirical observations [16, 256]. A landmark contribution in this direction was a 1991 paper by Mantegna: it was the first published in a physics journal and marked one of the earliest applications of physics-based techniques to financial data [257]. Shortly after, in 1995 the term "Econophysics" was coined by Stanley during the conference "Dynamics of Complex Systems" in Kolkata, a satellite meeting of the STATPHYS–19 conference planned in China [258, 259]. The name is the combination of economics and physics and acknowledged the growing role of physicists in economic research [260]. Today, the field is regarded as a branch of Complex Systems, focusing on the quantitative investigation of economic and financial data [259].

3.1 Research Motivation

Having briefly outlined the historical and disciplinary background, we now focus on the specific challenges and open questions that motivate the present work.

Financial markets are paradigmatic examples of complex systems: they involve a large number of interacting agents, each operating on the basis of their own expectations, beliefs, and information. These personal features introduce a high degree of unpredictability at the microscopic level. Yet,

at a macroscopic scale, statistical regularities emerge in financial time series, offering the potential for insight (and, in some cases, predictive power) regarding future dynamics. This approach relies on the assumption that evolution in financial markets occurs sufficiently slowly to be negligible in the near future [16]. Naturally, markets are not strictly stationary, and even this weak stability hypothesis breaks down during times of crisis. Nevertheless, the resulting models remain valuable tools for analysing complex financial phenomena.

The first quantitative models of price fluctuations, dating back to Bachelier [44] and later to the development of modern financial mathematics, typically assumed Gaussian statistics for returns, i.e., the relative changes in asset prices. However, the accumulation of high-quality empirical data has revealed a persistent departure from Gaussianity: return distributions display heavy tails, indicating that large fluctuations occur far more frequently than would be predicted by a normal distribution. This realisation has motivated the proposal of numerous non-Gaussian models (including Lévy-stable laws, truncated Lévy flights, Student's t, q-Gaussian, and modified Weibull distributions) each designed to capture the observed features of financial data. To date, no consensus has been reached on a single, universally accepted description.

On the basis of our original paper [3], the aim of this chapter is to provide a consistent and systematic comparison of some of the most widely studied heavy-tailed models in the econophysics literature. To our knowledge, no such direct comparison has been previously performed.

Consistent criteria are necessary for robust evaluation. Thus, we adopt parameter values grounded in empirical findings to ensure meaningful comparisons. Another key aspect of our study is the implementation of a simplified scenario for price dynamics, in line with the statistical physics tradition of gradually introducing complexity, so as to isolate the essential mechanisms at work before incorporating more realistic features. Based on both qualitative and quantitative analyses, we conclude that these models should not be considered as mutually exclusive alternatives, but rather as compatible tools yielding consistent parametrisations of real market dynamics.

The analyses presented here require substantial computational resources, especially for large-scale Monte Carlo simulations. This work was made possible thanks to the support provided by INFN, which is gratefully acknowledged. Selected codes, along with the datasets employed in the present study, are available at `github.com/fdedo/MSNGcodes`.

The chapter is organised as follows. Section 3.2 provides a concise overview of key financial concepts and the stylised facts of return distributions, laying the

groundwork for the discussion that follows. In Section 3.3, we present Gaussian and non-Gaussian models proposed for return distributions, which are qualitatively compared in Section 3.4, revealing strong similarities in their shapes. Section 3.5 describes techniques for generating random numbers from non-Gaussian distributions, a prerequisite for the quantitative analysis of price dynamics under simplified assumptions. We examine measures such as kurtosis and convergence to limit distributions (including convergence rates), further underscoring the similarities among the non-Gaussian models under analysis. Finally, Section 3.6 applies these models to option pricing, beginning with the standard diffusion framework for the dynamics and introducing a set of assumptions designed to enable a consistent comparison with the Gaussian benchmark. Similar behaviours are once again evident across the non-Gaussian models examined. Finally, we summarise the main findings and outline possible extensions for future work in Section 3.7.

3.2 Key Concepts in Econophysics

One of the most compelling aspects of econophysics (and financial theory more broadly) is its direct relevance to everyday life. Interest rates, currency exchange, and asset pricing all have a deep impact on individuals' lives. Yet, while these concepts permeate economic decision-making, their underlying mechanisms are often poorly understood outside specialist circles. For example, while many Italians are highly versed in cultural subjects such as music or art, it is often more difficult to find individuals who can explain the concept of compound interest [261]. This gap in practical financial understanding is concerning, especially given the growing complexity of modern financial systems.

In an attempt to address at least a fraction of this educational shortfall and to prepare the ground for the discussion that follows, the glossary below provides a concise overview of key financial concepts necessary for interpreting the models and results presented in this chapter. It does not aim to be exhaustive and the focus will be on the definitions, assumptions, and roles of the variables most directly connected to asset price dynamics, following the conventions in [16], unless stated otherwise.

- Market: Structured environment where buyers and sellers interact to exchange goods, services, or financial instruments, typically governed by rules that determine how prices are formed and transactions are executed. A market in which prices reflect all the available information at any time is said to be efficient [262]. This assumption underpins many

classical models, though empirical data often reveal deviations [263].

- Exchange: An organized and regulated marketplace (physical or electronic) where participants trade standardised financial instruments according to established rules. Exchanges centralise orders, provide transparent pricing, and ensure the clearing and settlement of transactions.

- Asset: Resource with economic value that can be owned or controlled with the expectation of future benefit. In finance, this typically refers to instruments like stocks, bonds, commodities, or currencies.
  Its market-determined value at a given time is denoted with $S(t)$ and it can be referred to as the spot price.

- Return: Relative change in asset price over a time interval Δt, namely

$$\frac{S(t+\Delta t)-S(t)}{S(t)}. \tag{3.1}$$

- Log-return: Logarithm of relative price changes over a time period Δt, which is

$$\ln\left(\frac{S(t+\Delta t)}{S(t)}\right). \tag{3.2}$$

  In the high-frequency limit (with time intervals down to milliseconds or even microseconds), returns and log-returns are approximately equivalent [264]. In fact, denoting

$$r=\frac{S(t+\Delta t)-S(t)}{S(t)}, \tag{3.3}$$

  we have that the first-order Taylor expansion of the logarithm holds:

$$\ln(1+r)\approx r, \tag{3.4}$$

  which leads to

$$\ln\left(\frac{S(t+\Delta t)}{S(t)}\right)\approx\frac{S(t+\Delta t)-S(t)}{S(t)}. \tag{3.5}$$

- Derivative: Financial contract whose value depends on underlying asset. Initially developed for agricultural hedging, derivatives now include options, defined below.

- Option: A call option grants the right (but not the obligation) to purchase an asset at a fixed price (the strike price) on or before a specified date (the expiration date). Analogously, a put option gives the right to sell under the same conditions. These contracts are often likened to insurance policies: a call option offers protection against rising prices, while a put safeguards against declines. While such instruments are commonly employed as hedging tools to mitigate specific risks, they are also widely used for speculative purposes. A more detailed classification is presented in Sec. 3.6.

- Stock (or share): Fractional ownership of a company. Publicly traded stocks are listed on exchanges, while private shares are exchanged through direct negotiation.

- Stock index: Aggregated measure of selected stock performances, published by exchanges (e.g., NYSE) or private firms (e.g., S&P, Dow Jones). Financial derivatives often track these indices, with the S&P 500 being a prominent example.

- Interest rate: Expressed as percentage, it measures the cost of borrowing or return on saving. It is typically associated with currencies, not countries.

- Volatility: Statistical measure of the magnitude of price fluctuations over a given time horizon, often used as a proxy for financial risk. In most models, its temporal evolution is described by a stochastic differential equation, reflecting the fact that volatility is not constant but time-varying and random. Volatility plays a dual role in finance: it is both an empirical object of measurement and a theoretical component in pricing and risk models. In empirical terms, it is typically treated as a latent process and must be estimated either from past returns (yielding the so-called historical volatility) or inferred from market prices of options (leading to implied volatility) [263, 265, 266].

- Time series: A collection of observations of a given asset recorded over a specific time window. Such data are sampled at discrete intervals defined either by market activity (tick-by-tick) or by physical time (e.g., seconds). Tick time advances when a price changes, while event time increments with each market event (e.g., order submission). This finer resolution enables the analysis of patterns such as intra-day seasonality or sudden spikes triggered by news. The high granularity of modern data (sometimes down to milliseconds) has become a cornerstone of financial modelling [267].

- Data Availability: Financial data can be proprietary (e.g., Bloomberg) or publicly available (e.g., Yahoo Finance). The completeness and quality of data are crucial: missing entries due to bankruptcy or mergers can bias analysis.

- Risk: The uncertainty associated with future financial outcomes, particularly the possibility of incurring losses. While often quantified through measures of volatility, such measures capture only typical price fluctuations and do not account for rare, high-impact events such as market crashes or systemic failures. A comprehensive view of risk encompasses both routine variability and extreme scenarios capable of destabilizing financial systems. Common categories include market risk (due to asset price fluctuations), credit risk (default by a counterparty), liquidity risk (difficulty in trading without significant price impact), operational risk (internal errors or failures), and systemic risk (widespread market breakdown triggered by institutional failures) [268].

3.2.1 Stylised Facts of Financial Time Series

Empirical analyses of return time series have revealed a number of robust statistical regularities that appear across different markets, asset classes, and time periods. These properties are often referred to as stylised facts and emerge as universal features of financial systems and are widely interpreted as signatures of underlying complex dynamics. Consequently, theoretical frameworks aiming to model financial markets must be capable of reproducing these empirical features, at least in an approximate sense [269–274]. The principal stylised facts can be summarised as follows [267]:

a) Heavy tails: The empirical distribution of asset returns and log-returns deviates significantly from the Gaussian benchmark, particularly at high-frequency time intervals. Such distributions exhibit a leptokurtic profile, characterised by a sharp central peak and heavy tails, implying a higher probability of extreme events than would be expected under normal assumptions [52, 275]. A characterisation of heavy-tailed distributions is provided in Sec. 1.5.

b) Absence of autocorrelations for returns: Returns computed over short time intervals (typically one to five minutes) exhibit negligible linear autocorrelations, indicating that they can be treated as approximately uncorrelated random variables, with past returns offering no predictive power for future ones [267, 273]. Minor deviations from this behaviour may occur at very short intra-day time scales (up to about 20

minutes), where weak correlations arise due to market microstructure effects rather than genuine informational inefficiencies [272].

c) Volatility clustering and non-linear dependence: The absence of linear autocorrelations in returns does not imply that returns are independent or identically distributed. In particular, transformations such as absolute or squared returns reveal a slowly decaying autocorrelation structure, often referred to as long memory. This phenomenon, known as volatility clustering, was first noted by Mandelbrot, who famously observed that "large changes tend to be followed by large changes (of either sign) and small changes tend to be followed by small changes" [52].

d) Aggregational normality: As the time scale over which returns are measured increases, the distribution gradually converges toward a Gaussian shape. This transition from a leptokurtic profile at high frequency to an approximately normal distribution at lower frequency is known as aggregational Gaussianity and is consistent with the Central Limit Theorem (Sec. 1.4.2) applied to weakly dependent variables. This behaviour is evident from the analysis of the Pareto exponent of the return distribution across different time scales, which reveals a systematic change in tail behaviour [276].

The key concepts outlined above provide both a shared vocabulary and an empirical benchmark for modelling financial time series. Real markets, however, introduce additional layers of complexity: trading is confined to specific hours, leading to discontinuities such as overnight gaps; volatility and trading activity follow pronounced intra-day and intra-week patterns; global asynchronicity can distort price formation.

In what follows, we examine Gaussian and non-Gaussian models of return distributions, assessing how well each framework captures the stylised facts and market complexities discussed here.

3.3 Gaussian vs. Non-Gaussian Distributions

Financial returns are typically represented as realisations of stochastic processes (Sec. 1.3) and their statistical modelling has a long and well-documented history, beginning with the Gaussian framework, which served as the principal reference throughout much of the twentieth century. In this chapter, we start by revisiting this model, in recognition of its foundational role and because its simplifying assumptions make it a natural benchmark for subsequent developments. While any representation of empirical data necessarily

involves a degree of abstraction, decades of research have produced alternative formulations that more accurately capture observed regularities, most notably the heavy-tailed nature of return distributions. Thus, we introduce a selection of non-Gaussian models specifically constructed to reflect the empirical characteristics of financial time series more faithfully.

3.3.1 The Standard Model of Finance and Beyond

The Gaussian framework traces its origins to the early work of Bachelier, who proposed modelling asset prices using arithmetic Brownian motion (Eq. (1.5)), at the beginning of the twentieth century [44]. This formulation implied normally distributed price changes, but it also allowed for negative prices, a feature that is clearly unrealistic in financial markets [277].

To overcome this limitation, the first widely accepted alternative was introduced independently by Samuelson [278] and Osborne [279], based on the assumption that asset prices evolve according to a geometric Brownian motion (GBM), given in Eq. (1.6). Therefore, this process preserves the Gaussian nature of returns while ensuring strictly positive prices. The stochastic differential equation governing this process is:

$$\mathrm{d}S(t) = \mu\, S(t)\, \mathrm{d}t + \sigma\, S(t)\, \mathrm{d}W(t)\,, \tag{3.6}$$

where $\mathrm{d}W(t)$ is the increment of a Wiener process, μ is the drift rate and σ denotes the volatility. In Samuelson's formulation, both μ and σ are assumed to be constant; this is a simplification that, despite its limitations, we adopt in our discussion for analytical tractability.

Under the GBM, returns satisfy the equation:

$$\frac{\mathrm{d}S(t)}{S(t)} = \mu\, \mathrm{d}t + \sigma\, \mathrm{d}W(t)\,. \tag{3.7}$$

Applying Itô's lemma (Sec. 1.3.3) to the logarithm of the price process yields the dynamics [277]:

$$\mathrm{d}\ln S(t) = \left\{\mu S(t)\frac{\partial \ln S(t)}{\partial S(t)} + \frac{1}{2}\sigma^2\, S^2(t)\frac{\partial^2 \ln S(t)}{\partial S^2(t)}\right\} + \sigma\, S(t)\, \frac{\partial \ln S(t)}{\partial S(t)}\, \mathrm{d}W(t).$$

Using the derivatives:

$$\frac{\partial \ln S(t)}{\partial S(t)} = \frac{1}{S(t)},$$
$$\frac{\partial^2 \ln S(t)}{\partial S^2(t)} = -\frac{1}{S^2(t)}\,.$$

the expression simplifies to

$$\mathrm{d}\ln S(t) = \left(\mu - \frac{\sigma^2}{2}\right)\,\mathrm{d}t + \sigma\,\mathrm{d}W(t). \tag{3.8}$$

Integrating this stochastic differential equation between the two times t and T ($t < T$) gives

$$\ln S_T - \ln S_t = \left(\mu - \frac{\sigma^2}{2}\right)(T-t) + \sigma\big[W(T) - W(t)\big]. \tag{3.9}$$

Since the Wiener process is normally distributed with zero mean and variance equal to the length of the time interval (Sec. 1.3.1), it follows that $\ln S_T$ is normally distributed with

$$\mathbb{E}[\ln S_T] = \left(\mu - \frac{\sigma^2}{2}\right)(T-t), \tag{3.10}$$

$$\mathrm{Var}[\ln S_T] = \sigma^2(T-t). \tag{3.11}$$

The corresponding probability density function for the log-returns is:

$$p\left(\ln\frac{S_T}{S_t}\right) = \frac{1}{\sqrt{2\,\pi\,\sigma^2\,(T-t)}}\exp\left(-\frac{\left[\ln\left(\frac{S_T}{S_t}\right) - \left(\mu - \frac{\sigma^2}{2}\right)(T-t)\right]^2}{2\,\sigma^2\,(T-t)}\right), \tag{3.12}$$

From Eq. (3.7), returns themselves follow a normal distribution with

$$\mathbb{E}\left[\frac{\mathrm{d}S(t)}{S(t)}\right] = \mu\,(T-t), \tag{3.13}$$

$$\mathrm{Var}\left[\frac{\mathrm{d}S(t)}{S(t)}\right] = \sigma^2(T-t). \tag{3.14}$$

This result is fully consistent with the earlier derivation for log-returns: since $\ln S_T$ is normally distributed, the price S_T itself is log-normally distributed according to

$$p(S_T) = \frac{1}{S_T}\,p\left(\ln\frac{S_T}{S_t}\right), \tag{3.15}$$

where $p(\cdot)$ denotes the Gaussian density given in Eq. (3.12). The corresponding moments are

$$\mathbb{E}[S_T] = S_t\,\exp\left[\mu(T-t)\right], \tag{3.16}$$

$$\mathrm{Var}[S_T] = S_t^2\,\exp\left[2\mu(T-t)\right]\left(\exp[\sigma^2(T-t)] - 1\right). \tag{3.17}$$

While the GBM framework avoids the unrealistic possibility of negative prices, it remains grounded in assumptions that are often inconsistent with empirical observations. In particular, it presumes constant volatility, independent and identically distributed (i.i.d.) returns, and normally distributed increments with finite variance. However, financial markets frequently exhibit volatility clustering, temporal correlations, and heavy-tailed return distributions that deviate significantly from Gaussian behaviour.

In a seminal 1963 study, Mandelbrot challenged the adequacy of the normal distribution in financial modelling. He demonstrated that empirical returns are better captured by stable Paretian (Lévy-stable) distributions (Sec. 1.4) with tail exponents around $\alpha \simeq 1.7$ [52]. His analysis, initially based on cotton prices, was soon extended to other products, including wheat, grains, and financial securities. Two years later, Fama provided strong empirical support for Mandelbrot's hypothesis [280, 281]. While reasonable in principle, this approach does not accurately reproduce the empirical characteristics of outliers and therefore falls short as a complete modelling framework.

A major refinement came from the work of Mantegna and Stanley [273, 282], who analysed high-frequency (one-minute) returns of the S&P 500 index over the period January 1984 to December 1989. Their examination of the full distribution revealed a leptokurtic profile, sharply peaked at the centre with heavy tails. This feature identified as the first stylised fact in Sec. 3.2.1. They proposed Lévy stable model with tail exponent $\alpha \simeq 1.4$, which improved the fit in the central region but still underestimated the empirical tail decay.

Notably, the complementary cumulative distribution function exhibits power-law scaling (Sec. 1.5.1) with tail exponent $\alpha \approx 3$, observed over timescales ranging from 1 to 120 minutes. This scaling has been documented for stock price variations [264], market indices [275], and individual company returns [283], all of which lie well outside the Lévy stable regime ($0 < \alpha < 2$) originally proposed.

The empirical inadequacy of the proposed models has led to the development of several non-Gaussian distributions designed to better capture the behaviour of high-frequency returns and log-returns. While no universal consensus has emerged [267], several distributions have gained widespread attention, including Student's t, q-Gaussian, Truncated Lévy, Modified Weibull, hyperbolic and normal inverse Gaussian distributions. These models exhibit power-law tails, with tail exponents $\alpha > 2$, ensuring a finite second moment (i.e., finite variance) though the precise value of α remains debated.

In this dissertation, following our original contribution [3], we examine

and compare four non-Gaussian models, adopting consistent criteria to enable a meaningful evaluation. Specifically, our focus lies on probability distribution functions that are centred (i.e., have zero mean), exhibit finite variance, positive excess kurtosis, and negligible skewness, all features that align with key empirical findings. Given our primary objective of performing a coherent comparative analysis, particular emphasis is placed on ensuring consistency in both variance and excess kurtosis across the models. The latter, in particular, is a statistical measure that quantifies the heaviness of the tails of a distribution, thereby reflecting the likelihood of extreme deviations from the mean [284].

In the Pearson formulation, kurtosis is defined as the ratio between the fourth central moment and the square of the second central moment:

$$\kappa = \frac{\mathbb{E}[(x-\mu)^4]}{(\mathbb{E}[(x-\mu)^2])^2} = \frac{\mu_4}{\mu_2^2}, \tag{3.18}$$

where μ denotes the mean of the distribution, and μ_i represents its i-th central moment.

An alternative and commonly used metric is the Fisher definition, or excess kurtosis, given by:

$$\kappa_E = \frac{\mathbb{E}[(x-\mu)^4]}{(\mathbb{E}[(x-\mu)^2])^2} - 3 = \frac{\mu_4}{\mu_2^2} - 3. \tag{3.19}$$

Under this convention, the Gaussian distribution has an excess kurtosis of zero, making κ_E a useful measure of deviation from normality. Distributions with $\kappa_E > 0$ are defined leptokurtic, indicating heavier tails and a sharper peak compared to the normal distribution, while platykurtic distributions, with $\kappa_E < 0$, appear flatter and exhibit lighter tails.

It is worth noting that the term kurtosis is often used in the literature to refer to excess kurtosis, sometimes without explicit clarification.

Having outlined the key statistical properties observed in empirical data, we now focus on specific probability models that have been proposed to account for such features. In what follows, the variable x will denote returns or, in the high-frequency limit, log-returns, as the distinction between the two becomes negligible.

3.3.2 Student's *t*-Distribution

We first examine the generalised Student's t-distribution, originally proposed in econophysics by Bouchaud and Potters as a suitable framework for modelling financial returns [16, 285–287]. It is defined by the probability density

function:

$$p_S(x) = \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\sqrt{\nu\,\pi}\,\hat{\sigma}\,\Gamma\left(\frac{\nu}{2}\right)}\left[1+\frac{1}{\nu}\left(\frac{x}{\hat{\sigma}}\right)^2\right]^{-\frac{\nu+1}{2}}, \tag{3.20}$$

where $\hat{\sigma}$ is a scale parameter (modulating the width of the distribution), ν denotes the number of degrees of freedom (shaping its form) and Γ is the Gamma function, satisfying $\Gamma(n) = (n-1)!$. The location parameter is set to zero, consistent with the standard assumption for high-frequency financial returns, which exhibit negligible mean or are reported to zero accordingly.

The tail behaviour of the distribution follows a power-law decay:

$$p_S(x) \xrightarrow{|x|\to\infty} |x|^{-(\nu+1)}, \tag{3.21}$$

indicating heavy tails. The distribution approaches the Gaussian form as $\nu \to \infty$, with convergence effectively occurring for $\nu \gtrsim 30$.

The variance of the distribution is given by:

$$\mathrm{Var_S\{x\}} = \frac{\hat{\sigma}^2\,\nu}{\nu-2}, \qquad \text{for}\,\nu > 2. \tag{3.22}$$

More generally, the existence of moments is governed by the asymptotic decay: the k-th moment exists only for $\nu > k$. In particular, the fourth central moment, required for computing the kurtosis, is:

$$\mathbb{E}[x^4] = \int_{-\infty}^{\infty} x^4\,p(x)\,\mathrm{d}x. \tag{3.23}$$

Approximating the integrand for large $|x|$ gives:

$$\mathbb{E}[x^4] \approx \int_{|x|\to\infty} \frac{x^4}{|x|^{\nu+1}}\,\mathrm{d}x = \int_{|x|\to\infty} |x|^{4-(\nu+1)}\,\mathrm{d}x. \tag{3.24}$$

which converges only if $4-(\nu+1) < -1$, or equivalently $\nu > 4$. Therefore, the fourth moment is finite only for $\nu > 4$, and the kurtosis diverges for $\nu \leq 4$. When finite, the excess kurtosis takes the form:

$$\kappa_S = \frac{6}{\nu-4} \qquad \text{for}\,\nu > 4. \tag{3.25}$$

Empirical studies on asset returns typically report values around $\nu \simeq 3$ for intra-day data [16,288,289], while daily returns are often better described by $\nu \simeq 4$ [16,287].

3.3.3 q-Gaussian Distribution

We now turn to the q-Gaussian distribution [290, 291], introduced by Tsallis in the context of non-extensive statistical mechanics [292] and later developed in collaboration with his co-authors as a realistic model for financial data [293]. It is defined as:

$$p_q(x) = \frac{\sqrt{\beta}}{C_q}\, e_q(-\beta\, x^2), \tag{3.26}$$

where β is a scale parameter, q governs the shape of the distribution and $e_q(x)$ denotes the q-exponential function, given by:

$$e_q(x) = [1 + (1-q)\, x]^{\frac{1}{1-q}} \qquad \text{for } q \neq 1.$$

The normalisation constant $C(q)$ ensures that the distribution integrates to one. In the range of interest $1 < q < 3$, it takes the form:

$$C_q = \sqrt{\frac{\pi}{q-1}}\, \frac{\Gamma\left(\frac{3-q}{2\,(q-1)}\right)}{\Gamma\left(\frac{1}{q-1}\right)}.$$

For $q = 1$, the q-exponential reduces to the standard exponential, and the Gaussian distribution is recovered.
The asymptotic behaviour for $q > 1$ follows a power-law decay:

$$p_q(x) \xrightarrow{|x|\to\infty} |x|^{-2/(q-1)}.$$

The variance is finite for $q < 5/3$ and is given by:

$$\mathrm{Var}_q\{x\} = \frac{1}{\beta(5-3\,q)} \qquad \text{for } q < 5/3. \tag{3.27}$$

The existence of higher moments is again governed by the asymptotic tail. In particular, the excess kurtosis is finite only for $q < 7/5$, and, when defined, it takes the form:

$$k_q = \frac{6\,(q-1)}{7-5\,q}. \tag{3.28}$$

Empirical analyses typically find $q \simeq [1.4, 1.5]$ for both intra-day and daily returns [293, 294].

An analytical equivalence exists between the q-Gaussian distribution in Eq.(3.26)

and the Student's t-distribution in Eq.(3.20), provided the parameters are related by:

$$q = \frac{\nu + 3}{\nu + 1} \quad \text{with } \beta = \frac{1}{(3-q)\,\hat{\sigma}^2}. \tag{3.29}$$

This correspondence underscores the close relationship between the two distribution families, both of which have been proposed as effective models for capturing the heavy tails observed in empirical financial data.

3.3.4 Modified Weibull Distribution

As a third candidate model, we now focus on the Modified Weibull Distribution (MWD), which was proposed by Sornette and collaborators [295–298]. Its probability density function is given by:

$$p_{MWD}(x) = \frac{1}{2\sqrt{\pi}} \frac{c}{\chi} \left(\frac{|x|}{\chi}\right)^{\frac{c}{2}-1} \exp\left[-\left(\frac{|x|}{\chi}\right)^{c}\right], \tag{3.30}$$

where $\chi > 0$ is a scale parameter and $c > 0$ is a shape parameter. The distribution is symmetric and exhibits tails that decay faster or slower than exponential depending on the value of c. For $c < 1$, the density decays more slowly than an exponential, and is often referred to as a stretched exponential. In the special case $c = 2$ and $\chi = \sqrt{2}$, the standard Gaussian is recovered.

The variance is finite for all $c > 0$ and is given by:

$$\mathrm{Var}_{MWD}\{x\} = \chi^2 \cdot \frac{\Gamma\left(\frac{1}{2} + \frac{2}{c}\right)}{\sqrt{\pi}}. \tag{3.31}$$

The excess kurtosis, when finite, takes the form:

$$k_{MWD} = \frac{\Gamma\left(\frac{1}{2} + \frac{4}{c}\right)}{\left[\Gamma\left(\frac{1}{2} + \frac{2}{c}\right)\right]^2} \sqrt{\pi} - 3. \tag{3.32}$$

This expression can be derived by recognizing that the Modified Weibull is a symmetric case of the Generalised Gamma distribution, with domain $x \in \mathbb{R}$, a broader family encompassing several well-known continuous models. The corresponding probability density function is:

$$p(x) = \frac{k}{a\,\Gamma\left(\frac{d}{k}\right)} \left(\frac{x}{a}\right)^{d-1} \exp\left[-\left(\frac{x}{a}\right)^{k}\right], \quad x > 0, \tag{3.33}$$

where $a > 0$ is the scale parameter, $d > 0$ a shape parameter and $k > 0$ an exponent. By setting $a = \chi$, $d = c/2$, and $k = c$, which is a parametrisation that reproduces the Modified Weibull (up to normalisation and symmetry), we obtain $d/k = 1/2$, from which the relevant raw moments follow:

$$\mathbb{E}[x^2] = \chi^2 \, \frac{\Gamma\left(\frac{1}{2} + \frac{2}{c}\right)}{\Gamma\left(\frac{1}{2}\right)}, \tag{3.34}$$

$$\mathbb{E}[x^4] = \chi^4 \, \frac{\Gamma\left(\frac{1}{2} + \frac{4}{c}\right)}{\Gamma\left(\frac{1}{2}\right)}. \tag{3.35}$$

The kurtosis κ is then obtained as:

$$\kappa_{MWD} = \Gamma\left(\frac{1}{2}\right) \cdot \frac{\Gamma\left(\frac{1}{2} + \frac{4}{c}\right)}{\left[\Gamma\left(\frac{1}{2} + \frac{2}{c}\right)\right]^2}. \tag{3.36}$$

Empirical studies suggest typical values of c in the range $[0.6, 0.9]$ for financial returns [295–297].

3.3.5 Truncated Lévy Distribution

We now turn our attention to the Truncated Lévy distribution (TLD), a model that differs significantly from the previous ones due to the absence of a closed-form analytical expression for its probability density function [299–301]. Instead, it is defined through its characteristic function. The form considered here, originally proposed by Koponen [302], is given by:

$$\varphi_{TLD}(k) = \exp\left\{ -\gamma \, \frac{(\lambda^2 + k^2)^{\alpha/2} \cos\left[\alpha \, \arctan\left(\frac{|k|}{\lambda}\right)\right] - \lambda^\alpha}{\cos\left(\frac{\pi\alpha}{2}\right)} \right\} \tag{3.37}$$

where $\gamma > 0$ is a scale parameter, $\lambda > 0$ is the truncation parameter, and $\alpha \in (0, 2]$, $\alpha \neq 1$, is the Lévy characteristic exponent. Unlike the previously discussed distributions which are characterized by two parameters, the TLD involves three.

This distribution exhibits power-law behaviour in an intermediate regime, followed by an exponential decay at large arguments. In the limiting case $\lambda \to 0$, the characteristic function of an α-stable Lévy distribution is recovered. Notably, for $\alpha = 1$ the distribution converges to the Cauchy distribution, while for $\alpha = 2$ it becomes Gaussian (see Sec. 1.4.1).

The corresponding probability density function is obtained via the inverse Fourier transform of the characteristic function:

$$p_{TLD}(x) = \frac{1}{2\,\pi} \int \mathrm{d}k \, e^{-i\,k\,x} \, \varphi_{TLD}(k). \tag{3.38}$$

No additional normalisation is required, since $\varphi_{TLD}(0) = 1$. The asymptotic behaviour of the distribution is:

$$p_{TLD}(x) \xrightarrow{|x|\to\infty} |x|^{-(1+\alpha)} \exp\{-\lambda |x|\},$$

showing a truncated power-law tail with exponential cut-off. Empirical fits to financial return data often suggest a characteristic exponent around $\alpha = 3/2$. In this case, the variance of the Truncated Lévy distribution takes the form (see [16] for derivation):

$$\mathrm{Var}_{TLD}\{x\} = \frac{3}{2\sqrt{2}} \frac{\gamma}{\sqrt{\lambda}}, \tag{3.39}$$

while the excess kurtosis is

$$k_{TLD} = \frac{1}{\sqrt{2}} \frac{1}{\gamma \lambda^{3/2}}. \tag{3.40}$$

Given the above characterisation of the distributions, we now turn to comparing the four models.

3.4 A Comparative Analysis of Models

Our aim is to implement a consistent comparison among the four non-Gaussian distributions introduced above. To ensure comparability, we consider standardised probability density functions, i.e., distributions with zero mean and unit variance. This approach reflects common practice in fitting financial data, where return series are normalised by subtracting the sample mean and dividing by the sample standard deviation.

Zero mean has already been assumed. As for unit variance, we fix the shape parameters based on empirical values reported in the literature for intra-day returns, which best capture the behaviour in the tails. For the two-parameter distributions, the chosen values are:

- $\nu = 3$ for the Student's t-distribution;
- $q = 1.5$ for the q-Gaussian;
- $c = 0.75$ for the Modified Weibull.

Once the shape parameters are fixed, the scale parameters are determined by imposing unit variance. This yields:

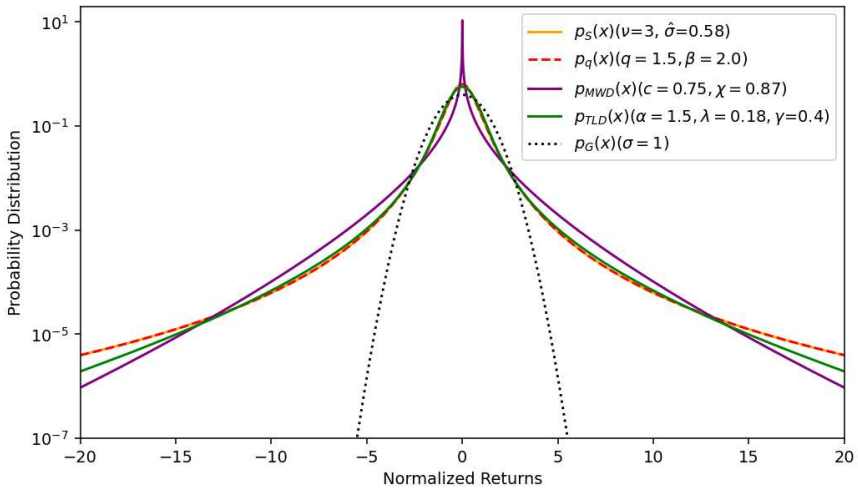

Figure 3.1: Comparison between standardised non-Gaussian return distributions and the Gaussian (dotted line) in the high-frequency limit over the interval $[-20, +20]$. The logarithmic scale on the y-axis highlights similarities among the non-Gaussian models, particularly in the tails, as well as their differences from the Gaussian.

- $\hat{\sigma} = 0.58$ for the Student's t (from Eq. (3.22));
- $\beta = 2.0$ for the q-Gaussian (from Eq. (3.27));
- $\chi = 0.87$ for the Modified Weibull (from Eq. (3.31)).

The strategy adopted for the Truncated Lévy distribution, which involves three parameters, follows a similar logic. Given empirical evidence reported in the literature [16], the characteristic exponent is set to $\alpha = 3/2$. The scale parameter γ is then obtained from Eq. (3.39) by requiring unit variance:

$$\gamma = 2\sqrt{2}\,\frac{\sqrt{\lambda}}{3}. \tag{3.41}$$

Finally, we determine the truncation parameter λ from values reported in high-frequency financial data, consistent with a specific λ–γ pairing reported in prior studies. The final parameter set used for comparison is:

- $\lambda = 0.18$;
- $\gamma = 0.4$.

As discussed earlier, the analytical form of the TLD in real space is not available in closed form. To address this, we compute the corresponding

probability density numerically by evaluating the inverse Fourier transform of the characteristic function $\varphi_{TLD}(k)$, as given in Eq. (3.38). This inversion is performed pointwise adopting numerical quadrature integration over a finite grid of real-space values. The resulting discrete set of values, which approximates the probability density function, is subsequently interpolated via cubic spline interpolation, a method employing piecewise third-degree polynomials [303]. This provides a smooth and continuous representation of the density, suitable for both graphical visualization and sampling.

Specifically, given a dataset $(x_i, y_i)_{i=0}^{n}$ with $a = x_0 < x_1 < \ldots < x_n = b$, the cubic spline function $S(x)$ satisfies the interpolation condition $S(x_i) = y_i$ and is twice continuously differentiable on $[a, b]$. On each subinterval $[x_i, xi + 1]$, the interpolant has the form:

$$S_i(x) = a_i + b_i\, x + c_i\, x^2 + d_i\, x^3.$$

The coefficients a_i, b_i, c_i, d_i are computed for each subinterval, ensuring global smoothness across the domain. In our implementation, we used the `Spline Interpolation` module from the `SciPy` Python package.

This procedure yields a graphical representation of the TLD, allowing for visual comparison with other models, as illustrated in Fig. 3.1.

Having established a consistent parameterisation for the non-Gaussian distributions under consideration, we proceed with a comparative analysis, presented in Fig. 3.1, focusing on the range $[-20, +20]$ which resembles realistic fluctuations observed in financial markets. To the best of our knowledge, no direct side-by-side evaluation of these models under consistent conditions has previously been reported.

By construction, all distributions are standardised to have zero mean and unit variance, ensuring that differences in shape (particularly in the tails) are attributable solely to their intrinsic functional forms rather than parameter scaling. Due to our consistent parameter selection, we observe that the four non-Gaussian models, each proposed as a plausible candidate for describing high-frequency financial returns, exhibit striking similarities in their leptokurtic profiles, both in the central region and in the tails. Notably, the closest resemblance is observed between the Student's t and the q-Gaussian distributions, in line with the analytical equivalence discussed in Eq. (3.29). These results suggest that, despite originating from distinct theoretical motivations, the models can yield consistent parametrisations of market dynamics when calibrated under uniform statistical constraints.

Finally, we remark that this consistency persists even when alternative parameter values are adopted and that empirically hold for daily returns. Specifically, we verified that analogous conclusions can be drawn when using

$\nu = 4$ for the Student's t, $q = 1.4$ for the q-Gaussian, $\lambda = 0.26$ (with $\alpha = 3/2$) for the Truncated Lévy Distribution, and $c = 0.85$ for the Modified Weibull Distribution. This robustness further supports the practical equivalence of the proposed models in capturing the statistical structure of financial return distributions.

3.5 Numerical Simulations

The initial comparison presented above highlights the qualitative similarities among the four non-Gaussian models, providing a foundation for the more quantitative analyses that follow. Specifically, our study now proceeds with the implementation of several Monte Carlo simulations aimed at investigating the statistical properties of returns and log-returns, as well as their time evolution. In this context, pseudo-random deviates (hereafter referred to simply as random, as explained in Sec. 1.7.1), are drawn from the different distributions and employed both as synthetic return data and as inputs for modelling non-Gaussian fluctuations in dynamical processes. Random numbers are generated within the interval $[-30, +30]$, chosen to replicate the magnitude of extreme fluctuations typically observed in empirical financial data. The generation of such deviates requires appropriate random number sampling techniques, as detailed below.

3.5.1 Random Number Generation

Modern computational environments such as `Python` provide built-in functionality for generating random deviates from a wide range of probability distributions. In particular, libraries like `numpy.random` and `scipy.stats` offer efficient and reliable methods to sample from standard distributions [304, 305]. Here, the term standard refers to the canonical form of a distribution, typically centred at zero and with unit scale. These functions enable the efficient generation of large samples of random variables for simulation, estimation, or hypothesis testing, without manually implementing the underlying transformation algorithms. However, in this work we focus on generalised, non-Gaussian distributions for which libraries do not provide built-in sampling routines. As a result, custom methods are required. Specifically, for the Student's t, q-Gaussian and Modified Weibull distributions (whose analytical forms are known) we implement the acceptance–rejection algorithm to generate independent random samples [32].

The acceptance-rejection sampling algorithm is a general-purpose method for generating random deviates from a target probability density function

$p(x)$ defined over a finite interval $[a, b]$. Let $p_{\max}$ denote the maximum value of the density in this interval. Given two independent and uniformly distributed random numbers $\xi_1, \xi_2 \in [0, 1]$, the algorithm proceeds as follows:

a) Generate a candidate point $x' \in [a, b]$ via the transformation $x' = a + \xi_1 (b - a)$;

b) Compute the value $p(x')$;

c) Define a vertical coordinate $y' \in [0, h]$ as $y' = \xi_2 h$, where $h \geq p_{max}$;

d) If $y' \leq p(x')$, accept x' as a sample from $p(x)$; otherwise, reject it;

e) Repeat the process until the desired number of sample is obtained.

The efficiency of the algorithm corresponds to the proportion of accepted samples relative to the total number of candidates generated. This corresponds to the ratio between the area under $p(x)$ and the area of the rectangle $[a, b] \times [0, h]$. A common improvement to this method involves adopting a proposal distribution $f(x)$ that better approximates $p(x)$, rather than a simple bounding box; this significantly increases the acceptance rate and improves the algorithm's overall efficiency [32, 306, 307]. Notably, care must be taken in choosing the interval $[a, b]$: if the interval is too large, the efficiency drops for heavy-tailed distributions; if too narrow, important tail events may be excluded, introducing bias into the statistical analysis.
In this context, acceptance–rejection sampling is employed in the analysis instead of inverse transform sampling for several reasons. While the latter is in principle applicable whenever the cumulative distribution function is known, its practical implementation would require the numerical inversion of the cumulative distribution function at each draw. For the non-Gaussian and heavy-tailed distributions considered in this work, such inversions are either not available in closed form or can be numerically unstable and computationally expensive, especially in the tails. Acceptance–rejection sampling avoids these issues by relying solely on the evaluation of the probability density function and provides a flexible and robust framework that can be applied uniformly across all distributions analysed.

For the q-Gaussian distribution (and, by virtue of the equivalence in Eq.(3.29), also for the Student's t), we can implement the generalised Box–Muller algorithm introduced by Thistleton, Marsh, Nelson, and Tsallis [308, 309]. This method generalises the classical Box–Muller algorithm[1] by replacing the nat-

[1]The standard Box–Muller algorithm is a widely adopted method for generating pairs of independent standard normal random variables $z_{1,2}$ from two independent uniform

ural logarithm with the q-logarithm:

$$\log_q(x) = \frac{x^{1-q} - 1}{1-q}, \qquad x > 0$$

and its inverse, the q-exponential:

$$e_q(x) = \begin{cases} [1 + (1-q)\,x]^{\frac{1}{(1-q)}}, & \text{if } 1 + (1-q)\,x \geq 0 \\ 0 & \text{otherwise.} \end{cases}$$

Given two independent uniform random variables $u_1, u_2 \in (0, 1]$, the algorithm generates a pair of random deviates:

$$\begin{aligned} z_1 &= \sqrt{-2\,\log_q u_1}\,\cos(2\,\pi\,u_2), \\ z_2 &= \sqrt{-2\,\log_q u_1}\,\sin(2\,\pi\,u_2), \end{aligned} \tag{3.42}$$

which follow a q-Gaussian distribution with:

$$q' = \frac{3\,q - 1}{q + 1}, \qquad \beta = \frac{1}{3 - q'}.$$

Note that q is the deformation parameter used in the logarithmic transformation (i.e., in $\log_q$ and e_q), while q' characterises the distribution of the generated deviates.

The q'-Gaussian obtained directly from the generalised Box–Muller transformation has a natural scale parameter given by

$$\beta_0 = \frac{1}{3 - q'}. \tag{3.43}$$

Substituting β_0 into the variance formula in Eq. (3.27) yields the variance of the raw deviates generated by the algorithm:

$$\mathrm{Var}_0 = \frac{3 - q'}{5 - 3q'}. \tag{3.44}$$

To obtain standardised deviates with unit variance, the output of the transformation should be rescaled:

$$z'_{1,2} = z_{1,2}\,\sqrt{\frac{5 - 3q'}{3 - q'}}. \tag{3.45}$$

---

random variables $u_{1,2}$ over the interval $[0, 1]$. It is based on the transformation: $z_1 = \sqrt{-2\ln u_1}\cos(2\pi u_2)$ and $z_2 = \sqrt{-2\ln u_1}\sin(2\pi u_2)$ [310].

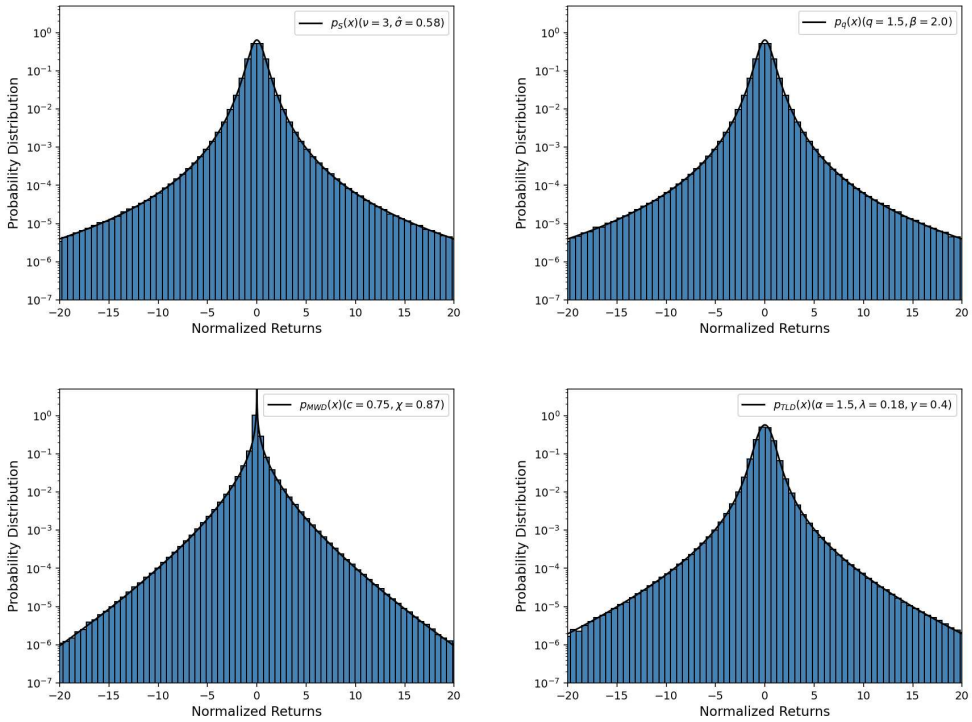

Figure 3.2: Logarithmic-scale comparison between the normalised non-Gaussian random number distributions (blue histograms) and their corresponding analytical standardised probability density functions (solid lines), over the interval $[-20, +20]$. From top-left, clockwise: Student's t, q-Gaussian, Truncated Lévy, Modified Weibull distribution. Parameters are selected as described in the main text.

This adjustment ensures that the generated samples are centred at zero and have variance exactly equal to one. Unlike the standard Box–Muller algorithm, the generalised version does not produce independent samples, as the joint distribution of z_1 and z_2 does not factorise into the product of marginals, indicating statistical dependence. Nonetheless, the two deviates remain uncorrelated and the method is efficient, yielding two sample per iteration. It therefore provides a practical and effective means for sampling from non-Gaussian distributions across a broad range of $q \in (-\infty, 3)$. The acceptance-rejection algorithm and the generalised Box-Mueller method have advantages and limitations. The latter is more efficient, because it supplies two random deviates starting from two uniform random numbers. In contrast, the acceptance-rejection efficiency is typically around 30%. However, as we sample within the finite interval $[-30; +30]$, the acceptance-rejection method remains preferable.

For the Modified Weibull distribution, the application of the acceptance-rejection algorithm is straightforward.

In contrast, for the Truncated Lévy distribution, this approach is not directly applicable, as its analytical form in real space is unknown. To overcome this, we first compute its characteristic function numerically over a finite grid, and then apply cubic spline interpolation to obtain a smooth approximation of the probability density $p_{TLD}(x)$ (as described above), which can subsequently be used as the target distribution in the acceptance–rejection algorithm. To validate this sampling procedure, we verify that in the limit $\lambda \to 0$ the method correctly recovers α-stable distributions by comparing the histogram of the generated random numbers with their corresponding analytical expressions [311,312]. In particular, we confirm that for $\alpha = 2$ the Gaussian distribution is recovered and for $\alpha = 1$ the Cauchy distribution, in agreement with theoretical expectations.

Given these generation techniques, for each non-Gaussian distribution we generate random deviates (denoted as ξ_{NG}), in order to obtain samples of size $M = 10^8$. Sampling is restricted to the interval $[-30; +30]$, in order to reflect realistic outliers observed in financial time series. It is worth noting that this truncation induces a slight distortion in the variance, which may deviate from unity. To correct for this, we rescale each dataset by dividing by its empirical standard deviation s, namely

$$\xi_{\mathrm{NG}} \to \frac{\xi_{\mathrm{NG}}}{s}. \tag{3.46}$$

This final normalisation ensures that all generated samples are standardised and consistent with the unit variance assumption, making them suitable as stylised models of financial returns. Figure 3.2 displays the resulting samples alongside the corresponding analytical probability density functions, thereby validating the accuracy of our numerical procedures.

3.5.2 Complementary Cumulative Distribution Functions

We have now at our disposal samples of size $M = 10^8$, consisting of standardised random deviates (with zero mean and unit variance) drawn from each of the four non-Gaussian distributions under consideration. These samples are defined over the range $[-30; +30]$, a choice motivated by the goal of reproducing realistic extreme fluctuations observed in financial time series.

Despite their rarity, tail events can induce outsized price movements, making them critically important for understanding and modelling financial

market dynamics. Accordingly, the study of tail behaviour is of particular interest. To this end, we examine the complementary cumulative distribution function (CCDF), as defined in Eq. (1.17), which quantifies the probability of observing a return greater than a given threshold. Specifically, we evaluate the CCDF for the four non-Gaussian models under investigation over the interval $[0.01, 20]$, corresponding to the empirical domain of impactful fluctuations in financial returns. For reference, the CCDF of the standard Gaussian distribution is also considered, given by:

$$\Phi_c(x) = \frac{1}{2}\,\mathrm{erfc}\left(\frac{x}{\sqrt{2}}\right), \tag{3.47}$$

where $\mathrm{erfc}(x)$ is the complementary error function, defined as:

$$\mathrm{erfc}(x) = 1 - \mathrm{erf}(x), \tag{3.48}$$

where

$$\mathrm{erf}(x) = \frac{2}{\sqrt{\pi}} \int_0^x e^{-t^2}\, dt. \tag{3.49}$$

The results, shown in Fig. 3.3, confirm the consistency among the proposed models and are in line with the insights presented earlier. Specifically, the Student's t and q-Gaussian distributions are known to be mathematically equivalent under the mapping introduced in Eq. (3.29). In particular, the shape parameter values $\nu = 3$ and $q = 1.5$ yield identical tail behaviour, both exhibiting power-law decay of the form:

$$p_{s,q}(x) \xrightarrow{x\to\infty} x^{-4}, \tag{3.50}$$

which implies a CCDF asymptotically governed by:

$$P_{C_{s,q}}(x) \xrightarrow{x\to\infty} x^{-3}. \tag{3.51}$$

This inverse cubic law is a well-documented empirical feature of high-frequency financial return distributions across a wide range of assets and markets [264]. While some empirical studies suggest deviations from this law under specific conditions [313,314], it nonetheless remains a broadly accepted benchmark for characterizing tail risk in financial time series.

Beyond this equivalence, both the Student's t and the Truncated Lévy Distribution (TLD) have been shown to offer comparably accurate fits to empirical return distributions. For instance, in the analysis of 30-minute log-returns of the S&P 500 index, the CCDF of a Student's t-distribution with $\nu \simeq 3$ was found to match that of a TLD with parameter $\alpha = 3/2$, yielding a best-fit parameter combination $\gamma^{2/3}\lambda = 0.096$ [16]. In the present work, we

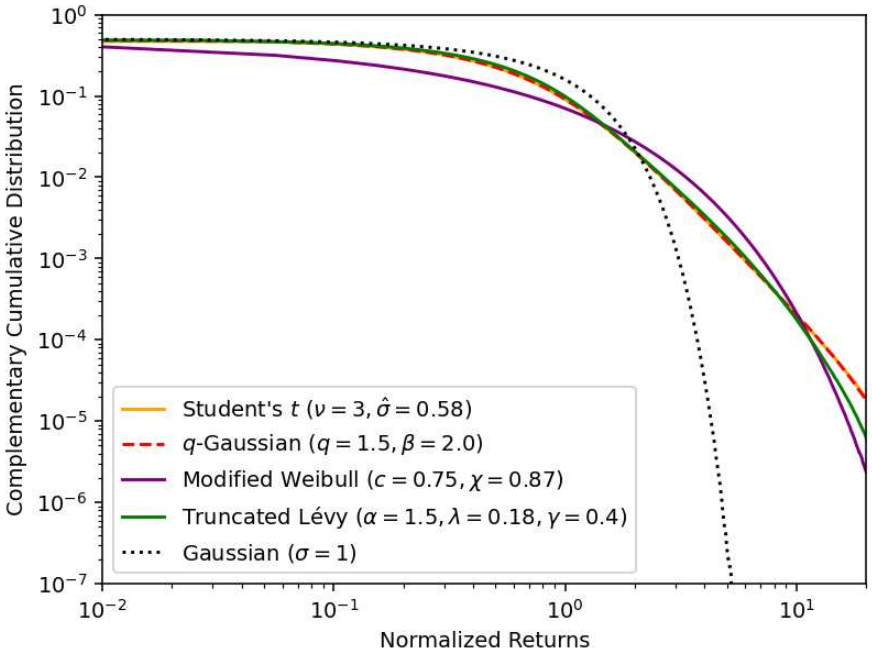

Figure 3.3: Log–log comparison between the complementary cumulative distribution functions of standardised non-Gaussian models and the Gaussian distribution (dotted line), computed for normalised returns in the range $[0.01, 20]$. All curves are obtained from $M = 10^8$ Monte Carlo samples generated over the interval $[-30, 30]$ with parameters chosen to ensure consistent comparison.

adopt the values $\lambda = 0.18$ and $\gamma = 0.4$, selected to ensure unit variance (see Eq. (3.39)), which result in $\gamma^{2/3}\lambda \simeq 0.098$, in excellent agreement with the empirical estimate.

Regarding the Modified Weibull Distribution, its CCDF with shape parameters in the range $c \in [0.6, 0.9]$ (with $c = 0.75$ adopted here) has been show to provide an accurate description of extreme fluctuations observed in traded currency exchange rates. Within this parameter range, distinguishing between stretched exponential and power-law decay becomes particularly challenging. As discussed in [296], this ambiguity arises because stretched exponential tails can approximate power-law behaviour in the limit $c \to 0$. The Modified Weibull is thus a flexible and effective model for financial applications.

Overall, this comparison based on the complementary cumulative distributions provides further evidence that the non-Gaussian models exam-

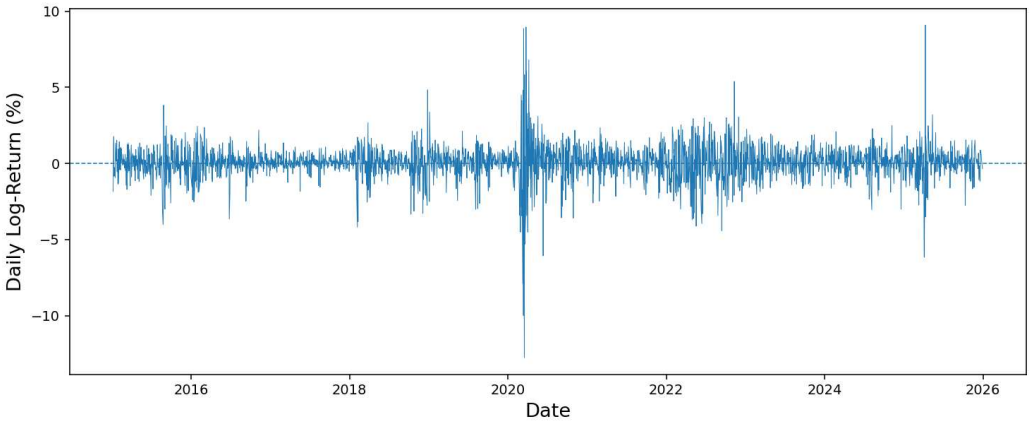

Figure 3.4: Daily log-returns of the S&P 500 index, expressed in percent, as a function of time from 02/01/2015 to 29/12/2025. Data source: Google Finance.

ined in this work offer compatible representations of extreme events in high-frequency financial data, provided that a consistent choice of parameters is adopted. The similarities observed, both in analytical behaviour and in simulated samples, reinforce the conclusion that these models are not alternatives, but rather represent equivalent parametrisations of the same underlying empirical regularities.

Before turning our attention to a more quantitative comparison of the four non-Gaussian models introduced in this chapter, we first illustrate their empirical relevance through an application to real financial data. Specifically, we consider daily returns of the S&P 500 index, spanning the period from 02/01/2015 to 29/12/2025 and obtained from Google Finance, which provide a canonical example of market dynamics characterised by large fluctuations and extreme events.
From the daily closing prices of the S&P 500 index, we construct the series of daily log-returns and rescale it to zero mean and unit variance. The resulting signal (Fig. 3.4) exhibits strongly intermittent behaviour, with long periods of relatively small fluctuations punctuated by rare but extreme events. A particularly prominent feature is observed around early 2020, where the onset of the COVID-19 pandemic produces an abrupt spike in the magnitude of returns, reflecting a sudden breakdown of market stability and a sharp increase in uncertainty. This episode exemplifies how exogenous shocks can trigger collective market responses with amplitudes far exceeding those expected under normal conditions, reinforcing the view of financial markets as complex systems operating far from equilibrium.

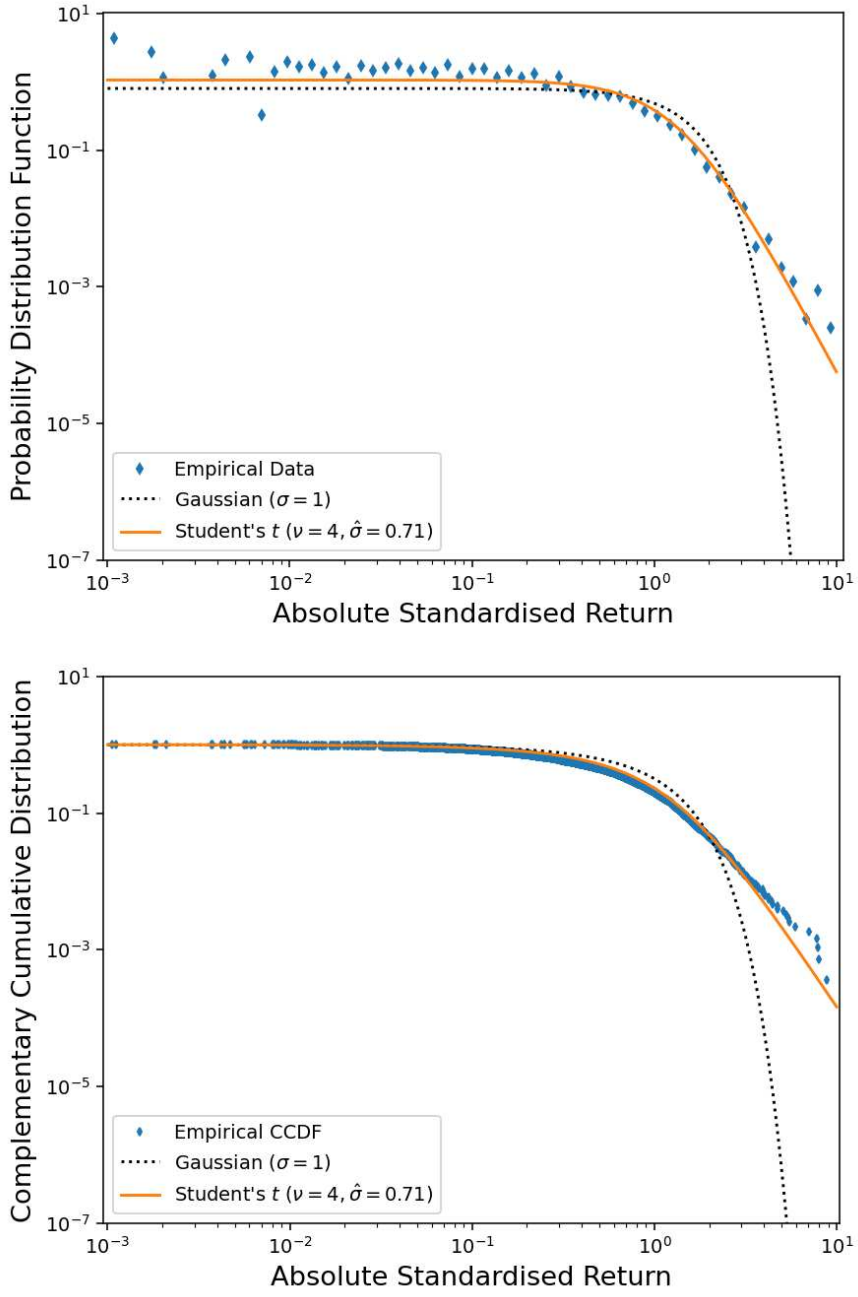

Figure 3.5: Empirical distribution of the absolute standardised daily log-returns of the S&P 500 index. Empirical data (blue markers) are compared with a Gaussian benchmark (dotted line) and with a standardised Student's t-distribution with $\nu = 4$ (orange line). Top panel: probability density function. Bottom panel: complementary cumulative distribution function.

To quantify these fluctuations, we estimate both the empirical probability density function and the complementary cumulative distribution function of the absolute standardised returns (Fig. 3.5). These empirical distributions are compared with a Gaussian benchmark and with a Student's t-distribution with fixed degrees of freedom $\nu = 4$, rescaled to unit variance to allow for a direct comparison with standardised data. It is important to note that no parameter fitting is performed and the characteristic parameters are those commonly adopted for daily returns (as discussed in Secs. 3.3.2 and 3.4). Moreover, the Student's t-distribution is chosen as a representative element of the four non-Gaussian models under consideration.

As expected, the Gaussian hypothesis fails dramatically to reproduce the empirical data (especially along the tails), underestimating the probability of large fluctuations by several orders of magnitude. In contrast, the Student's t-distribution captures the empirical decay of the tails with good accuracy over the entire observed range, providing a faithful description of extreme events. This agreement offers clear evidence that financial return fluctuations are intrinsically non-Gaussian and are more appropriately described by heavy-tailed, scale-free distributions, in line with the well-established stylised facts documented in econophysics. Given Fig. 3.3, it also provides additional support for the empirical equivalence of the four non-Gaussian models under study.

3.5.3 Kurtosis as a Comparative Metric

A further comparative measure is the excess kurtosis, as defined in Eq. (3.19), here evaluated computationally, theoretically and in light of empirical findings. As a benchmark, we consider the excess kurtosis reported for 30-minute returns of the S&P 500 index, estimated at $k_E \simeq 20$ in [16, 284, 315, 316]. More generally, for highly liquid markets and frequently traded assets (both at intra-day and daily timescales), excess kurtosis values typically fall in the range $k_E \in [10, 40]$.

From this point onward, we refer to excess kurtosis simply as kurtosis. For the Student's t distribution (Eq. (3.25)) and the q-Gaussian (Eq. (3.28)), the theoretical kurtosis is undefined because the fourth moment diverges for the shape parameters adopted here. However, since our analysis is restricted to the finite domain $[-30, +30]$, the computed sample kurtosis remains finite and informative.

In contrast, the kurtosis for the Modified Weibull (Eq. (3.32)) and the Truncated Lévy distributions (Eq. (3.40)) is finite by construction. The key distinction lies in the tail behaviour: MWD and TLD exhibit power-law

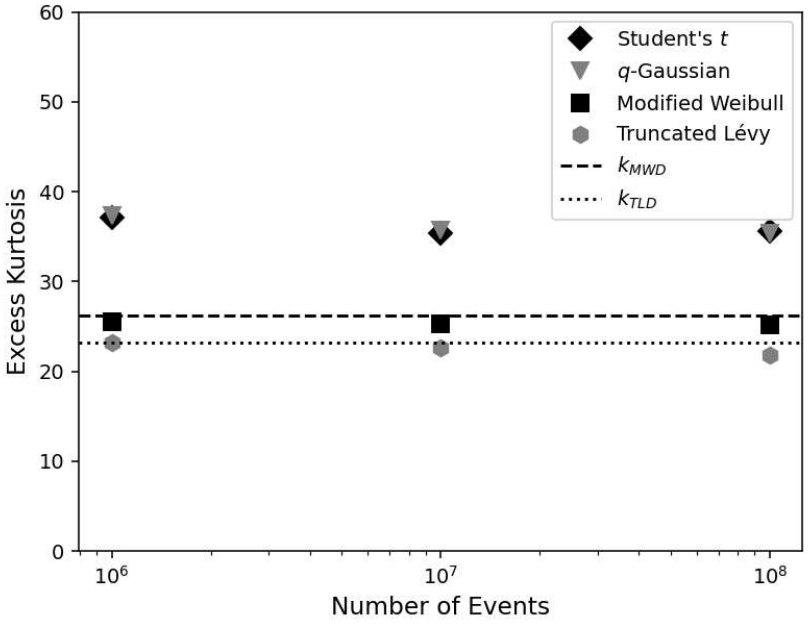

Figure 3.6: Excess kurtosis values (markers) for non-Gaussian return models as a function of the number of generated deviates $M = 10^6, 10^7, 10^8$ over the interval $[-30, +30]$. Dashed and shaded lines indicate theoretical predictions for the TLD ($\alpha = 3/2, \lambda = 0.18$) and MWD ($c = 0.75$), with $k_{\mathrm{TLD}} = 23.2$ and $k_{\mathrm{MWD}} = 26.2$. Student's t ($\nu = 3$) and q-Gaussian ($q = 1.5$) yield kurtosis around 35, in line with empirical values for high-frequency returns.

decay modulated by an exponential or stretched exponential cutoff, which guarantees the existence of all moments. On the other hand, the Student's t and q-Gaussian distributions follow a pure power-law decay, leading to the divergence of higher-order moments, such as kurtosis, when defined over the full real line. However, in practice, empirical financial return data span a finite domain. Within such a bounded range, these divergences are effectively regularised, allowing the sample kurtosis to remain finite and well-defined. A similar regularisation approach is applied to the variance in [317, 318].

Fig. 3.6 shows the estimated kurtosis for each distribution as a function of the Monte Carlo sample size $M = 10^6, 10^7, 10^8$, using random deviates generated over the interval $[-30, +30]$. The parameters are consistent with those adopted in Sec. 3.4 to model intra-day returns. The TLD and MWD closely reproduce their theoretical predictions, with $k_{\mathrm{TLD}} = 23.2$ and $k_{\mathrm{MWD}} = 26.2$, respectively. In contrast, the Student's t and q-Gaussian distributions yield higher values (approximately $k \simeq 35$), yet still fall within the empirical range

typically observed for high-frequency financial returns.

We also investigate a narrower domain, $[-20, +20]$, more representative of typical intra-day returns. Under this constraint, the kurtosis for the Student's t and q-Gaussian distributions falls to $k \simeq 20$, aligning more closely with the TLD and MWD. This result supports the argument that, under realistic bounds, all four models exhibit consistent statistical behaviour.

Importantly, the kurtosis for TLD and MWD remains stable across different sampling intervals, provided these are sufficiently broad. By contrast, the kurtosis for the Student's t and q-Gaussian is more sensitive to the truncation range. In fact, simulations performed adopting the generalised Box–Muller method to draw samples from these distributions (over an unbounded domain) reveal that the estimated kurtosis continues to grow with sample size, without evidence of convergence. This trend reflects the theoretical divergence of the fourth moment in these two heavy-tailed models. However, such divergence is not problematic in practice, as real-world financial returns are naturally bounded.

In conclusion, the observed convergence in kurtosis across the four non-Gaussian return models arises from a consistent and empirically grounded selection of shape and scale parameters. Since these parameters accurately reflect the statistical behaviour of large market fluctuations, the kurtosis analysis further supports the practical equivalence of the four distributions in modelling high-frequency financial returns.

3.5.4 Model Dynamics

Random variates play a central role in the numerical simulation of asset price dynamics and the associated log-returns. Recalling Eqs. (3.6) and (3.8), these continuous-time stochastic processes can be represented by substituting the Wiener increment $\mathrm{d}W(t)$ with its discrete counterpart $\Delta W(t)$, defined as

$$\Delta W(t) = \xi_G \sqrt{\Delta t}, \tag{3.52}$$

where ξ_G are independent standard Gaussian deviates, i.e., $\xi_G \sim \mathcal{N}(0, 1)$, and Δt denotes the time step. Within this framework, asset price and log-return trajectories can be simulated using the Euler–Maruyama method, a numerical scheme in which the stochastic process evolves in small discrete time increments [319]. Specifically, this scheme corresponds to the Euler discretisation of a Langevin equation with additive Gaussian white noise. As discussed in detail in [320], integrating the stochastic differential equation over a finite time step Δt yields a deterministic contribution proportional to

Δt and a stochastic contribution arising from the time integral of the noise term. For Gaussian white noise, this integral results in a random increment with zero mean and variance proportional to Δt, which can be written as $\sqrt{\Delta t}\,\xi_G$. This procedure directly leads to Eq. (3.52) and clarifies the origin and scaling of the noise term appearing in the discrete-time evolution.

In our case, the approximation over discrete time intervals $t_0, t_1, \ldots, t_N$ with uniform step size Δt leads to:

$$S(t+\Delta t) = S(t) + S(t)\,\mu\,\Delta t + S(t)\,\sigma\,\xi_G\,\sqrt{\Delta t}, \tag{3.53}$$

$$\ln S(t+\Delta t) = \ln S(t) + \left(\mu - \frac{\sigma^2}{2}\right)\Delta t + \sigma\,\xi_G\,\sqrt{\Delta t}. \tag{3.54}$$

In both cases, the accuracy of the numerical scheme depends on choosing a sufficiently small time step Δt to ensure that the discrete-time approximation remains faithful to the underlying continuous-time dynamics.

To generalise this approach, we substitute the Gaussian deviates ξ_G with non-Gaussian ones ξ_{NG}, sampled from the standardised distributions, as discussed earlier. As they have zero mean and unit variance by construction, this substitution preserves compatibility with the original Gaussian framework while introducing heavy-tailed characteristics into the dynamics.

The results of these simulations are shown in Fig. 3.7, where we compare asset price trajectories generated by the four non-Gaussian models under consideration (with the same choice of parameters as before) against the Gaussian case. As expected, non-Gaussian paths exhibit occasional sharp fluctuations, more closely mirroring the empirical behaviour observed in real financial markets.

This phenomenon is further highlighted in Fig. 3.8, which compares the log-return series generated from the Truncated Lévy distribution with that of a standard Gaussian process. While the TLD is shown as a representative case, analogous results hold for the other non-Gaussian distributions under scrutiny.

It is important to emphasise that, within this modelling framework, we implicitly assume that both returns and log-returns, although characterised by heavy-tailed distributions, still evolve according to a standard diffusion process. In particular, this means that the variance grows linearly in time, consistent with the behaviour of a Wiener process. Such an assumption allows for a direct and coherent comparison between Gaussian and non-Gaussian stochastic walks under the same diffusive dynamics. At the same time, empirical studies have consistently shown that returns themselves are well approximated as memoryless processes, since their autocorrelations decay rapidly and become negligible beyond very short time lags (a feature

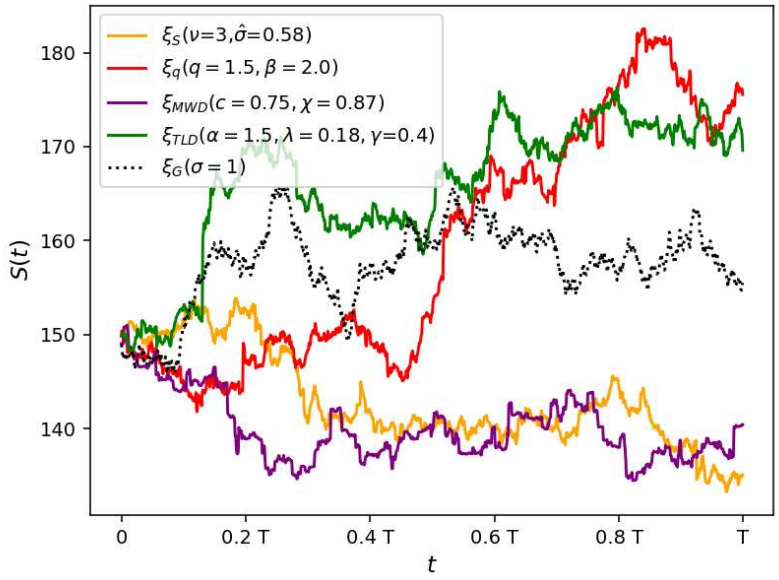

Figure 3.7: Simulated trajectories for asset price $S(t)$ based on non-Gaussian random walks, compared to the Gaussian benchmark (black dot). Parameters for the non-Gaussian distributions are those selected in the previous section to ensure consistency across models. Simulation settings are: $\Delta t = 10^{-3}$, total steps $N = 10^3$, leading to a final time horizon $T = N\Delta t = 1$, initial asset price $S(0) = 150$, drift $\mu = 0.01$, volatility $\sigma = 0.1$.

commonly attributed to the action of arbitrage mechanisms [272, 321]). Deviations from standard diffusive behaviour, including long-range dependence and anomalous scaling, are instead observed in functions of returns, such as volatility, absolute returns, or squared returns, which exhibit clustering and slowly decaying correlations over extended time horizons [16, 322]. In this sense, anomalous diffusion should be associated with volatility-related observables rather than with returns [323]. While the current implementation does not explicitly account for these effects, the framework can be extended in future work to incorporate such anomalous scaling properties.

3.5.5 Distributional Convergence

We now turn our attention to the convergence properties of random walks toward their corresponding limiting distributions. As introduced in Secs. 1.4.2 and 3.3.1, the additive process governing the log-return dynamics under Gaussian noise leads to a Gaussian distribution, while the corresponding multiplicative process for asset prices yields a log-normal distribution [324].

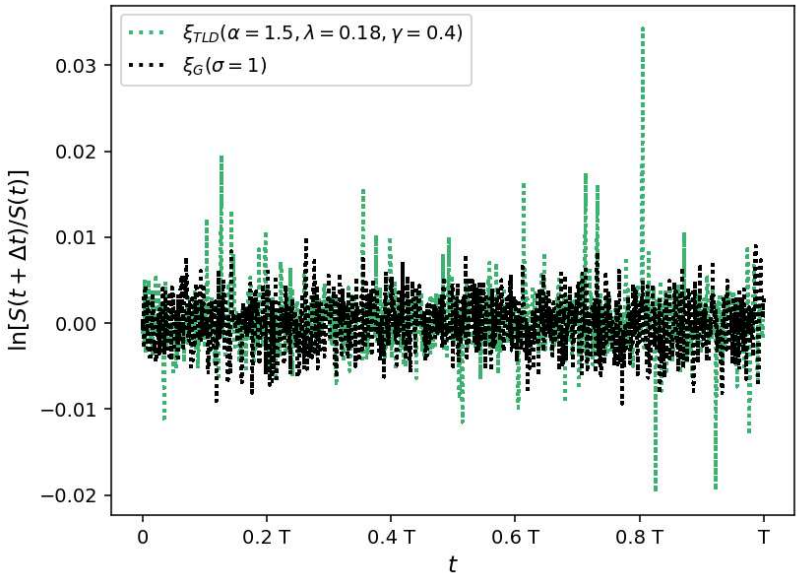

Figure 3.8: Log-return time series generated with Truncated Lévy noise (green trace) compared with the Gaussian one (black trace). The non-Gaussian model better captures the presence of extreme fluctuations.

These results extend naturally to non-Gaussian random deviates, provided that the variables are independent, identically distributed and possess finite variance, as well as finite higher-order moments. Indeed, under these assumptions, the additive and multiplicative versions of the Central Limit Theorems apply. Notably, one of the stylised facts presented in Sec. 3.2.1 is aggregational Gaussianity: after a characteristic time scale of approximately 10 to 16 trading days [283], the distributions of returns and log-returns gradually converge toward a Gaussian profile, despite being heavy-tailed over short horizons [325]. Crucially, the rate of convergence toward the limiting distributions is strongly influenced by the statistical moments of the underlying variables, whether Gaussian or non-Gaussian.

To investigate this convergence behaviour, we proceed under the following simplifying assumptions:

a) Among the four non-Gaussian models examined in this work, we choose the Truncated Lévy distribution as a representative case; nevertheless, the conclusions reached here extend to the other distributions under consideration.

b) We assume the random deviates to be i.i.d., although it is well known

that financial returns are typically uncorrelated but not independent (Sec. 3.2.1). This assumption can be relaxed through stochastic volatility models (see Sec. 3.7).

c) We fix the model parameters as

$$\mu = \frac{\sigma^2}{2}, \qquad \sigma = 1. \tag{3.55}$$

which simplifies the expressions for the limiting distributions and their moments without loss of generality.

Under these settings, Monte Carlo simulations are performed for the additive stochastic equation in discrete time, following:

$$\ln S(t + \Delta t) = \ln S(t) + \xi_i \sqrt{\Delta t} \tag{3.56}$$

and for its multiplicative counterpart according to:

$$S(t + \Delta t) = S(t) + \frac{1}{2} S(t)\, \Delta t + S(t)\, \xi_i \sqrt{\Delta t}, \tag{3.57}$$

where $i = G, NG$ denotes the Gaussian and non-Gaussian (TLD) deviates, respectively, drawn from standardised distributions. The same dataset adopted for ξ_i in the analyses of the complementary cumulative distribution function, kurtosis, and dynamical properties is employed here.

The simulation parameters are:

- Time step: $\Delta t = 10^{-3}$ to approximate the continuous-time limit;
- Number of random variables (i.e., iterations) added or multiplied: $N = 10, 10^2, 10^3$;
- Corresponding total times: $t = N\,\Delta t = 0.01,\ 0.1,\ 1$, where $T = 1$ represents the maximum simulation horizon (not tied to any physical time scale);
- Initial condition: $S(0) = 1$;
- Sample size (number of independent simulations): $M = 10^5$.

Considering the above settings, for the additive process in Eq. (3.56), the limiting distribution is the normal distribution:

$$p(\ln S_t, t) = \frac{1}{\sqrt{2\,\pi\, t}} \exp\left(-\frac{(\ln S_t - t/2)^2}{2\, t}\right), \tag{3.58}$$

where $S(t) = S_t$ and where the mean grows linearly with time while the variance scales as $\mathrm{Var}[\ln S_t] = t$.

As shown in Fig. 3.9, simulations under Gaussian noise rapidly converge to the limiting distribution, even for small N. In contrast, as represented in Fig. 3.10, convergence is significantly slower under TLD noise, requiring approximately $N \sim 10^2$ iterations for satisfactory agreement, particularly along the tails.

For the multiplicative process in Eq. (3.57), the limiting distribution is log-normal. Specifically, if $\ln S_t$ is normally distributed, then S_t follows:

$$p(S_t, t) = \frac{1}{S_t\sqrt{2\pi t}} \exp\left(-\frac{(\ln S_t)^2}{2\,t}\right), \tag{3.59}$$

which has strictly positive support and a peak that shifts and broadens over time due to the underlying stochastic exponential growth. The n-th moment is:

$$\mathbb{E}[(S_t)^n] = \exp\left[\frac{n^2 t}{2}\right], \tag{3.60}$$

demonstrating super-exponential growth and the increased likelihood of large deviations, especially under heavy-tailed noise. Simulation results are shown in Figs. 3.11 and 3.12.

It is worth emphasising that the numerical simulation of stochastic processes admitting rare events and extreme fluctuations is intrinsically limited by finite sampling. For heavy-tailed distributions, extreme events contribute significantly to tail statistics and higher-order moments, yet they occur with very low probability and may therefore not be observed even over long simulation horizons [326]. As a result, convergence to the asymptotic distribution can be exceedingly slow, already in the absence of temporal correlations. The presence of memory effects or strong temporal correlations further worsens this issue. In correlated dynamics, successive observations are not statistically independent, which effectively reduces the number of independent samples available and slows down convergence even further. This reduction in the effective sample size implies that rare events may lie well beyond the accessible numerical time scales, despite the apparent length of the simulated trajectories. From a statistical standpoint, this limitation can be expressed in terms of a characteristic time scale $\tau \sim \ln M$, where M denotes the number of independent realisations. This scaling reflects the fact that increasing the ensemble size only logarithmically improves the probability of observing extreme fluctuations. In our case ($M = 10^5$, $T = 1$), this threshold lies beyond the simulation horizon. These considerations highlight a fundamental

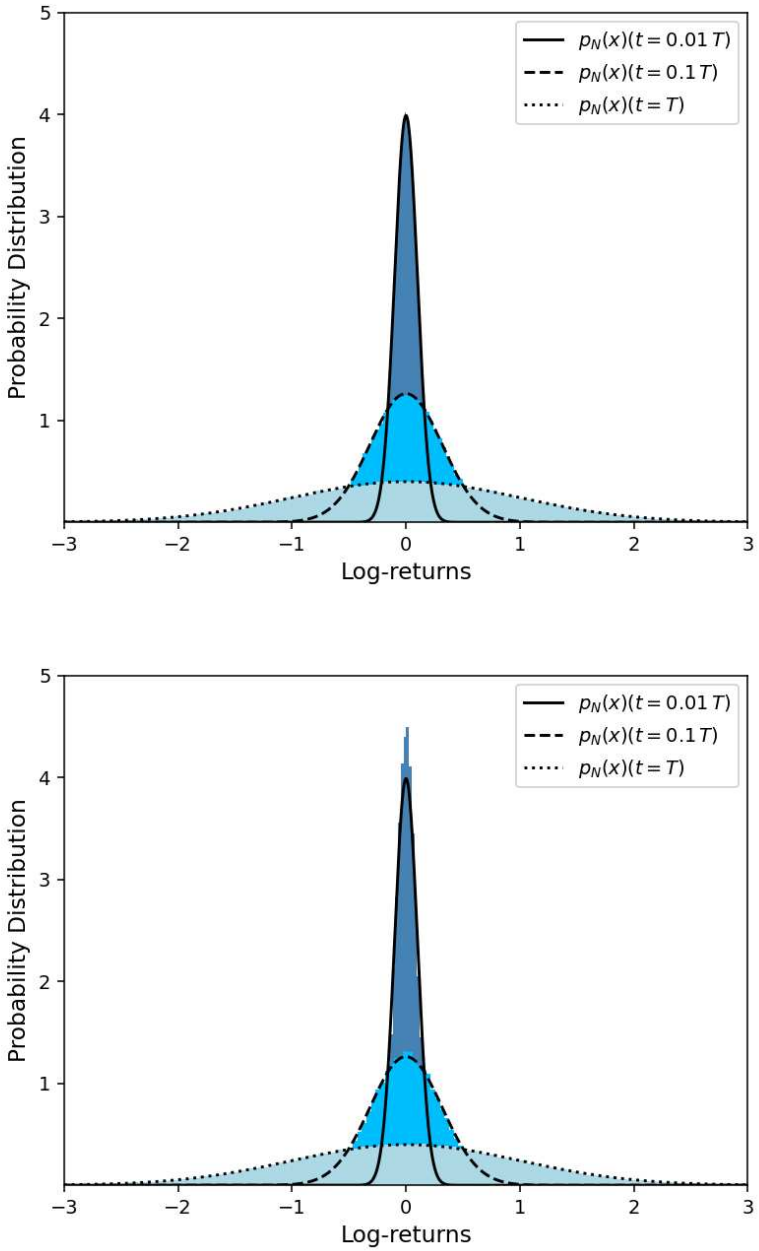

Figure 3.9: Normalised histograms for the sum of i.i.d. random variables drawn from a standardised normal (top) and Truncated Lévy (bottom) distribution, compared to the Gaussian limit (solid line). Simulation parameters are described in the text.

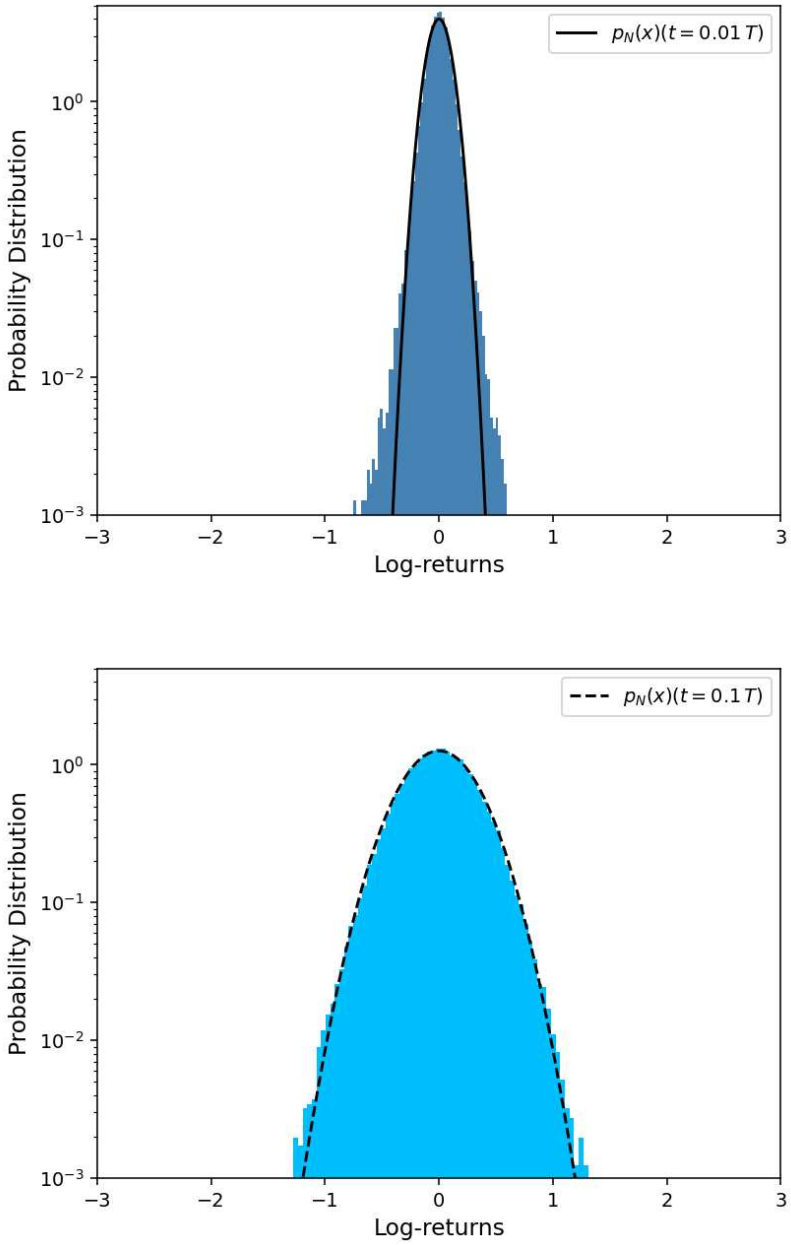

Figure 3.10: Normalised histograms for the sum of i.i.d. random variables drawn from Truncated Lévy distributions, compared to the Gaussian limit (solid line). As expected, non-Gaussian deviates require more iterations to achieve convergence, as seen in the different cases $N = 10$ (top) and $N = 10^2$ (bottom). Simulation parameters are described in the text.

constraint of numerical simulations and call for caution when interpreting results for strongly correlated systems or processes characterised by heavy tails and rare events.

In addition, our results are consistent with previous analyses showing that TLD converges toward a Gaussian distribution once the number of summed variables exceeds its kurtosis, a condition met here [16].

This convergence is also confirmed by kurtosis values, which approach the Gaussian value as the time horizon increases [327].

In summary, both Gaussian and non-Gaussian random walk dynamics converge to their respective limiting distributions, yet convergence is markedly slower in the presence of heavy-tailed noise. Notably, for TLD deviates, a visible agreement emerges only after $N \sim 100$ steps. To quantify this convergence, we invoke the Berry–Esseen theorem [328, 329], which provides an upper bound on the Kolmogorov–Smirnov distance between the considered model and theoretical Gaussian cumulative distribution functions. While the Central Limit Theorem ensures that the distribution of normalised sums of independent and identically distributed random variables approaches the Gaussian limit asymptotically, the Berry–Esseen theorem quantifies the rate, making it an essential tool in practical settings such as financial modelling and Monte Carlo simulations.

Convergence Rate: The Berry–Esseen Theorem

The Berry–Esseen theorem characterises how rapidly the convergence to limiting distributions occurs for finite sample sizes [16, 273, 330].

Given a sequence of independent and identically distributed (i.i.d.) random variables $X_1, X_2, \ldots, X_N$, each with:

- zero mean;
- finite variance σ^2;
- finite third absolute moment ρ,

consider the normalised sum:

$$\widehat{S}_N = \frac{X_1 + X_2 + \cdots + X_N}{\sigma\sqrt{N}}. \tag{3.61}$$

The Berry–Esseen theorem guarantees the existence of an universal constant $C > 0$ such that the Kolmogorov–Smirnov distance between the cumulative

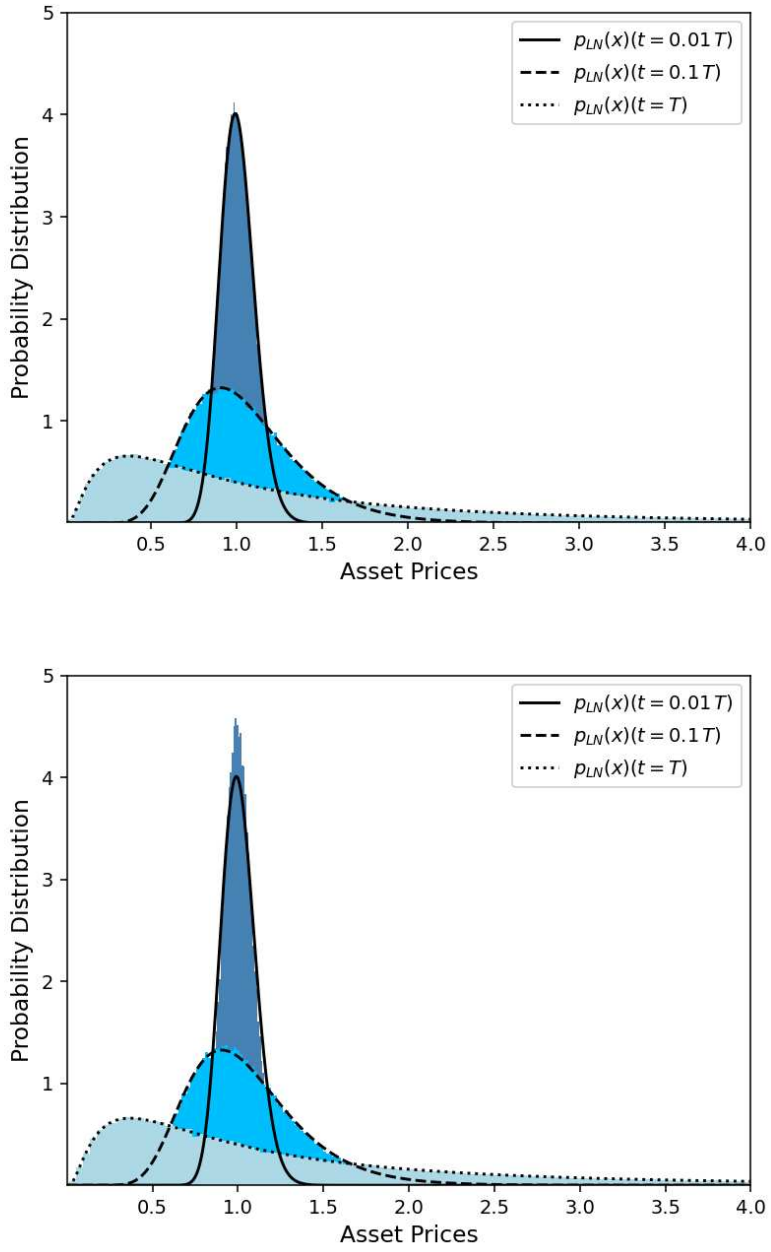

Figure 3.11: Normalised histograms for the multiplicative evolution of i.i.d. random variables drawn from a standardised normal (top) and Truncated Lévy (bottom) distributions, compared to the log-normal limit (solid line). Simulation parameters are reported in the text.

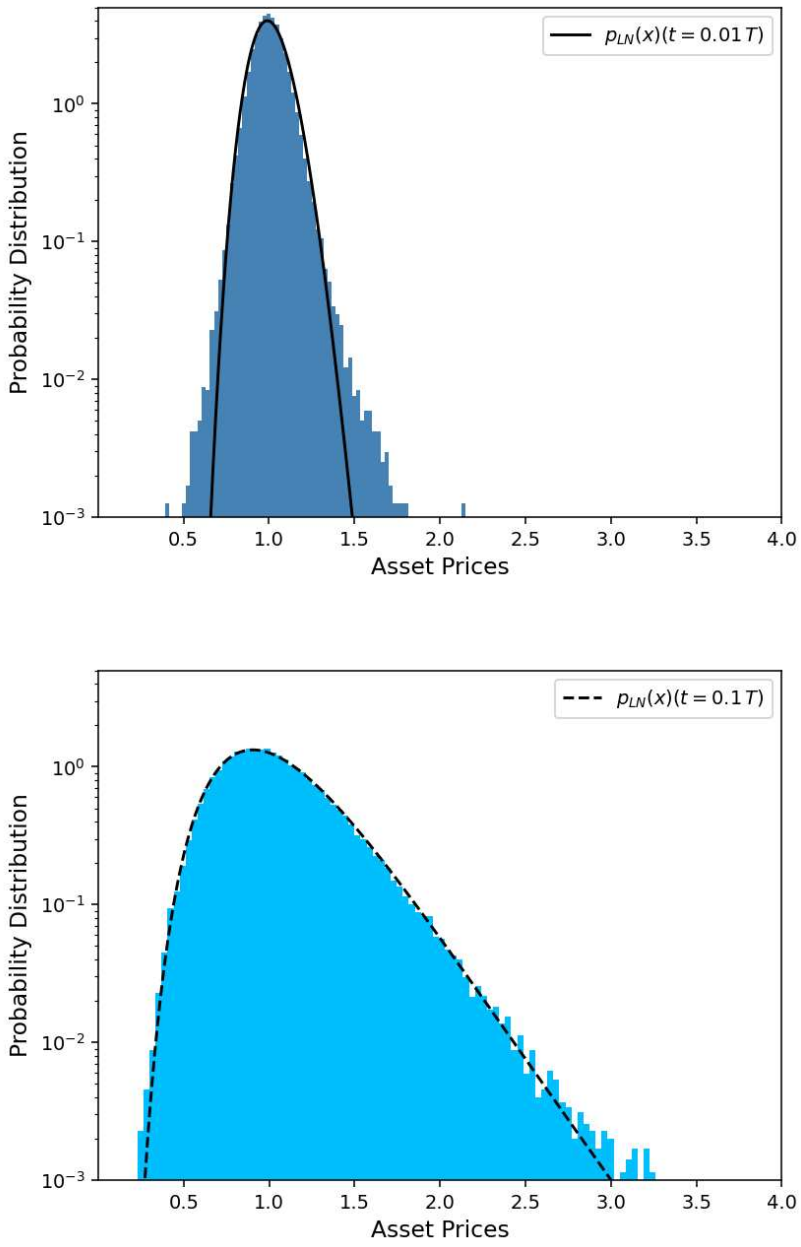

Figure 3.12: Normalised histograms for the multiplicative evolution of i.i.d. random variables drawn from Truncated Lévy distributions, compared to the log-normal limit (solid line). The heavy-tailed TLD deviates show slower convergence, as indicated for $N = 10$ (top) and $N = 10^2$ (bottom). Simulation parameters are reported in the text.

distribution function (CDF) $F_N(x)$ of $\widehat{S}_N$ and the standard normal CDF $\Phi(x)$ satisfies

$$\sup_x |F_N(x) - \Phi(x)| \leq C \frac{\beta_3}{\sqrt{N}}, \tag{3.62}$$

where $\beta_3 = \rho/\sigma^3$ is the standardised third absolute moment and

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-t^2/2}\, dt \tag{3.63}$$

is the standard normal cumulative distribution function.

In essence, this bound ensures that for any set of i.i.d. variables satisfying the above moment conditions, the distance between the distribution of their normalised sum and the Gaussian limit decreases at least as fast as $1/\sqrt{N}$.

Since its original formulation in the 1940s by Berry [331] and Esseen [332], several refinements have been proposed for the constant C. The sharpest estimate is due to Shevtsova [333], who demonstrated that:

$$C < 0.4748. \tag{3.64}$$

In our specific setup, the summed considered in the Berry–Esseen analysis is

$$X_k = \xi_{i,k}\sqrt{\Delta t}, \tag{3.65}$$

with $k = 1, 2, \ldots, N$ and the terms $\xi_{i,k}$ are i.i.d. random variables drawn from a standardised distribution, either Gaussian or TLD. These stochastic components correspond to the discrete-time increments used in the log-return trajectories, as defined in Eq. (3.56). Each X_k thus has zero mean, variance Δt and finite third absolute moment, satisfying the conditions of the Berry–Esseen theorem. Substituting these into Eq. (3.61), the normalised sum becomes

$$\widehat{S}_N = \frac{X_1 + X_2 + \cdots + X_N}{\sqrt{\Delta t}\sqrt{N}}. \tag{3.66}$$

To evaluate the convergence rate, we perform simulations for a number of iterations N in the range $[10, 10^3]$, with a fixed sample size of $M = 10^5$ and initial condition $S(0) = 1$, ensuring zero mean for the resulting log-returns. For each value of N, we compute the left-hand side of Eq. (3.62), representing the Kolmogorov–Smirnov distance between the empirical and the standard normal cumulative distribution functions. This is then compared to the theoretical upper bound on the right-hand side of the inequality, evaluated using the standardised third absolute moments of the variables under consideration:

$$\beta_{3,G} = \frac{2\sqrt{2}}{\sqrt{\pi}}, \quad \beta_{3,TLD} = 3.4. \tag{3.67}$$

The constant $C = 0.4748$, which provides the best known bound for the Berry–Esseen inequality, is adopted throughout.

As illustrated in Fig. 3.13, the simulation results align well with theoretical predictions. For Gaussian deviates (top panel), the empirical distance remains approximately constant at $\sim 3 \cdot 10^{-3}$ even for small N, reflecting the rapid convergence expected from Gaussian variables. In contrast, for TLD deviates (bottom panel), the Kolmogorov–Smirnov distance decreases steadily with N (from 10^{-2} to 10^{-3}), reaching values comparable to the Gaussian case only for $N \simeq 100$. These findings are consistent with the convergence patterns observed in Figs. 3.9, 3.10, 3.11 and 3.12.

Convergence: Analysis via Mean First Passage Time

The variable defined in Eq. (3.66) is further employed to study convergence via first-passage statistics. Specifically, we focus on the mean first passage time (MFPT), which quantifies the average time required for a random process to reach a specified threshold for the first time. MFPT is a fundamental metric in the study of stochastic dynamics across economic and social systems [334]. In finance, for instance, the MFPT can represent the expected time for an asset price to cross a critical level or for volatility to breach a risk threshold [335]. More broadly, in socio-economic models it is adopted to study tipping points, adoption thresholds in collective behaviour, or escape times from metastable equilibria. For processes driven by Gaussian fluctuations, such as classical Brownian motion, MFPTs can often be computed analytically. However, when non-Gaussian statistics are involved, passage dynamics become significantly more complex. These deviations are particularly relevant in empirical financial and social systems, where rare events and bursty dynamics are common [336].

In our study, the process under investigation is the normalised sum of i.i.d. random variables $\widehat{S}_N$, as defined in Eq. (3.66). The thresholds U are selected to represent both central and tail regions of the target distribution, enabling us to assess how quickly different regions of the distribution are explored under Gaussian and heavy-tailed dynamics. In the latter case, Truncated Lévy deviates are adopted as a representative model for non-Gaussian behaviour. The MFPT is computed as:

$$\tau = \mathbb{E}\left\{\min\left[t \geq 0 \ \text{ such that } \widehat{S}_N \geq U\right]\right\}, \tag{3.68}$$

i.e., the expected first time step t at which $\widehat{S}_N$ crosses the barrier U.

Simulations are performed according to the following setup:

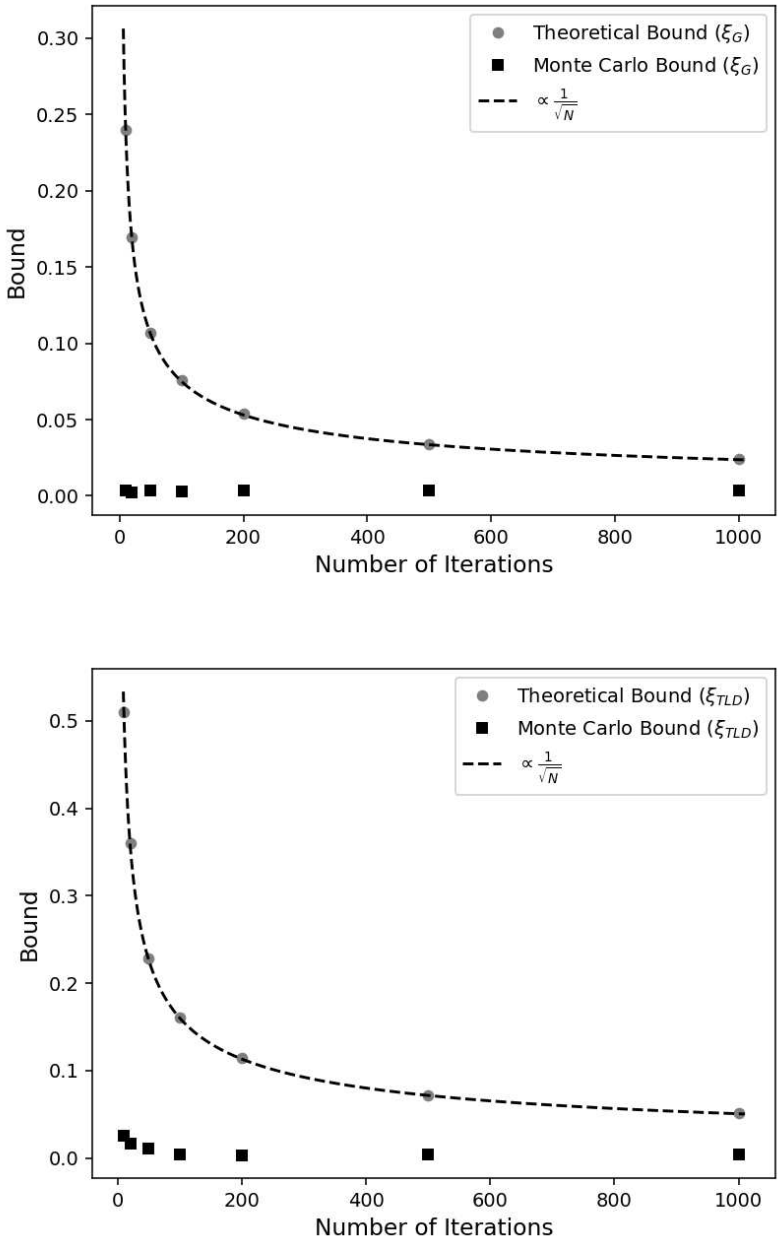

Figure 3.13: Convergence to the Gaussian limiting distribution evaluated via the Berry-Esseen theorem. Square markers represent the empirical Kolmogorov-Smirnov distance for Gaussian (top) and Truncated Lévy (bottom) deviates, while dot markers indicate the corresponding theoretical upper bound given by the Berry–Esseen inequality. The dashed line, proportional to $1/\sqrt{N}$, highlights the expected convergence rate. Simulation parameters are detailed in the main text.

Table 3.1: Mean First Passage Time τ_i and number of threshold-crossing events H_i for the barrier U, evaluated for the quantity $\widehat{S}_N$ adopting Gaussian ($i = G$) and Truncated Lévy ($i = \text{TLD}$) deviates. Simulation parameters are given in the text.

U	N	τ_G	H_G	τ_{TLD}	H_{TLD}
±1	10	0.0029	77793	0.0034	62547
±1	50	0.0061	93536	0.0089	85922
±1	100	0.0082	96363	0.0128	91541
±1	500	0.0139	99209	0.0257	97862
±1	1000	0.0167	99543	0.0323	98895
±3	10	0.0044	1469	0.0033	4166
±3	50	0.0147	2924	0.0099	5918
±3	100	0.0265	3570	0.0163	6747
±3	500	0.0982	5213	0.0611	8288
±3	1000	0.1817	5911	0.1147	8776

- MFPTs τ_G and τ_{TLD} are computed with standardised Gaussian and Truncated Lévy deviates, respectively, as inputs for $\widehat{S}_N$;

- The sample size is fixed at $M = 10^5$ trajectories;

- All trajectories are initialised with $\widehat{S}_N(0) = 0$;

- Symmetric thresholds are set at $U = \pm1, \pm3$, to probe both the central region and the tails;

- Iteration counts range from $N = 10$ to $N = 10^3$ corresponding to $t \in [0.01, 1]$, under the time convention introduced earlier;

- For each configuration, both the MFPT τ_i and the number of trajectories H_i that hit the barrier are recorded, with $i = G, TLD$.

The results, summarised in Table 3.1, clearly distinguish between Gaussian and non-Gaussian dynamics, as expected, and align with our earlier observations on convergence toward the limiting distribution. For thresholds at $U = \pm1$, Gaussian summands reach the barrier both more quickly and more frequently than their TLD counterparts, as indicated by lower values of τ_G and higher H_G. This behaviour is consistent with the peaked shape of Gaussian distributions compared to the broader leptokurtic profile of the TLD.

Conversely, for thresholds at $U = \pm3$, the TLD summands exhibit shorter MFPTs and more crossings than those generated by Gaussian variables. This is consistent with the heavy-tailed nature of the TLD distribution, which

increases the likelihood of large fluctuations and early extreme events. Consequently, such dynamics tend to oversample the tails of the limiting distribution during early stages, yielding a non-representative picture before convergence toward the Gaussian regime.

Overall, our findings show that the heavy-tailed stochastic processes under examination converge toward their respective Gaussian-related limits (normal for log-returns and log-normal for prices), under both additive and multiplicative dynamics, albeit slowly. This gradual transition mirrors empirical financial data, where the leptokurtic shape of return and log-return distributions evolves slowly toward Gaussianity and where the kurtosis plays a decisive role in setting the timescale of convergence. The Berry–Esseen analysis quantifies this convergence rate by establishing bounds on how rapidly the distribution of normalised sums approaches Gaussianity, while the mean first passage time results offer a complementary dynamical perspective by revealing how different regions of the distribution are explored over time. These insights are directly relevant to applications such as option pricing, where both the rate of convergence toward Gaussian behaviour and the finite-time over-representation of extreme events critically impact pricing value. This motivates the application explored in the following section.

3.6 Option Pricing

Building on the stochastic dynamics introduced earlier, we examine their application to the pricing of financial options. As introduced in Sec. 3.2, an option is a financial derivative that grants its holder the right (but not the obligation) to buy or sell a specified quantity of an underlying asset at a predetermined price (referred to as the strike price) on or before a specified expiration date (i.e., the maturity). The two fundamental types of options are defined as follows:

- A call option grants the right to buy the underlying asset;

- A put option grants the right to sell the underlying asset.

The party issuing the option (referred to as the writer) receives a premium as compensation for accepting the obligation to fulfill the contract, should the holder decide to exercise the option.

Based on the rules governing exercise, options are typically classified into the following categories:

- European-style: exercisable only at the expiration date;

- American-style: exercisable at any time up to and including maturity;

- Exotic options: featuring more intricate payoff structures or conditions, such as barrier features or path dependence (e.g., Asian options).

Options play a pivotal role in financial markets, serving as instruments for hedging, speculation, and risk management [337]. Their valuation is intrinsically tied to the stochastic modelling of the underlying asset's price dynamics and the principle of arbitrage-free pricing, which asserts that no riskless profit can be systematically extracted from market inefficiencies.

In the following, we first introduce the Black-Scholes framework, i.e., the canonical model for option pricing, and subsequently extend the analysis to incorporate non-Gaussian features in the underlying dynamics.

3.6.1 Black-Scholes Model

The valuation of financial derivatives (specifically, options) represents a central challenge in financial theory. This problem, which involves determining a fair price for a contract acceptable to both buyer and seller, was first rigorously solved in 1973 by Black and Scholes [338] and independently by Merton [339]. They introduced a quantitative framework that transformed modern finance, laying the foundation for contemporary option pricing theory. In recognition of their work, Scholes and Merton were awarded the Nobel Prize in Economic Sciences in 1997 "for a new method to determine the value of derivatives" (Black had passed away and was thus ineligible) [340]. Their methodology was groundbreaking: it allowed for the valuation of options throughout the life of the contract, from initiation to expiration.

The Black–Scholes model remains a cornerstone of financial modelling and is widely adopted in practice, despite its simplifying assumptions. In fact, the model is formulated under a set of idealised conditions:

a) Perfect and frictionless markets: assets can be traded without transaction costs, taxes, bid-ask spreads, or liquidity constraints;

b) No credit risk: all parties are assumed to honor their contractual obligations;

c) Market efficiency: asset prices fully and instantaneously reflect all available information;

d) Continuous trading and infinite divisibility: the underlying asset can be traded continuously in time and in arbitrarily small quantities;

e) No arbitrage: riskless profit opportunities cannot exist through simultaneous buying and selling;

f) No dividends: the underlying asset does not pay dividends over the option's lifetime, that is, periodic payments distributed by the company to its shareholders are assumed to be absent;

g) Geometric Brownian Motion (GBM): asset prices follow a stochastic process with constant drift μ and volatility σ (Eq. (3.6)):

$$\mathrm{d}S(t) = \mu\, S(t)\, \mathrm{d}t + \sigma\, S(t)\, \mathrm{d}W(t),$$

where $W(t)$ is a standard Wiener process.

In essence (see [34,38] for a detailed discussion), the Black–Scholes model proposes a continuously rebalanced portfolio composed of a fraction $\Delta(t)$ of the underlying asset and a cash position $\pi(t)$. The total value of this portfolio, $\mathcal{W}(t) = \pi(t) + \Delta(t)S(t)$, is designed to replicates the option price $O(S,t)$. By applying Itô's lemma (Sec. 1.3.3) and imposing the delta-hedging condition $\Delta = \partial O/\partial S$, the portfolio becomes riskless. This leads to a deterministic evolution governed by the celebrated Black–Scholes equation:

$$\frac{\partial O}{\partial t} + \frac{1}{2}\sigma^2 S^2 \frac{\partial^2 O}{\partial S^2} + rS\frac{\partial O}{\partial S} = rO, \tag{3.69}$$

where r is the constant risk-free interest rate. This parabolic partial differential equation describes the evolution of the option price under the no-arbitrage assumption, a constant risk-free interest rate r and volatility σ. Its solution, under the final condition specified by the payoff function, yields a closed-form expression for the price of European-style options.

For a call option (which is the one of interest here), the terminal payoff at maturity T is:

$$C(S_T, T) = \max\{S_T - X, 0\}, \tag{3.70}$$

where X is the strike price set by the contract.

Solving the Black–Scholes equation under this condition results in the well-known pricing formula:

$$C(S_t, t) = S_t\, \Phi(d_1) - X\, e^{-r(T-t)}\, \Phi(d_2), \tag{3.71}$$

with

$$\begin{aligned} d_1 &= \frac{\ln(S_t/X) + (r + \frac{\sigma^2}{2})(T-t)}{\sigma\sqrt{T-t}}, \\ d_2 &= d_1 - \sigma\sqrt{T-t}, \end{aligned} \tag{3.72}$$

and $\Phi(x)$ denoting the cumulative distribution function of the standard normal distribution.

Alternatively, the same result can be derived using a probabilistic approach based on the Feynman–Kac theorem, which links the solution of the Black–Scholes partial differential equation (Eq. (3.69)) to an expectation under a risk-neutral probability measure. In this formulation, the drift term μ is replaced by the risk-free rate r in the underlying asset dynamics:

$$\mu \to r.$$

Consequently, the value of an European call option at time t can be expressed as:

$$C(S_t, t) = e^{-r(T-t)} \int_0^\infty \mathrm{d}S_T \, p^*(S_T|S_t) \, \Pi(S_T), \tag{3.73}$$

where $p^*(S_T|S_t)$ denotes the risk-neutral log-normal conditional probability density (indicated by the asterisk), obtained from the GBM distribution in Eq. (3.15) under the substitution $\mu \to r$. The function $\Pi(S_T)$ represents the payoff at maturity, which, for an European call option, is given by Eq. (3.70).

A more detailed discussion of the Black–Scholes model and its derivation can be found in [38, 337].

Importantly, the volatility σ in the Black–Scholes formula is not directly observable and must be inferred from market option prices. This implied volatility plays a central role in practice, serving both as a pricing input and as a proxy for market expectations of future uncertainty. A striking empirical observation is that implied volatility is not constant across strike prices or maturities, as assumed in the Black–Scholes framework. Instead, it typically displays a smile or skew pattern, which is commonly referred to as the volatility smile or volatility surface. Such discrepancies have motivated the development of more sophisticated pricing models that relax the assumptions of GBM. Among them, stochastic volatility models have gained prominence, offering a more realistic representation of asset dynamics.

3.6.2 Option Pricing under Non-Gaussian Dynamics

The Black–Scholes framework provides an elegant and analytically tractable solution under Gaussian assumptions. Indeed, modelling asset prices via geometric Brownian motion allows for the direct application of Itô's lemma and, in addition, the assumption of costant parameters (namely the drift μ, the volatility σ and the risk-free rate r) further simplifies the mathematical treatment. However, empirical data from financial markets reveal substantial deviations from these idealised conditions and features such as volatility

clustering, heavy tails and abrupt jumps are commonly observed. In such settings, the assumption of normally distributed returns is no longer valid, and closed-form solutions to the pricing equation often become inaccessible.

To address these limitations, we adopt a Monte-Carlo approach to price European options under generalised stochastic processes with non-Gaussian noise [341]. Specifically, we retain the risk-neutral valuation framework introduced in Eq.(3.73), which generalises to:

$$C(S_t, t) = e^{-r(T-t)} \int_0^\infty \mathrm{d}S_T \, p^*_{\mathrm{NG}}(S_T|S_t) \, \Pi(S_T), \tag{3.74}$$

where $p^*_{NG}(S_T|S_t)$ denotes the risk-neutral conditional probability density associated with the selected non-Gaussian process. As in the Gaussian scenario, the drift term is replaced by $\mu \to r$, under the risk neutral measure. The payoff function $\Pi(S_T)$ is weighted by the appropriate probability distribution function, which, after repeated iterations of the stochastic dynamics in Eq. (3.53), converges to the non-Gaussian distribution determined by the underlying noise process.

The option price is then estimated as the discounted expected payoff over an ensemble of simulated trajectories:

$$C(S_t, t) \simeq e^{-r(T-t)} \, \langle \, \Pi(S_T) \, \rangle \, , \tag{3.75}$$

where the average is computed over M independent realizations. This strategy is particularly suitable for simulations and enables a direct and coherent comparison between Gaussian and non-Gaussian models under an unified pricing framework, following a common approach in both financial engineering and econophysics [342, 343].

Importantly, this methodology represents a considerable simplification. In general, extending option pricing to non-Gaussian settings requires deeper considerations of market completeness and the mathematical structure of the underlying stochastic process [344]. Nonetheless, we adopt the working hypothesis that pricing under non-Gaussian dynamics can be conducted using the same risk-neutral formalism as in the Gaussian case. A similar strategy is adopted in [345] for the Truncated Lévy scenario.

To implement this approach, we compute the fair value of European-style options under both Gaussian and non-Gaussian dynamics with the following setup:

a) The Monte Carlo estimator in Eq. (3.75) is employed, based on the dynamics specified in Eq. (3.53), where deviates ξ_i (with $i = G, NG$) are drawn from the random samples adopted in the previous analyses;

b) The stochastic dynamics is discretised with a time step $\Delta t = 10^{-3}$, appropriate for a continuous-time approximation;

c) The time to maturity is given by $T = N\,\Delta t$ and different values of N are explored to study the temporal evolution of option pricing behaviour;

d) Realistic input parameters are used to ensure practical relevance:

$$r = 0.01 \quad \sigma = 0.1, \quad S(0) = 150, \tag{3.76}$$

where the units of measure are $[\text{time}]^{-1}$ and $[\text{time}]^{-1/2}$ for r and σ, respectively, and a given currency for $S(0)$.

e) The number of Monte Carlo paths is fixed at $M = 10^5$;

f) Error bars represent the standard error of the mean, i.e., the sample standard deviation divided by $\sqrt{M}$. All results are reported with 1σ confidence intervals.

As an initial validation step, we verify that the Monte Carlo estimate under Gaussian dynamics accurately reproduces the analytical Black–Scholes result (Eq. (3.71)) for plain vanilla European call options. The analysis then proceeds with the study of both European call options and knock-out options under two distinct scenarios. The first focuses on short maturities, in order to highlight the impact of heavy tails at small values of N across the four non-Gaussian models under consideration. The second explores the effect of increasing maturity to examine convergence behaviour. In this latter case, the Truncated Lévy Distribution (TLD) is adopted as a representative example, although comparable results are obtained for the other non-Gaussian models.

We proceed with the analysis of European call options. The payoff function introduced in Eq. (3.75) is given by:

$$\Pi(S_T) = \max\{S_T - X, 0\}. \tag{3.77}$$

We consider at-the-money conditions by setting the strike price to $X = 150$, consistent with the initial asset price $S(0) = 150$. This choice yields the most informative comparison, given the parameter values set.

Fig. 3.14 presents the result of this analysis. The top panel shows that, at short maturities, all non-Gaussian models produce consistently lower option prices than the Gaussian benchmark, while remaining in close agreement with one another. Notably, the Black–Scholes model tends to systematically

overestimate the price of at-the-money options. This discrepancy suggests that the effective volatility for heavy-tailed models is lower, which is consistent with empirical findings.

The bottom panel illustrates that, as maturity increases, the pricing outputs of Gaussian and non-Gaussian models converge, as expected from the Multiplicative Central Limit Theorem (see Sec. 1.4.2 and 3.5.5), which ensures that price distributions approach log-normality over long time horizons. Accordingly, the implied volatility surface flattens with increasing maturity, as observed in real markets.

We also briefly examined in-the-money and out-of-the-money scenarios, finding that the Gaussian model tends to underestimate option prices relative to the non-Gaussian models. This again reflects the greater probability of extreme returns in heavy-tailed distributions, which increases the chance of large payoffs.

We next consider knock-out barrier options, defined by the following payoff:

$$\Pi(S_T) = \begin{cases} 0 & \text{if } \exists\, S_{t_i} > U \quad \text{for some } t_i \in [t, T], \\ \max\{S_T - X,\, 0\} & \text{otherwise,} \end{cases} \tag{3.78}$$

where U is the upper barrier and $X = 140$ is the strike price, here chosen to represent in-the-money conditions. These path-dependent contracts are voided if the asset price exceeds the barrier $U = 152$ at any point before maturity T.

The results, shown in Fig. 3.15, indicate that non-Gaussian models yield higher option values than the Gaussian counterpart at short maturities (top panel). This is likely due to the sharply peaked central regions of leptokurtic distributions, which dominate the valuation of barrier options at small N: since the payoff in such options is highly sensitive to the central portion of the return distribution, the influence of this central peak is pronounced. As maturity increases (bottom panel), the pricing gap narrows, again due to the convergence toward log-normal behaviour.

Across all test cases, the various non-Gaussian models yield consistent valuations, suggesting robustness and coherence of the pricing framework. This further supports our claim that the such models can serve as viable alternatives within a unified theoretical structure.

In conclusion, while the Black–Scholes model remains a milestone in financial theory, its reliance on Gaussian assumptions limits its ability to capture key empirical features of market dynamics. By incorporating non-Gaussian stochastic processes via Monte Carlo simulations, we introduce a flexible

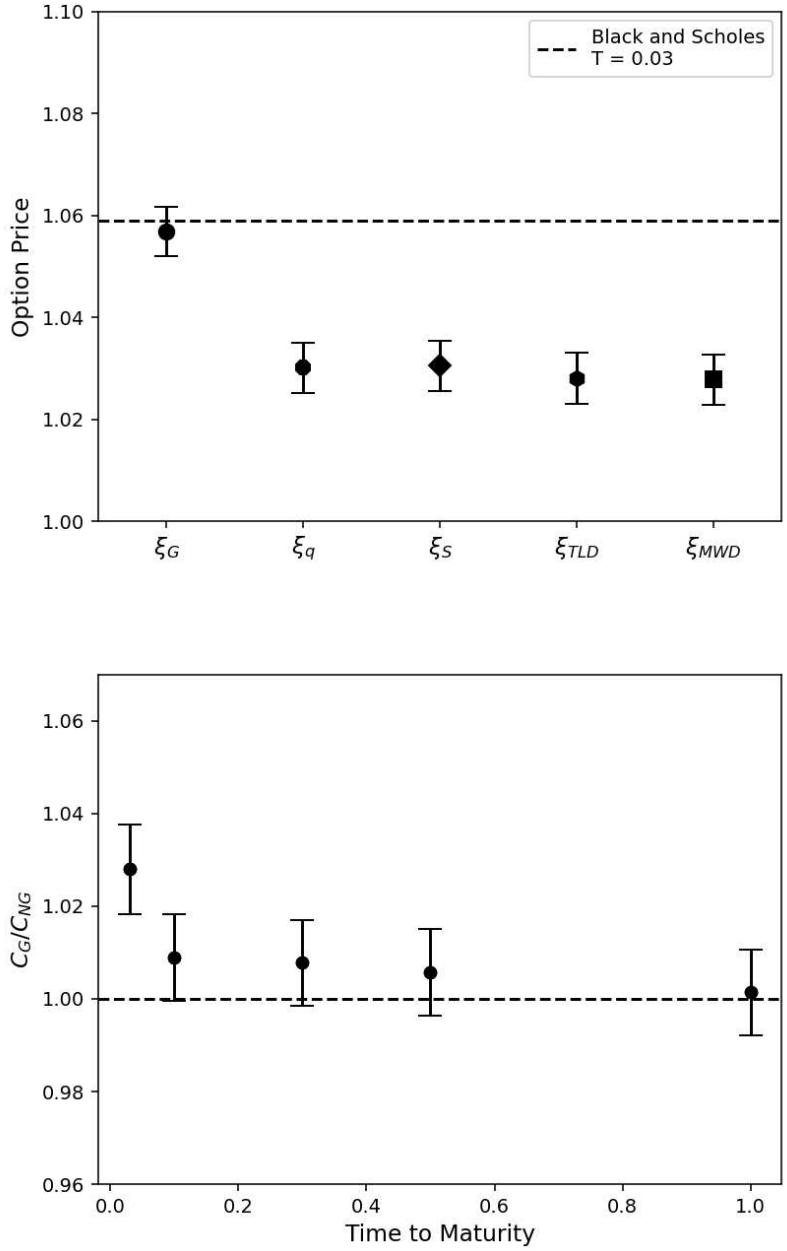

Figure 3.14: Top panel: Monte Carlo estimates of at-the-money European call option prices at short maturity ($T = 0.03$) under different stochastic return models. The dashed line represents the Black–Scholes price. Horizontal labels refer to the noise type: Gaussian (ξ_G), q-Gaussian (ξ_q), Student's t (ξ_S), Truncated Lévy (ξ_{TLD}), and Weibull (ξ_{MWD}). Bottom panel: Ratio between Gaussian C_G and Truncated Lévy C_{NG} option prices versus maturity (T from 0.03 to 1). Simulation parameters are specified in the main text.

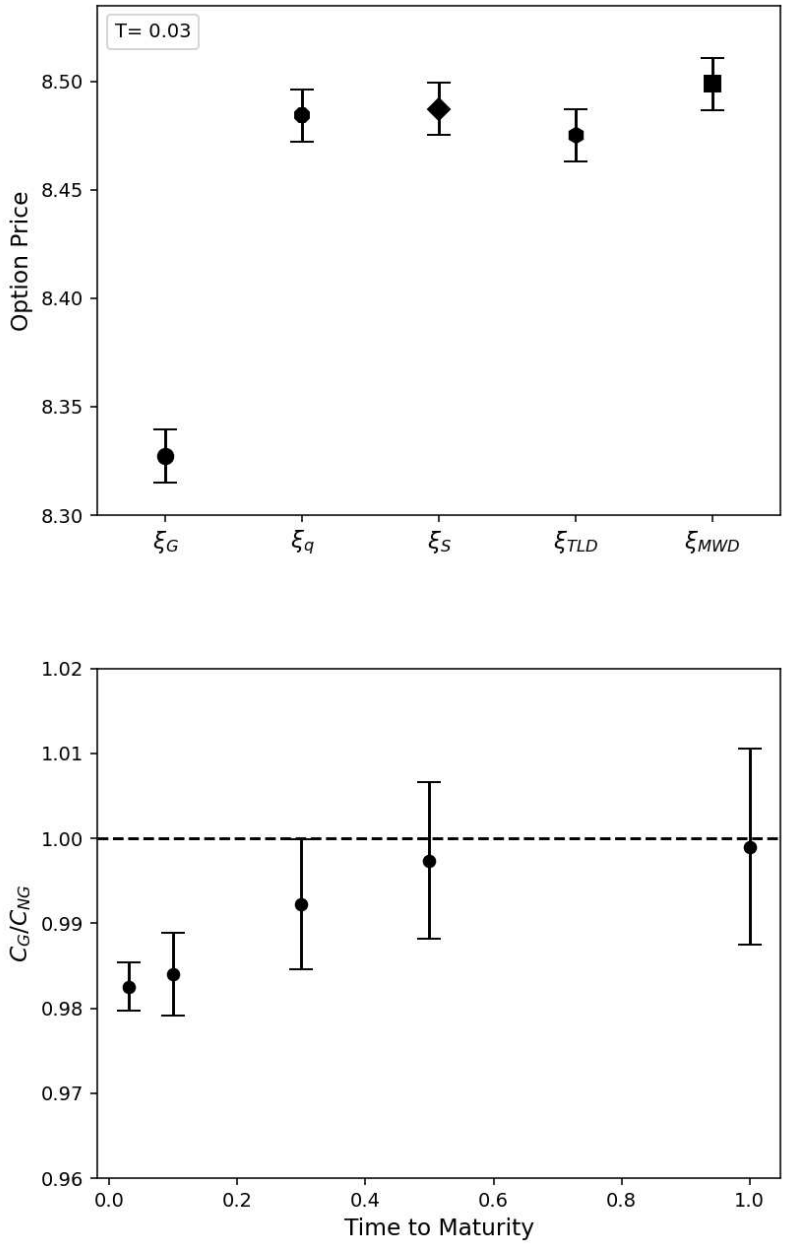

Figure 3.15: Same analysis as in Fig. 3.14, now applied to knock-out call options with $X = 140$ and barrier $U = 152$. Non-Gaussian models again yield systematically different valuations at short maturities compared to the Gaussian benchmark.

and empirically grounded extension of classical option pricing theory. This methodology, rooted in the probabilistic framework of statistical physics, remains computationally efficient and enables consistent pricing across a broad range of non-Gaussian models. It thus offers a powerful tool for analysing risk and valuation in complex, realistic market environments.

3.7 Summary and Discussion

The work presented in [3], and more extensively developed in this chapter, focuses on heavy-tailed distributions that have been proposed over the years as suitable models for describing high frequency returns and log-returns in financial markets. Specifically, we examine the Student's t, the q-Gaussian, the Modified Weibull and the Truncated Lévy distributions and compare them against the Gaussian benchmark. While the Gaussian model has historically served as the standard framework, empirical evidence has shown that non-Gaussian noise offers more accurate description of large fluctuations and rare events in asset price dynamics.

We introduce a consistent comparative analysis of the four aforementioned leptokurtic models. Central to this investigation is the careful calibration of the characteristic parameters for each distribution. To ensure fair comparability, all models are standardised to zero mean (reflecting the common practice in finance where the mean is negligible or subtracted) and unit variance. Shape parameters are selected based on empirical estimates for intra-day returns and the scale parameters are subsequently determined to satisfy the unit variance constraint.

A qualitative comparison, supported by graphical representations, reveals strong agreement among the four non-Gaussian models in their ability to capture heavy tails and large deviations from the mean. These findings suggest that the models should not be viewed as competing alternatives, but rather as compatible tools that yield consistent parametrisations of real market dynamics. In fact, each of them exhibits a substantial agreement in reproducing the universal statistical features of extreme deviations from Gaussian behaviour.

Our analysis then extends to a more quantitative comparison. To this aim, large samples of random deviates are generated for each distribution through the acceptance-rejection algorithms over a bounded range, chosen to reflect the typical domain of financial fluctuations observed in real markets. For the Truncated Lévy distributions (whose probability density function lacks a closed analytical form) a dedicated numerical procedure is implemented to enable sampling. Treating these synthetic datasets as proxies for

empirical return data, we observe the expected power-law decay in the tails through the analysis of the complementary cumulative distribution functions. A consistent behaviour across models emerges in capturing extreme outliers.

We then evaluate the kurtosis as a statistical indicator of tail heaviness. While formally undefined for the Student's t and the q-Gaussian distributions under certain parameter choices, the restriction to a bounded range (chosen to mimic realistic fluctuations) ensures that kurtosis remains finite and computable. The measured values are consistent across all the four models and in good agreement with the theoretical predictions available for the Truncated Lévy and Modified Weibull distributions, as well as with empirical data.

These generated random deviates are then employed as stochastic inputs in the dynamics of prices and log-returns, allowing us to study convergence to limiting distributions under both additive and multiplicative processes, given the assumption of independent and identically distributed variables. As expected, the Gaussian benchmark converges substantially faster than the heavy-tailed models, which show similar behaviour. This property is further examined via the Berry–Esseen theorem (which provides theoretical bounds on the convergence rate) and by examining mean first-passage times. The latter highlights distinct occupation patterns between leptokurtic and Gaussian scenarios. Together, these analyses reinforce our earlier conclusions about the practical equivalence of the four models and support the empirical stylised fact of aggregational Gaussianity.

The stochastic dynamics is subsequently applied to the problem of option pricing for both European plain vanilla and barrier options, under a simplified risk-neutral framework based on standard diffusion. Monte Carlo simulations are implemented to evaluate option prices under the four non-Gaussian models, enabling direct comparison with the Black–Scholes benchmark. Once again, the leptokurtic models under examination display strong mutual agreement and better adherence to empirical observations, particularly in the short-maturity regime. As maturity increases, convergence toward Gaussian pricing is observed, in line with the central-limit tendencies of heavy-tailed distributions at longer time scales.

Importantly, our findings show that the choice among these four non-Gaussian models does not materially affect the option pricing outcome, provided the parameters are correctly calibrated. This coherence across models supports their use within a unified statistical–dynamical framework for financial analysis.

These findings open several avenues for further research. One natural extension would be to test the robustness of these conclusions across different asset classes or market indices. The use of non-parametric goodness-of-fit

tests could help assess the empirical plausibility of each model in different contexts [16]. Additional statistical measures could also be introduced to quantify model compatibility.

Further refinement of the underlying dynamics represents another promising direction. Replacing standard Brownian motion with fast diffusion processes offer a more realistic description of high-frequency price evolution. For instance, the use of fractional Brownian motion introduces long-range dependence and self-similarity into the price dynamics, enabling explicit modelling of memory effects in financial time series, a feature that standard Brownian motion fails to capture [346]. However, such models pose notable theoretical challenges (particularly the breakdown of Itô calculus), necessitating alternative stochastic integration frameworks. In parallel, incorporating stochastic volatility models, such as those proposed in [345, 347–350], would relax the assumption of independence in returns and allow for a more accurate representation of volatility clustering, a well-documented empirical feature of financial time series. Such extensions would significantly influence the convergence properties of both asset price distributions and option valuations, thus meriting systematic investigation.

Lastly, this framework may be extended to incorporate risk measures, to assess the practical implications of non-Gaussian modelling in risk management and regulatory contexts [351, 352].

In summary, this work demonstrates that a variety of non-Gaussian stochastic models (namely the Student's t, q-Gaussian, Modified Weibull, and Truncated Lévy distributions) offer statistically consistent and empirically plausible descriptions of asset return dynamics. By standardising key parameters (zero mean and unit variance) and aligning shape and scale parameters with empirical evidence, we establish a unified framework in which these models lead to equivalent behaviour across both qualitative and quantitative analyses. Despite their different functional forms, all four models yield coherent outcomes in pricing analyses. As such, the choice among them becomes less a matter of model superiority and more a question of analytical convenience or computational tractability. This conclusion highlights the value of a principled and empirically grounded model selection process. Rather than seeking a single distribution, we argue for a flexible modelling approach that acknowledges the compatibility of multiple non-Gaussian frameworks under consistent calibration.

Chapter 4

Conclusions and Perspectives

This thesis explores two distinct yet methodologically connected frameworks: opinion dynamics in social systems and probabilistic modelling in financial markets. Both lines of research rely on statistical physics and highlight its ability to offer valuable insights into emergent collective behaviours arising from individual-level interactions. The original contributions presented here are based on three studies: two [1,2] addressing decision-making under social influence and one [3] investigating return distributions in financial markets. All three works combine Monte Carlo simulations with analytical methods. In the following, we first provide a synthesis of the main results in Section 4.1. Then, in Section 4.2, we discuss the limitations of the proposed approaches and possible directions for future research. We finally conclude in Section 4.3 by summarising the broader contributions of this work.

4.1 Synthesis of Results

The first two contributions (extensively described in Chap. 2) rely on agent-based models and focus on how social influence shapes individual decision-making and emergent collective behaviour. Indeed, human behaviour is profoundly shaped by social influence: individuals seek approval from peers and, when their opinions align with those of others, the resulting sense of social validation tends to reinforce existing beliefs; conversely, exposure to opposing views can induce discomfort, weakening attachment to prior convictions. As a consequence, people may adopt prevailing beliefs (even when suboptimal) by conforming to dominant norms or trends. This tendency can lead to the neglect of personal information, the suppression of independent judgement or even the misrecognition of talent. Our models offer deeper insight into the limitations and unintended consequences of imitative behaviour.

In the first study [1], we investigate the trade-off between social conformity and external information. We implement an agent-based model in which agents adopt one of two discrete states, representing binary opinions or behaviours (e.g., voting preferences or lifestyle choices such as smoking or dropping out of school). Each agent receives a payoff associated with their choice, which balances an external signal (representing a ground truth) and peer influence, modulated by a parameter that controls the weight of social influence. This payoff determines their decision-making via a stochastic Fermi rule, which incorporates a sensitivity parameter governing how strongly agents respond to payoff differences. The model captures realistic bounded rationality: although agents tend to favour strategies with higher payoffs, they may occasionally choose suboptimal options due to stochastic fluctuations.

We observe a sharp transition in the system's dynamics. When agents are highly sensitive to payoff differences and social influence is low, they tend to follow the external signal and switch states regularly, resulting in ergodic behaviour. However, as social influence increases, the system undergoes a critical transition: beyond a certain threshold, it becomes non-ergodic, and agents remain locked into a single collective state. This phenomenon (reminiscent of phase transitions in physics) demonstrates how strong peer influence can suppress responsiveness to objective evidence.

When memory is introduced into the dynamics, the external signal is smoothed through time-averaging, leading to earlier transitions and path dependence. This setting allows for a mean-field approximation, under which we analytically derive the critical threshold observed in simulations. In the high-sensitivity regime, memory induces irreversible flipping: once the time-averaged signal exceeds a threshold, agents switch state and remain there. This mechanism captures how prolonged exposure to biased information can lead to persistent behavioural change, relevant to phenomena such as political polarisation or cultural shifts. We characterise the flipping probability both analytically and numerically.

In addition, we test the robustness of the model by implementing the dynamics across different network topologies (Erdős–Rényi, Barabási–Albert and Stochastic Block Models). The results closely resemble those of the baseline scenarios, suggesting that the underlying mechanism is structurally stable and not dependent on the specific network microstructure.

Our model has potential applications across a range of empirical domains, from electoral dynamics to social habits and collective action. It helps explain sudden shifts in collective behaviour, such as financial crises, viral trends, rapid mobilisation, or abrupt changes in public sentiment. Relevant examples include school dropout, smoking, voting preferences, community crime

rates, and the observed stronger influence of peers over parents on students' academic performance. In financial markets, it captures herding behaviour, where individuals often prefer to follow other investors' decisions rather than rely on their own private information.

The second model [2] examines the interplay between imitation, randomness and rankings in decision-making. Rankings simplify complexity by ordering individuals or institutions according to performance, which can facilitate resource allocation but also introduces several potential drawbacks. One such limitation is that individuals often imitate high performers in the hope of replicating their success, even though pursuing their own strategy is frequently more effective. Our results reinforce earlier findings on this issue and offer further confirmation that top-ranked entities are not necessarily the most talented and that self-reinforcing dynamics can transform early advantages into long-lasting success.

We implement an agent-based model in which agents choose from a fixed set of actions, each associated with a payoff. The probability of remaining in a given field depends on both the agent's individual skill and the popularity of the action, reinforcing persistence in rewarding activities. Agents may also switch: either by imitating others (selecting actions in proportion to their popularity, thus capturing sensitivity to prestige and rankings) or by exploring a new field at random. A single parameter governs the balance between imitation and exploration, reflecting the weight agents assign to popularity versus novelty: low values favour imitation, leading agents to follow popular choices, while high values promote exploration and encourage independent decision-making.

To analyse the system's behaviour, we track several observables. When imitation dominates, aggregate payoff reaches its maximum; however, inequality (measured by Gini coefficient) rises sharply, as only a small number of agents contribute significantly to the total. The correlation between individual payoff and talent reveals that these successful agents are not necessarily the most skilled. The participation ratio further shows that the population converges on a single action, which is not necessarily beneficial for society as a whole. Thus, strong imitation leads to unequal and non-meritocratic outcomes, with talented agents failing to achieve proportional success. This effect originates in the early dynamics: some agents select actions aligned with their abilities and quickly attract imitators. Their early advantage amplifies into disproportionate popularity, reinforcing convergence and discouraging exploration. As a result, most agents remain locked into poorly matched actions. While imitation benefits a small minority, it generally produces suboptimal outcomes for the majority, including those with

higher intrinsic talent.

By contrast, when randomness dominates, the total payoff is lower but more evenly distributed, and agents are spread more uniformly across actions. The correlation between payoff and talent becomes positive, resulting in a more meritocratic society. This regime reflects a form of serendipity, where randomness fosters beneficial outcomes, and echoes findings from other domains (such as financial trading) where random strategies can outperform more sophisticated ones.

Between the imitation and randomness regimes, we observe sharp transitions across all measures. Such threshold behaviour is common in social systems driven by conformity, with parallels in social climbing, market dynamics, and the sudden emergence of traffic jams.

We test the model's robustness by adding learning from repeated actions, incorporating societal benefit and splitting the population into subgroups. In all frameworks, results remain qualitatively consistent with the baseline scenario, showing that the model, despite its simplicity, captures stable features of collective dynamics. Its abstraction from any single domain makes it broadly applicable to systems shaped by individual decision-making.

This finding aligns with previous studies and suggests that societies guided more by randomness than by social influence may achieve greater equality and meritocracy.

The third study [3] (presented in Chap. 3) focuses on financial markets as complex systems, where numerous interacting agents operate based on diverse beliefs and information. While this heterogeneity introduces unpredictability at the microscopic level, financial time series exhibit robust statistical regularities at larger scales.

In modelling these emergent phenomena, early approaches typically assumed that returns (i.e., relative price changes) followed a Gaussian distribution. However, empirical evidence consistently reveals heavy-tailed distributions, where extreme fluctuations occur far more frequently than predicted by the Gaussian benchmark. This has led to the proposal of various non-Gaussian models, such as the Student's t, q-Gaussian, Modified Weibull and Truncated Lévy distributions, though no consensus has emerged on a single best-fitting model.

To fill this gap in the literature, we provide a systematic comparison of the four leptokurtic models. To ensure consistency, all distributions are reported to zero mean and unit variance, following standard financial practice. Shape parameters are calibrated from empirical intra-day returns and scale parameters are then chosen to satisfy the variance constraint.

A qualitative comparison shows strong agreement among the four models

in capturing heavy tails and large deviations from the mean. For a quantitative test, we generate synthetic return data via acceptance–rejection sampling over a realistic bounded range. The complementary cumulative distributions confirm power-law decay in the tails. In addition, excess kurtosis analysis provides a consistent measure of tail heaviness. Although undefined for Student's t and q-Gaussian under some parameters, bounding the range makes the excess kurtosis finite and computable. Measured values are consistent across all four models and in agreement with theoretical predictions as well as empirical data.

We introduce the generated random deviates as stochastic inputs for price and log-return dynamics and study their convergence under both additive and multiplicative processes. As expected, Gaussian returns converge faster than heavy-tailed models, which behave similarly to one another. We further examine this aspect through the Berry–Esseen theorem and mean first-passage times, which reveal distinct occupation patterns between Gaussian and leptokurtic cases. These results reinforce the practical equivalence of the four models and support the empirical stylised fact of aggregational Gaussianity.

We apply the stochastic dynamics to option pricing, focusing on European plain vanilla and barrier options, within a simplified diffusion framework. Relying on Monte Carlo simulations, we compare prices across the four non-Gaussian models and the Black–Scholes benchmark. The leptokurtic models exhibiting strong mutual agreement and align more closely with empirical data, particularly at short maturities; as maturity increases, option prices gradually converge toward the Gaussian result, consistent with central limit theorems.

In summary, these findings demonstrate that the four non-Gaussian models yield consistent parametrisations of market dynamics. The Student's t, q-Gaussian, Modified Weibull and Truncated Lévy distributions each offer plausible description of asset returns and produce coherent results for both statistical properties and pricing. The choice among them is therefore less about identifying a superior model than about analytical convenience or computational efficiency. This highlights the value of a flexible, empirically grounded approach that treats multiple non-Gaussian frameworks as compatible under consistent calibration.

4.2 Limitations and Outlook

The results presented in this thesis are based on essential models designed to isolate and analyse collective behaviour in social and economic systems. This modelling approach reflects a deliberate choice: by focusing on large-

scale patterns rather than individual details, it becomes possible to uncover emergent regularities and traits arising from interactions among individuals. Accordingly, the emphasis is on a qualitative and mechanism-oriented approach, which naturally entails certain limitations.

A first important limitation concerns the level of abstraction at which agents are modelled. In the agent-based frameworks introduced in [1, 2], individuals are characterised by a restricted number of parameters and follow relatively simple behavioural rules. This simplification is intentional, as it enables the identification of clear dynamical regimes, such as phase transitions, path dependence, and ergodicity breaking. At the same time, it implies that specific preferences, agents own learning strategies, or peculiar cognitive processes are only partially captured. Indeed, real social systems are shaped by a wide range of factors, including history, expectations, strategic behaviour, and endogenous feedback from the environment, which are not explicitly represented in the present models.

Relatedly, the interaction structures among agents considered in this work are deliberately simplified. They are often implemented through homogeneous or fully connected topologies, and network structure is typically fixed and idealised. While this choice facilitates analytical investigation and highlights the role of social influence independently of detailed network effects, it limits the ability to capture phenomena arising from more elaborated social ties. Introducing explicit network heterogeneity, adaptive connections, or time-dependent measures therefore represents a natural extension of the present framework and may offer deeper insight into how changing environments and social structures shape collective outcomes, particularly close to critical regimes.

A further limitation concerns empirical validation. As discussed in Sec. 2.2.5, agent-based models do not rely on a single validation procedure; instead, validation must be tailored to the purpose and scope of the model. In addition, the frameworks developed in this thesis are intended as stylised representations aimed at uncovering generic mechanisms, rather than at reproducing a specific empirical dataset. While qualitative correspondences with observed phenomena can be identified, a specific calibration against data would require additional assumptions, data sources, and methodological tools. In particular, empirically estimating critical thresholds associated with phase transitions could help identify conditions under which real systems shift between flexibility, consensus, and polarisation. A systematic empirical validation of this kind is therefore beyond the scope of the present work and is left to future studies.

Given these limitations, several directions for further investigation are opened. A possible extension concerns the incorporation of more complex

interaction structures (such as heterogeneous or adaptive networks), to assess how social topology affects the impact of imitation, memory, and randomness on the dynamics. Moreover, both models could be empirically informed by calibrating critical thresholds of phase transitions through surveys, behavioural experiments, or analyses of digital data.

Similar considerations apply to the financial-market models discussed in [3]. Testing model robustness across different asset classes or market indices could help assess empirical plausibility in different contexts, while additional metrics and non-parametric goodness-of-fit tests may also be introduced to quantify model compatibility. A further natural direction concerns refining the underlying dynamics by incorporating faster diffusion processes or more realistic volatility dynamics, which could improve the description of high-frequency price evolution and may affect convergence behaviour. These extensions may further enable applications in areas such as risk management.

In general, understanding the limitations of simplified models is therefore essential not only for their correct interpretation, but also for identifying which features are most relevant when a more realistic setting is considered. In this sense, the present work should be seen as a foundation for further theoretical and empirical studies.

4.3 Conclusions

In summary, the works [1–3], situated at the intersection of statistical physics, opinion dynamics in social systems and probabilistic modelling in financial markets, contribute to the growing body of research investigating how individual decision-making gives rise to emergent collective phenomena. Within this interdisciplinary framework, our models not only reproduce established patterns of collective behaviour (such as consensus formation), but also identify novel conditions under which societies converge toward either accurate or distorted shared beliefs. This approach enables a deeper understanding of the mechanisms behind large-scale opinion shifts and highlights the structural parameters that govern their dynamics. Importantly, our findings indicate that imitation does not necessarily constitute the optimal strategy, neither for individuals, who may remain unrecognised despite genuine merit, nor for society, which may lose its ability to align with objective truth. Relying solely on others' opinions, driven by social pressure or perceived rankings, may ultimately undermine both talent and success. In contrast, decisions based on individual information (or even random exploration) can sometimes lead to fairer and more effective collective outcomes.

Our work extends naturally to financial markets, where many interacting agents with limited information give rise to statistical regularities in asset returns and pricing. In this context, by comparing non-Gaussian models under a unified calibration scheme, we show that they yield consistent and equivalent representations of market dynamics, both qualitatively and quantitatively.

Bibliography

[1] Federica De Domenico, Fabio Caccioli, and Giacomo Livan. Peer influence breaks ergodicity in an opinion dynamics model with external information. *Journal of Physics: Complexity*, 6(4):045007, 2025.

[2] Federica De Domenico, Fabio Caccioli, Giacomo Livan, Guido Montagna, and Oreste Nicrosini. Imitation versus serendipity in ranking dynamics. *Royal Society Open Science*, 11, 2024.

[3] Federica De Domenico, Giacomo Livan, Guido Montagna, and Oreste Nicrosini. Modeling and simulation of financial returns under non-Gaussian distributions. *Physica A: Statistical Mechanics and its Applications*, 622:128886, 2023.

[4] James Ladyman, James Lambert, and Karoline Wiesner. What is a complex system? *European Journal for Philosophy of Science*, 3:33–67, 2013.

[5] Giorgio Parisi. Complex systems: a physicist's viewpoint. *arXiv preprint cond-mat/0205297*, 2002.

[6] Richard Spencer-Smith. Reductionism and emergent properties. In *Proceedings of the Aristotelian Society*, pages 113–129. JSTOR, 1995.

[7] Philip W Anderson. More is different: broken symmetry and the nature of the hierarchical structure of science. *Science*, 177(4047):393–396, 1972.

[8] Jarosław Kwapień and Stanisław Drożdż. Physical approach to complex systems. *Physics Reports*, 515(3-4):115–226, 2012.

[9] Hidetoshi Nishimori and Gerardo Ortiz. *Elements of phase transitions and critical phenomena*. Oup Oxford, 2010.

[10] Mark EJ Newman. Complex systems: A survey. *arXiv preprint arXiv:1112.1440*, 2011.

[11] Alessandro Vespignani, Huaiyu Tian, Christopher Dye, James O Lloyd-Smith, Rosalind M Eggo, Munik Shrestha, Samuel V Scarpino, Bernardo Gutierrez, Moritz UG Kraemer, Joseph Wu, et al. Modelling COVID-19. *Nature Reviews Physics*, 2(6):279–281, 2020.

[12] NobelPrize.org. The Nobel Prize in Physics 2021, 2025. Accessed: 2025-06-30.

[13] Janusz A Hołyst, Krzysztof Kacperski, and Frank Schweitzer. Social impact models of opinion dynamics. *Annual Reviews Of Computational PhysicsIX*, pages 253–273, 2001.

[14] Herbert A Simon. The architecture of complexity. In *The Roots of Logistics*, pages 335–361. Springer, 2012.

[15] James R Lehane and AA Ekdale. Fractal analysis of graphoglyptid trace fossils. *Palaios*, 28(1):23–32, 2013.

[16] Jean-Philippe Bouchaud and Marc Potters. *Theory of financial risk and derivative pricing: from statistical physics to risk management*. Cambridge university press, 2003.

[17] Diego Vidaurre, Stephen M Smith, and Mark W Woolrich. Brain network dynamics are hierarchically organized in time. *Proceedings of the National Academy of Sciences*, 114(48):12827–12832, 2017.

[18] György Buzsáki. *Rhythms of the Brain*. Oxford university press, 2006.

[19] Gustavo Deco, Viktor K Jirsa, and Anthony R McIntosh. Emerging concepts for the dynamical organization of resting-state activity in the brain. *Nature reviews neuroscience*, 12(1):43–56, 2011.

[20] Olaf Sporns. *Networks of the Brain*. MIT press, 2016.

[21] Danielle S Bassett and Michael S Gazzaniga. Understanding complexity in the human brain. *Trends in cognitive sciences*, 15(5):200–209, 2011.

[22] Crawford S Holling. Understanding the complexity of economic, ecological, and social systems. *Ecosystems*, 4(5):390–405, 2001.

[23] Lev Davidovich Landau and Evgenii Mikhailovich Lifshitz. *Statistical Physics: Volume 5*, volume 5. Elsevier, 2013.

[24] Radu Balescu. *Statistical dynamics: matter out of equilibrium*. World Scientific, 1997.

[25] Eric Bertin. *Statistical Physics of Complex Systems*. Springer, 2021.

[26] Marco Baldovin, Giacomo Gradenigo, Angelo Vulpiani, and Nino Zanghì. On the foundations of statistical mechanics. *arXiv:2411.08709 [cond-mat.stat-mech]*, 2024.

[27] Erasmo Recami. L'articolo di Ettore Majorana su" Il valore delle Leggi Statistiche nella Fisica e nelle Scienze Sociali"(Ettore Majorana's article on" The value of Statistical Laws in Physics and in Social Sciences"). *arXiv preprint arXiv:0709.3537*, 2007.

[28] Claudio Castellano, Santo Fortunato, and Vittorio Loreto. Statistical physics of social dynamics. *Reviews of modern physics*, 81(2):591–646, 2009.

[29] Thomas C Schelling. Dynamic models of segregation. *Journal of mathematical sociology*, 1(2):143–186, 1971.

[30] Andrea De Martino and Matteo Marsili. Statistical mechanics of socioeconomic systems with heterogeneous agents. *Journal of Physics A: Mathematical and General*, 39(43):R465, 2006.

[31] Steven N Durlauf. How can statistical mechanics contribute to social science? *Proceedings of the national academy of sciences*, 96(19):10582–10584, 1999.

[32] Alberto Rotondi, Paolo Pedroni, and Antonio Pievatolo. *Probabilità, Statistica e Simulazione*. Springer, 2021.

[33] Grigorios A Pavliotis. Stochastic processes and applications. *Texts in applied mathematics*, 60, 2014.

[34] Crispin Gardiner. *Stochastic methods*, volume 4. Springer Berlin Heidelberg, 2009.

[35] Joseph L Doob. What is a stochastic process? *The American Mathematical Monthly*, 49(10):648–653, 1942.

[36] José Enrique Moyal. Stochastic processes and statistical physics. *Journal of the Royal Statistical Society. Series B (Methodological)*, 11(2):150–210, 1949.

[37] William Feller. Retracted chapter: On the theory of stochastic processes, with particular reference to applications. In *Selected Papers I*, pages 769–798. Springer, 2015.

[38] Wolfgang Paul and Jörg Baschnagel. Stochastic processes. *From Physics to Finance, Springer, Berlin*, 1999.

[39] Federica De Domenico. Modelling and Simulation of Non-Gaussian Financial Returns. Master's thesis, University of Pavia, 2022.

[40] Roberto Maiocchi. The case of Brownian motion. *The British journal for the history of science*, 23(3):257–283, 1990.

[41] Marian Von Smoluchowski. Zur kinetischen theorie der brownschen molekularbewegung und der suspensionen. *Annalen der physik*, 326(14):756–780, 1906.

[42] Paul Langevin et al. Sur la théorie du mouvement brownien. *CR Acad. Sci. Paris*, 146(530-533):530, 1908.

[43] Albert Einstein et al. On the motion of small particles suspended in liquids at rest required by the molecular-kinetic theory of heat. *Annalen der physik*, 17(549-560):208, 1905.

[44] Louis Bachelier. Théorie de la spéculation. In *Annales scientifiques de l'École normale supérieure*, volume 17, pages 21–86, 1900.

[45] Norbert Wiener. Differential-space. *Journal of Mathematics and Physics*, 2(1-4):131–174, 1923.

[46] Subrahmanyan Chandrasekhar. Stochastic problems in physics and astronomy. *Reviews of modern physics*, 15(1):1, 1943.

[47] Kiyosi Itô. On stochastic processes (I) Infinitely divisible laws of probability. In *Japanese journal of mathematics: transactions and abstracts*, volume 18, pages 261–301. The Mathematical Society of Japan, 1941.

[48] Kiyosi Ito, Kiyosi Itô, Kiyosi Itô, Japon Mathématicien, Kiyosi Itô, and Japan Mathematician. *On stochastic differential equations*, volume 4. American Mathematical Society New York, 1951.

[49] John P Nolan. *Univariate stable distributions*. Springer, 2020.

[50] Satya N Majumdar, Arnab Pal, and Grégory Schehr. Extreme value statistics of correlated random variables: a pedagogical review. *Physics Reports*, 840:1–32, 2020.

[51] Benoit Mandelbrot. The Pareto-Levy law and the distribution of income. *International economic review*, 1(2):79–106, 1960.

[52] Benoit Mandelbrot et al. The variation of certain speculative prices. *Journal of business*, 36(4):394, 1963.

[53] Mark EJ Newman. Power laws, Pareto distributions and zipf's law. *Contemporary physics*, 46(5):323–351, 2005.

[54] Constantino Tsallis. Nonextensive statistical mechanics, anomalous diffusion and central limit theorems. *Milan Journal of Mathematics*, 73(1):145–176, 2005.

[55] Constantino Tsallis. Lévy distributions. *Physics World*, 10(7):42, 1997.

[56] Lev Borisovič Klebanov. *Heavy tailed distributions*, volume 488. Matfyzpress Prague, 2003.

[57] Michael Mitzenmacher. A brief history of generative models for power law and lognormal distributions. *Internet mathematics*, 1(2):226–251, 2004.

[58] Herbert A Simon. On a class of skew distribution functions. *Biometrika*, 42(3/4):425–440, 1955.

[59] Derek J De Solla Price. Networks of scientific papers: The pattern of bibliographic references indicates the nature of the scientific research front. *Science*, 149(3683):510–515, 1965.

[60] Ye Sun, Fabio Caccioli, Xiancheng Li, and Giacomo Livan. The academic Great Gatsby Curve. *Journal of the Royal Society Interface*, 21(217):20240173, 2024.

[61] G Neukum and BA Ivanov. Crater size distributions and impact probabilities on Earth from lunar, terrestrial-planet, and asteroid cratering data. *Hazards due to Comets and Asteroids*, 359(1):359–416, 1994.

[62] RN Costa Filho, MP Almeida, JS Andrade, JE Moreira, et al. Scaling behavior in a proportional voting process. *Physical Review E*, 60(1):1067, 1999.

[63] Vilfredo Pareto. *Cours d'économie politique*. F. Rouge, Lausanne, 1896. Vol. I & II.

[64] Luigi Amoroso. Vilfredo Pareto. *Econometrica: Journal of the Econometric Society*, pages 1–21, 1938.

[65] Rosie Dunford, Quanrong Su, Ekraj Tamang, and Abigail Wintour. The Pareto principle. *The Plymouth Student Scientist*, 7(1):140–148, 2014.

[66] Oren S. Klass, Ofer Biham, Moshe Levy, Ofer Malcai, and Sorin Solomon. The Forbes 400 and the Pareto wealth distribution. *Economics Letters*, 90(2):290–295, 2006.

[67] Joseph L. Gastwirth. The estimation of the Lorenz Curve and Gini Index. *The Review of Economics and Statistics*, 54(3):306–316, 1972.

[68] Simone Pellegrino et al. The Gini coefficient: its origins. 2020.

[69] Maxim Pinkovskiy and Xavier Sala-i Martin. Parametric estimations of the world distribution of income. Technical report, National Bureau of Economic Research, 2009.

[70] Albert-László Barabási and Réka Albert. Emergence of scaling in random networks. *science*, 286(5439):509–512, 1999.

[71] Mark EJ Newman. Clustering and preferential attachment in growing networks. *Physical review E*, 64(2):025102, 2001.

[72] Robert K Merton. The Matthew effect in science: The reward and communication systems of science are considered. *Science*, 159(3810):56–63, 1968.

[73] Thomas Ising, Reinhard Folk, Ralph Kenna, Bertrand Berche, and Yurij Holovatch. The fate of Ernst Ising and the fate of his model. *arXiv preprint arXiv:1706.01764*, 2017.

[74] Richard IG Hughes. The Ising model, computer simulation, and universal physics. *Ideas In Context*, 52:97–145, 1999.

[75] Stephen G Brush. History of the Lenz-Ising model. *Reviews of modern physics*, 39(4):883, 1967.

[76] Barry M McCoy and Jean-Marie Maillard. The importance of the Ising model. *Progress of Theoretical Physics*, 127(5):791–817, 2012.

[77] Adam Lipowski. Ising model: Recent developments and exotic applications. *Entropy*, 24(12):1834, 2022.

[78] A Jiménez, KF Tiampo, and AM Posadas. An Ising model for earthquake dynamics. *Nonlinear Processes in Geophysics*, 14(1):5–15, 2007.

[79] Didier Sornette. Physics and financial economics (1776–2014): puzzles, Ising and agent-based models. *Reports on progress in physics*, 77(6):062001, 2014.

[80] Morikazu Nakamura, Kohei Kaneshima, and Takeo Yoshida. Petri net modeling for Ising model formulation in quantum annealing. *Applied Sciences*, 11(16):7574, 2021.

[81] Mauricio A Valle, Jaime F Lavín, and Nicolás S Magner. Equity market description under high and low volatility regimes using maximum entropy pairwise distribution. *Entropy*, 23(10):1307, 2021.

[82] W-X Zhou and Didier Sornette. Self-organizing Ising model of financial markets. *The European Physical Journal B*, 55:175–181, 2007.

[83] Richard Fitzpatrick. Introduction to computational physics. 2022.

[84] Julia M Yeomans. *Statistical mechanics of phase transitions*. Clarendon Press, 1992.

[85] Ginestra Bianconi. Mean field solution of the Ising model on a Barabási–Albert network. *Physics Letters A*, 303(2-3):166–168, 2002.

[86] Steven N Durlauf. Statistical mechanics approaches to socioeconomic behavior. In *The economy as an evolving complex system II*, pages 81–104. CRC Press, 2018.

[87] Serge Galam. What is sociophysics about? In *Sociophysics: A Physicist's Modeling of Psycho-political Phenomena*, pages 3–19. Springer, 2011.

[88] Parongama Sen and Bikas K Chakrabarti. *Sociophysics: an introduction*. OUP Oxford, 2014.

[89] Konstantin Klemm, Víctor M Eguíluz, Raúl Toral, and Maxi San Miguel. Nonequilibrium transitions in complex networks: A model of social interaction. *Physical review E*, 67(2):026120, 2003.

[90] Dietrich Stauffer. Sociophysics simulations. *arXiv preprint cond-mat/0210213*, 2002.

[91] Dietrich Stauffer. Sociophysics simulations II: opinion dynamics. *arXiv preprint physics/0503115*, 2005.

[92] Boris S Kerner and Hubert Rehborn. Experimental properties of phase transitions in traffic flow. *Physical Review Letters*, 79(20):4030, 1997.

[93] Natalie S Glance and Bernardo A Huberman. The outbreak of cooperation. *Journal of Mathematical sociology*, 17(4):281–302, 1993.

[94] Thomas C Schelling. Micromotives and macrobehavior ww norton & company. *New York, NY*, 1978.

[95] Richard Roll. The international crash of October 1987. *Financial analysts journal*, 44(5):19–35, 1988.

[96] Moshe Levy. Social phase transitions. *Journal of Economic Behavior & Organization*, 57(1):71–87, 2005.

[97] William A Brock and Steven N Durlauf. A formal model of theory choice in science. *Economic theory*, 14:113–130, 1999.

[98] Moshe Levy and Haim Levy. The danger of assuming homogeneous expectations. *Financial Analysts Journal*, 52(3):65–70, 1996.

[99] William A Brock. Pathways to randomness in the economy: emergent nonlinearity and chaos in economics and finance. *Estudios Economicos*, pages 3–55, 1993.

[100] Jean-Michel Lasry and Pierre-Louis Lions. Mean field games. *Japanese journal of mathematics*, 2(1):229–260, 2007.

[101] Josu Doncel, Nicolas Gast, and Bruno Gaujal. Discrete mean field games: Existence of equilibria and convergence. *arXiv preprint arXiv:1909.01209*, 2019.

[102] Rene Carmona. Applications of mean field games in financial engineering and economic theory. *arXiv preprint arXiv:2012.05237*, 2020.

[103] Aimé Lachapelle and Marie-Therese Wolfram. On a mean field game approach modeling congestion and aversion in pedestrian crowds. *Transportation research part B: methodological*, 45(10):1572–1589, 2011.

[104] W Brian Arthur. Inductive reasoning and bounded rationality. *The American economic review*, 84(2):406–411, 1994.

[105] Ole Peters and Alexander Adamou. The time interpretation of expected utility theory. *arXiv preprint arXiv:1801.03680*, 2018.

[106] Ole Peters. The ergodicity problem in economics. *Nature Physics*, 15(12):1216–1221, 2019.

[107] Yonatan Berman, Ole Peters, and Alexander Adamou. An empirical test of the ergodic hypothesis: Wealth distributions in the United States. *SSRN https://ssrn. com/abstract*, 2794830, 2017.

[108] David Meder, Finn Rabe, Tobias Morville, Kristoffer H Madsen, Magnus T Koudahl, Ray J Dolan, Hartwig R Siebner, and Oliver J Hulme. Ergodicity-breaking reveals time optimal decision making in humans. *PLoS computational biology*, 17(9):e1009217, 2021.

[109] Michael W Macy and Robert Willer. From factors to actors: Computational sociology and agent-based modeling. *Annual review of sociology*, 28(1):143–166, 2002.

[110] Adam M Johansen, Ludger Evers, and N Whiteley. Monte Carlo methods. *International encyclopedia of education*, pages 296–303, 2010.

[111] Dobriyan M Benov. The Manhattan Project, the first electronic computer and the Monte Carlo method. *Monte Carlo Methods and Applications*, 22(1):73–79, 2016.

[112] Nicholas Metropolis and Stanislaw Ulam. The Monte Carlo method. *Journal of the American statistical association*, 44(247):335–341, 1949.

[113] Nicholas Metropolis. The beginning of Monte Carlo. *Los Alamos Science Special Issue*, 15:125–130., 1987.

[114] Makoto Matsumoto and Takuji Nishimura. Mersenne twister: a 623-dimensionally equidistributed uniform pseudo-random number generator. *ACM Transactions on Modeling and Computer Simulation (TOMACS)*, 8(1):3–30, 1998.

[115] John Hammersley. *Monte Carlo methods*. Springer Science & Business Media, 2013.

[116] Alexander J McNeil, Rüdiger Frey, and Paul Embrechts. *Quantitative risk management: concepts, techniques and tools-revised edition*. Princeton university press, 2015.

[117] Kurt Binder, David M Ceperley, J-P Hansen, MH Kalos, DP Landau, D Levesque, H Mueller-Krumbhaar, D Stauffer, and J-J Weis. *Monte Carlo methods in statistical physics*, volume 7. Springer Science & Business Media, 2012.

[118] Mark EJ Newman and Gerard T Barkema. *Monte Carlo methods in statistical physics*. Clarendon Press, 1999.

[119] Young Hoon Kwak and Lisa Ingall. Exploring Monte Carlo simulation applications for project management. *Risk management*, 9:44–57, 2007.

[120] David Landau and Kurt Binder. *A guide to Monte Carlo simulations in statistical physics*. Cambridge university press, 2021.

[121] Jiaxin Zhang. Modern Monte Carlo methods for efficient uncertainty quantification and propagation: A survey. *Wiley Interdisciplinary Reviews: Computational Statistics*, 13(5):e1539, 2021.

[122] Michael Benzaquen and Jean-Philippe Bouchaud. Exploration of the parameter space in macroeconomic models. *Routledge International Handbook of Complexity Economics*, page 190, 2024.

[123] Charles M Macal and Michael J North. Tutorial on agent-based modeling and simulation. In *Proceedings of the Winter Simulation Conference, 2005.*, pages 14–pp. IEEE, 2005.

[124] Charles M Macal and Michael J North. Agent-based modeling and simulation. In *Proceedings of the 2009 winter simulation conference (WSC)*, pages 86–98. IEEE, 2009.

[125] Dirk Helbing. Agent-based modeling. In *Social self-organization: Agent-based simulations and experiments to study emergent social behavior*, pages 25–70. Springer, 2012.

[126] Nigel Gilbert. A simulation of the structure of academic science. *Sociological research online*, 2(2):91–105, 1997.

[127] Eric Bonabeau. Agent-based modeling: Methods and techniques for simulating human systems. *Proceedings of the national academy of sciences*, 99(suppl_3):7280–7287, 2002.

[128] Joshua M Epstein and Robert Axtell. *Growing artificial societies: social science from the bottom up*. Brookings Institution Press, 1996.

[129] Herbert A Simon. *The Sciences of the Artificial, reissue of the third edition with a new introduction by John Laird*. MIT press, 2019.

[130] Anirban Chakraborti, Ioane Muni Toke, Marco Patriarca, and Frédéric Abergel. Econophysics review: II. Agent-based models. *Quantitative Finance*, 11(7):1013–1041, 2011.

[131] Mark Newman, Albert-László Barabási, and Duncan J Watts. *The structure and dynamics of networks*. Princeton university press, 2011.

[132] Albert-László Barabási. Network science. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 371(1987):20120375, 2013.

[133] Duncan J Watts and Steven H Strogatz. Collective dynamics of 'small-world'networks. *nature*, 393(6684):440–442, 1998.

[134] Stefano Boccaletti, Vito Latora, Yamir Moreno, Martin Chavez, and D-U Hwang. Complex networks: Structure and dynamics. *Physics reports*, 424(4-5):175–308, 2006.

[135] Réka Albert and Albert-László Barabási. Statistical mechanics of complex networks. *Reviews of modern physics*, 74(1):47, 2002.

[136] Mark Newman. *Networks*. Oxford university press, 2018.

[137] Steven H Strogatz. Exploring complex networks. *nature*, 410(6825):268–276, 2001.

[138] Mark EJ Newman. The structure and function of complex networks. *SIAM review*, 45(2):167–256, 2003.

[139] Riccardo Marcaccioli and Giacomo Livan. A pólya urn approach to information filtering in complex networks. *Nature communications*, 10(1):745, 2019.

[140] Mark Granovetter. The strength of weak ties: A network theory revisited. *Sociological theory*, pages 201–233, 1983.

[141] Giovanni Bonanno, Guido Caldarelli, Fabrizio Lillo, Salvatore Micciche, Nicolas Vandewalle, and Rosario Nunzio Mantegna. Networks of equities in financial markets. *The European Physical Journal B*, 38(2):363–371, 2004.

[142] Andrew Cook, Henk AP Blom, Fabrizio Lillo, Rosario Nunzio Mantegna, Salvatore Miccichè, Damián Rivas, Rafael Vázquez, and Massimiliano Zanin. Applying complexity science to air traffic management. *Journal of Air Transport Management*, 42:149–158, 2015.

[143] Réka Albert, Hawoong Jeong, and Albert-László Barabási. Error and attack tolerance of complex networks. *nature*, 406(6794):378–382, 2000.

[144] Stefano Battiston, Michelangelo Puliga, Rahul Kaushik, Paolo Tasca, and Guido Caldarelli. Debtrank: Too central to fail? Financial networks, the fed and systemic risk. *Scientific reports*, 2(1):541, 2012.

[145] Duncan J Watts. A simple model of global cascades on random networks. *Proceedings of the National Academy of Sciences*, 99(9):5766–5771, 2002.

[146] Stefano Battiston and Guido Caldarelli. Systemic risk in financial networks. *Journal of Financial Management, Markets and Institutions*, 1(2):129–154, 2013.

[147] Paul Erdős, Alfréd Rényi, et al. On the evolution of random graphs. *Publications of the*, 1960.

[148] Albert-László Barabási and Eric Bonabeau. Scale-free networks. *Scientific american*, 288(5):60–69, 2003.

[149] Paul W Holland, Kathryn Blackmond Laskey, and Samuel Leinhardt. Stochastic blockmodels: First steps. *Social networks*, 5(2):109–137, 1983.

[150] Emmanuel Abbe. Community detection and stochastic block models: recent developments. *Journal of Machine Learning Research*, 18(177):1–86, 2018.

[151] Stefano Zapperi, Camille Varlet-Bertrand, Camille Bastidon, et al. Epidemiology models explain rumour spreading during France's Great Fear of 1789. *Nature*, 2025.

[152] George C Homans. *Social behavior: Its elementary forms*. Harcourt Brace Jovanovich, 1974.

[153] J. M. Keynes. *The General Theory of Employment, Interest and Money*. Macmillan, 1936. 14th edition, 1973.

[154] Maciej Lewenstein, Andrzej Nowak, and Bibb Latané. Statistical mechanics of social impact. *Physical Review A*, 45(2):763, 1992.

[155] Christian Borghesi and Jean-Philippe Bouchaud. Of songs and men: a model for multiple choice with herding. *Quality & quantity*, 41(4):557–568, 2007.

[156] Robert B Cialdini and Noah J Goldstein. Social influence: Compliance and conformity. *Annu. Rev. Psychol.*, 55(1):591–621, 2004.

[157] Serge Galam, Yuval Gefen (Feigenblat), and Yonathan Shapir and. Sociophysics: A new approach of sociological collective behaviour. I. mean-behaviour description of a strike. *The Journal of Mathematical Sociology*, 9(1):1–13, 1982.

[158] Michele Starnini, Fabian Baumann, Tobias Galla, David Garcia, Gerardo Iñiguez, Márton Karsai, Jan Lorenz, and Katarzyna Sznajd-Weron. Opinion dynamics: Statistical physics and beyond. *arXiv preprint arXiv:2507.11521*, 2025.

[159] Alina Sîrbu, Vittorio Loreto, Vito DP Servedio, and Francesca Tria. Opinion dynamics: models, extensions and external effects. In *Participatory sensing, opinions and collective awareness*, pages 363–401. Springer, 2016.

[160] Thomas C Schelling. *Micromotives and macrobehavior*. WW Norton & Company, 2006.

[161] Duncan J Watts and Peter Sheridan Dodds. Influentials, networks, and public opinion formation. *Journal of consumer research*, 34(4):441–458, 2007.

[162] Mohammad Afshar and Masoud Asadpour. Opinion formation by informed agents. *Journal of Artificial Societies and Social Simulation*, 13(4):5, 2010.

[163] Santo Fortunato, Vito Latora, Alessandro Pluchino, and Andrea Rapisarda. Vector opinion dynamics in a bounded confidence consensus model. *International Journal of Modern Physics C*, 16(10):1535–1551, 2005.

[164] David MJ Lazer, Alex Pentland, Duncan J Watts, Sinan Aral, Susan Athey, Noshir Contractor, Deen Freelon, Sandra Gonzalez-Bailon, Gary King, Helen Margetts, et al. Computational social science: Obstacles and opportunities. *Science*, 369(6507):1060–1062, 2020.

[165] Rama Cont and Jean-Philipe Bouchaud. Herd behavior and aggregate fluctuations in financial markets. *Macroeconomic dynamics*, 4(2):170–196, 2000.

[166] Thomas Lux and Michele Marchesi. Scaling and criticality in a stochastic multi-agent model of a financial market. *Nature*, 397(6719):498–500, 1999.

[167] Robert J Shiller. *Irrational exuberance: Revised and expanded third edition*. Princeton University Press, 2015.

[168] Petter Holme and Mark EJ Newman. Nonequilibrium phase transition in the coevolution of networks and opinions. *Physical Review E—Statistical, Nonlinear, and Soft Matter Physics*, 74(5):056108, 2006.

[169] William A Brock and Steven N Durlauf. Discrete choice with social interactions. *The Review of Economic Studies*, 68(2):235–260, 2001.

[170] Partha Dasgupta. The population problem: theory and evidence. *Journal of economic literature*, 33(4):1879–1902, 1995.

[171] Alessandro Pluchino, Alessio Emanuele Biondo, and Andrea Rapisarda. Talent versus luck: The role of randomness in success and failure. *Advances in Complex systems*, 21(03n04):1850014, 2018.

[172] Giacomo Livan. Don't follow the leader: how ranking performance reduces meritocracy. *Royal Society Open Science*, 6(11):191255, 2019.

[173] George C Homans. Fifty years of sociology. *Annual Review of Sociology*, 12(1):xiii–xxx, 1986.

[174] Paul L Krapivsky and Sidney Redner. Dynamics of majority rule in two-state interacting spin systems. *Physical Review Letters*, 90(23):238701, 2003.

[175] Peter Clifford and Aidan Sudbury. A model for spatial conflict. *Biometrika*, 60(3):581–588, 1973.

[176] Vishal Sood and Sidney Redner. Voter model on heterogeneous graphs. *Physical review letters*, 94(17):178701, 2005.

[177] Michael W Macy, Boleslaw K Szymanski, and Janusz A Hołyst. The Ising model celebrates a century of interdisciplinary contributions. *npj Complexity*, 1(1):10, 2024.

[178] Serge Galam. Spontaneous Symmetry Breaking and the transition to disorder in Physics. In *Cognitive Economics: An Interdisciplinary Approach*, pages 157–168. Springer, 2004.

[179] Mark Granovetter. Threshold models of collective behavior. *American journal of sociology*, 83(6):1420–1443, 1978.

[180] Serge Galam. Sociophysics: A review of Galam models. *International Journal of Modern Physics C*, 19(03):409–440, 2008.

[181] Serge Galam, Yuval Gefen, and Yonathan Shapir. Sociophysics: A new approach of sociological collective behaviour. i. mean-behaviour description of a strike. *Journal of Mathematical Sociology*, 9(1):1–13, 1982.

[182] Arkadiusz Jędrzejewski and Katarzyna Sznajd-Weron. Impact of memory on opinion dynamics. *Physica A: Statistical Mechanics and its Applications*, 505:306–315, 2018.

[183] Daniel McFadden. Conditional logit analysis of qualitative choice behavior. *Frontiers in Econometrics*, 1972.

[184] Serge Galam. Unifying local dynamics in two-state spin systems. *arXiv preprint cond-mat/0409484*, 2004.

[185] Sven Banisch and Eckehard Olbrich. Opinion polarization by learning from social feedback. *The Journal of Mathematical Sociology*, 43(2):76–103, 2019.

[186] Adrián Carro, Raúl Toral, and Maxi San Miguel. The noisy voter model on complex networks. *Scientific reports*, 6(1):24775, 2016.

[187] Roy J Glauber. Time-dependent statistics of the Ising model. *Journal of mathematical physics*, 4(2):294–307, 1963.

[188] Lawrence E Blume. The statistical mechanics of strategic interaction. *Games and economic behavior*, 5(3):387–424, 1993.

[189] Wolfgang Weidlich. *Sociodynamics: A systematic approach to mathematical modelling in the social sciences*. Courier Corporation, 2006.

[190] W Brian Arthur. Complexity and the economy. In *Handbook of Research on Complexity*. Edward Elgar Publishing, 2009.

[191] Jean-Philippe Bouchaud and Marc Mézard. Wealth condensation in a simple model of economy. *Physica A: Statistical Mechanics and its Applications*, 282(3):536–545, 2000.

[192] Daniela Morale, Vincenzo Capasso, and Karl Oelschläger. An interacting particle system modelling aggregation behavior: from individuals to populations. *Journal of mathematical biology*, 50:49–66, 2005.

[193] Janusz A Hołyst, Krzysztof Kacperski, and Frank Schweitzer. Phase transitions in social impact models of opinion formation. *Physica A: Statistical Mechanics and its Applications*, 285(1-2):199–210, 2000.

[194] Krzysztof Kacperski et al. Opinion formation model with strong leader and external impact: a mean field approach. *Physica A: Statistical Mechanics and its Applications*, 269(2-4):511–526, 1999.

[195] Janusz A Hołyst. Why does history surprise us? *Journal of Computational Science*, 73:102137, 2023.

[196] James P Gleeson. Binary-state dynamics on complex networks: Pair approximation and beyond. *Physical Review X*, 3(2):021004, 2013.

[197] Martin A Nowak, Corina E Tarnita, and Tibor Antal. Evolutionary dynamics in structured populations. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 365(1537):19–30, 2010.

[198] Harrison C White, Scott A Boorman, and Ronald L Breiger. Social structure from multiple networks. I. Blockmodels of roles and positions. *American journal of sociology*, 81(4):730–780, 1976.

[199] Damon Centola. The spread of behavior in an online social network experiment. *science*, 329(5996):1194–1197, 2010.

[200] Quentin Michard and J-P Bouchaud. Theory of collective opinion shifts: from smooth trends to abrupt swings. *The European Physical Journal B-Condensed Matter and Complex Systems*, 47:151–159, 2005.

[201] Jan Lorenz. Continuous opinion dynamics under bounded confidence: A survey. *International Journal of Modern Physics C*, 18(12):1819–1838, 2007.

[202] Rod Bond and Peter B Smith. Culture and conformity: A meta-analysis of studies using Asch's (1952b, 1956) line judgment task. *Psychological bulletin*, 119(1):111, 1996.

[203] Mehdi Moussaïd, Juliane E Kämmer, Pantelis P Analytis, and Hansjörg Neth. Social influence and the collective dynamics of opinion formation. *PloS one*, 8(11):e78433, 2013.

[204] Levi Boxell, Matthew Gentzkow, and Jesse M Shapiro. Cross-country trends in affective polarization. *Review of Economics and Statistics*, 106(2):557–565, 2024.

[205] Walter Quattrociocchi, Guido Caldarelli, and Antonio Scala. Opinion dynamics on interacting networks: media competition and social influence. *Scientific reports*, 4(1):4938, 2014.

[206] James N Druckman, Erik Peterson, and Rune Slothuus. How elite partisan polarization affects public opinion formation. *American political science review*, 107(1):57–79, 2013.

[207] Sara B Hobolt. The Brexit vote: a divided nation, a divided continent. *Journal of European public policy*, 23(9):1259–1277, 2016.

[208] Olivier Guedj and Jean-Philippe Bouchaud. Experts'earning forecasts: Bias, herding and gossamer information. *International Journal of Theoretical and Applied Finance*, 8(07):933–946, 2005.

[209] Flaminio Squazzoni and R Boero. Does empirical embeddedness matter?: Methodological issues on agent-based models for analytical social science. *JASSS*, 8(4):1–31, 2005.

[210] Paul Windrum, Giorgio Fagiolo, and Alessio Moneta. Empirical validation of agent-based models: Alternatives and prospects. *Journal of Artificial Societies and Social Simulation*, 10(2):8, 2007.

[211] Sushil Bikhchandani, David Hirshleifer, and Ivo Welch. A theory of fads, fashion, custom, and cultural change as informational cascades. *Journal of political Economy*, 100(5):992–1026, 1992.

[212] Ivo Welch. Sequential sales, learning, and cascades. *The Journal of finance*, 47(2):695–732, 1992.

[213] Andrea Baronchelli. The emergence of consensus: a primer. *Royal Society open science*, 5(2):172189, 2018.

[214] Edward L Glaeser, Bruce Sacerdote, and Jose A Scheinkman. Crime and social interactions. *The Quarterly journal of economics*, 111(2):507–548, 1996.

[215] Laurence Steinberg. *Beyond the classroom*. Simon and Schuster, 1997.

[216] Gerardo Iñiguez, Carlos Pineda, Carlos Gershenson, and Albert-László Barabási. Dynamics of ranking. *Nature communications*, 13(1):1646, 2022.

[217] Judith G Kelley and Beth A Simmons. The power of performance indicators: rankings, ratings and reactivity in international relations. In *APSA 2014 Annual Meeting Paper*, 2014.

[218] Aaron Clauset, Samuel Arbesman, and Daniel B Larremore. Systematic inequality and hierarchy in faculty hiring networks. *Science advances*, 1(1):e1400005, 2015.

[219] Péter Érdi. *Ranking: The unwritten rules of the social game we all play*. Oxford University Press, 2019.

[220] Weihua Li, Tomaso Aste, Fabio Caccioli, and Giacomo Livan. Early coauthorship with top scientists predicts success in academic careers. *Nature communications*, 10(1):5170, 2019.

[221] Mark EJ Newman. Who is the best connected scientist? a study of scientific coauthorship networks. In *Complex networks*, pages 337–370. Springer, 2004.

[222] Santo Fortunato, Carl T Bergstrom, Katy Börner, James A Evans, Dirk Helbing, Staša Milojević, Alexander M Petersen, Filippo Radicchi, Roberta Sinatra, Brian Uzzi, et al. Science of science. *Science*, 359(6379):eaao0185, 2018.

[223] Alan Mislove, Massimiliano Marcon, Krishna P Gummadi, Peter Druschel, and Bobby Bhattacharjee. Measurement and analysis of online social networks. In *Proceedings of the 7th ACM SIGCOMM conference on Internet measurement*, pages 29–42, 2007.

[224] Alexander Michael Petersen, Santo Fortunato, Raj K Pan, Kimmo Kaski, Orion Penner, Armando Rungi, Massimo Riccaboni, H Eugene Stanley, and Fabio Pammolli. Reputation and impact in academic careers. *Proceedings of the National Academy of Sciences*, 111(43):15316–15321, 2014.

[225] Matthew J Salganik, Peter Sheridan Dodds, and Duncan J Watts. Experimental study of inequality and unpredictability in an artificial cultural market. *science*, 311(5762):854–856, 2006.

[226] Charles AE Goodhart. Problems of monetary management: the uk experience. In *Monetary theory and practice: The UK experience*, pages 91–121. Springer, 1984.

[227] Christopher Mattson, Reamer L Bushardt, and Anthony R Artino Jr. When a measure becomes a target, it ceases to be a good measure, 2021.

[228] Wendy Nelson Espeland and Michael Sauder. Rankings and reactivity: How public measures recreate social worlds. *American journal of sociology*, 113(1):1–40, 2007.

[229] Marco Seeber, Mattia Cattaneo, Michele Meoli, and Paolo Malighetti. Self-citations as strategic response to the use of metrics for career decisions. *Research Policy*, 48(2):478–491, 2019.

[230] Isidro Aguillo, Judit Bar-Ilan, Mark Levene, and José Ortega. Comparing university rankings. *Scientometrics*, 85(1):243–256, 2010.

[231] Anthony FJ Van Raan. Fatal attraction: Conceptual and methodological problems in the ranking of universities by bibliometric methods. *Scientometrics*, 62(1):133–143, 2005.

[232] Sandro M Reia, Dieter Pfoser, and Paulo RA Campos. Group size matters: Synergistic effects and reduced inequality in performance rankings. *Physica A: Statistical Mechanics and its Applications*, 665:130496, 2025.

[233] Jerker Denrell and Chengwei Liu. Top performers are not the most impressive when extreme performance indicates unreliability. *Proceedings of the National Academy of Sciences*, 109(24):9331–9336, 2012.

[234] Ye Sun, Fabio Caccioli, and Giacomo Livan. Ranking mobility and impact inequality in early academic careers. *Proceedings of the National Academy of Sciences*, 120(34):e2305196120, 2023.

[235] Paul D Allison. Measures of inequality. *American sociological review*, pages 865–880, 1978.

[236] H Eugene Stanley. *Phase transitions and critical phenomena*. Clarendon Press, Oxford, 1971.

[237] Nigel Goldenfeld. *Lectures on phase transitions and the renormalization group*. CRC Press, 2018.

[238] Maurice G Kendall. A new measure of rank correlation. *Biometrika*, 30(1-2):81–93, 1938.

[239] Ohid Yaqub. Serendipity: Towards a taxonomy and a theory. *Research Policy*, 47(1):169–179, 2018.

[240] Oxford University Press. Serendipity, n.d. Accessed August 27, 2025.

[241] H Peyton Young. The evolution of conventions. *Econometrica: Journal of the Econometric Society*, pages 57–84, 1993.

[242] H Peyton Young. *Individual strategy and social structure: An evolutionary theory of institutions*. Princeton University Press, 2020.

[243] Damien Challet, Alessandro Pluchino, Alessio Emanuele Biondo, and Andrea Rapisarda. The origins of extreme wealth inequality in the talent versus luck model. *Advances in Complex Systems*, 23(02):2050004, 2020.

[244] Alessio Emanuele Biondo, Alessandro Pluchino, Andrea Rapisarda, and Dirk Helbing. Are random trading strategies more successful than technical ones? *PloS one*, 8(7):e68344, 2013.

[245] Claudio Castellano, Matteo Marsili, and Alessandro Vespignani. Nonequilibrium phase transition in a model for social influence. *Physical Review Letters*, 85(16):3536, 2000.

[246] Marco Bardoscia, Giancarlo De Luca, Giacomo Livan, Matteo Marsili, and Claudio J Tessone. The social climbing game. *Journal of statistical physics*, 151(3):440–457, 2013.

[247] Alan Kirman. Ants, rationality, and recruitment. *The Quarterly Journal of Economics*, 108(1):137–156, 1993.

[248] Takashi Nagatani. The physics of traffic jams. *Reports on progress in physics*, 65(9):1331, 2002.

[249] Filippo Radicchi, Santo Fortunato, Benjamin Markines, and Alessandro Vespignani. Diffusion of scientific credits and the ranking of scientists. *Physical Review E—Statistical, Nonlinear, and Soft Matter Physics*, 80(5):056103, 2009.

[250] Jaime A Teixeira da Silva, Panagiotis Tsigaris, and Mohammadamin Erfanmanesh. Publishing volumes in major databases related to Covid-19. *Scientometrics*, 126(1):831–842, 2021.

[251] Angelo Capodici, Aurelia Salussolia, Francesco Sanmarchi, Davide Gori, and Davide Golinelli. Biased, wrong and counterfeited evidences published during the COVID-19 pandemic, a systematic review of retracted COVID-19 papers. *Quality & quantity*, 57(5):4881–4913, 2023.

[252] Mark De Rond and Alan N Miller. Publish or perish: Bane or boon of academic life? *Journal of management inquiry*, 14(4):321–329, 2005.

[253] Seema Rawat and Sanjay Meena. Publish or perish: Where are we heading? *Journal of research in medical sciences: the official journal of Isfahan University of Medical Sciences*, 19(2):87, 2014.

[254] Panagiotis Tsolakidis, Naoum Mylonas, and Eugenia Petridou. The impact of imitation strategies, managerial and entrepreneurial skills on startups' entrepreneurial innovation. *Economies*, 8(4):81, 2020.

[255] Tegan Cruwys, Kirsten E Bevelander, and Roel CJ Hermans. Social modeling of eating: a review of when and why social influence affects food intake and choice. *Appetite*, 86:3–18, 2015.

[256] Bruno Biais, Pierre Hillion, and Chester Spatt. An empirical analysis of the limit order book and the order flow in the Paris Bourse. *the Journal of Finance*, 50(5):1655–1689, 1995.

[257] Rosario Nunzio Mantegna. Lévy walks and enhanced diffusion in Milan stock exchange. *Physica A: Statistical Mechanics and its Applications*, 179(2):232–242, 1991.

[258] H Eugene Stanley, Viktor Afanasyev, Luis A Nunes Amaral, Serguei V Buldyrev, Ary L Goldberger, Steve Havlin, Harry Leschhorn, Philipp Maass, Rosario N Mantegna, C-K Peng, et al. Anomalous fluctuations in the dynamics of complex systems: from DNAd and physiology to econophysics. *Physica A: Statistical Mechanics and its Applications*, 224(1-2):302–321, 1996.

[259] Victor M Yakovenko. Econophysics, statistical mechanics approach to. In *Complex Systems in Finance and Econometrics*, pages 247–273. Springer, 2009.

[260] Kausik Gangopadhyay. Interview with Eugene H. Stanley. *IIM Kozhikode Society & Management Review*, 2(2):73–78, 2013.

[261] Annamaria Lusardi and Olivia S Mitchell. Financial literacy around the world: an overview. *Journal of pension economics & finance*, 10(4):497–508, 2011.

[262] Eugene F Fama. Efficient capital markets: A review of theory and empirical work. *The journal of Finance*, 25(2):383–417, 1970.

[263] Robert J Shiller. *Market volatility*. MIT press, 1992.

[264] Parameswaran Gopikrishnan, Martin Meyer, LA Nunes Amaral, and H Eugene Stanley. Inverse cubic law for the distribution of stock price variations. *The European Physical Journal B-Condensed Matter and Complex Systems*, 3(2):139–140, 1998.

[265] Robert F Engle and Andrew J Patton. What good is a volatility model? In *Forecasting volatility in the financial markets*, pages 47–63. Elsevier, 2007.

[266] Torben G Andersen, Tim Bollerslev, Peter F Christoffersen, and Francis X Diebold. Volatility and correlation forecasting. *Handbook of economic forecasting*, 1:777–878, 2006.

[267] Anirban Chakraborti, Ioane Muni Toke, Marco Patriarca, and Frédéric Abergel. Econophysics review: I. Empirical facts. *Quantitative Finance*, 11(7):991–1012, 2011.

[268] John C Hull. *Risk management and financial institutions*. John Wiley & Sons, 2023.

[269] Dean Rickles. Econophysics and the complexity of financial markets. In *Philosophy of complex systems*, pages 531–565. Elsevier, 2011.

[270] Nicholas Kaldor. Capital accumulation and economic growth. In *The Theory of capital: proceedings of a conference held by the International Economic Association*, pages 177–222. Springer, 1961.

[271] Dominique M Guillaume, Michel M Dacorogna, Rakhal R Davé, Ulrich A Müller, Richard B Olsen, and Olivier V Pictet. From the bird's eye to the microscope: A survey of new stylized facts of the intra-daily foreign exchange markets. *Finance and stochastics*, 1(2):95–129, 1997.

[272] Rama Cont. Empirical properties of asset returns: stylized facts and statistical issues. *Quantitative finance*, 1(2):223, 2001.

[273] Rosario N Mantegna and H Eugene Stanley. *Introduction to econophysics: correlations and complexity in finance*. Cambridge university press, 1999.

[274] Eric Jondeau, Ser-Huang Poon, and Michael Rockinger. *Financial modeling under non-Gaussian distributions*. Springer, 2007.

[275] Parameswaran Gopikrishnan, Vasiliki Plerou, Luis A Nunes Amaral, Martin Meyer, and H Eugene Stanley. Scaling of the distribution of fluctuations of financial market indices. *Physical Review E*, 60(5):5305, 1999.

[276] L Kullmann, J Töyli, J Kertesz, A Kanto, and K Kaski. Characteristic times in stock market indices. *Physica A: Statistical Mechanics and its Applications*, 269(1):98–110, 1999.

[277] Johannes Voit. *The statistical mechanics of financial markets*. Springer, 2003.

[278] Paul A Samuelson. Rational theory of warrant pricing. In *Henry P. McKean Jr. Selecta*, pages 195–232. Springer, 1965.

[279] Maury FM Osborne. Brownian motion in the stock market. *Operations research*, 7(2):145–173, 1959.

[280] Eugene F Fama. The behavior of stock-market prices. *The journal of Business*, 38(1):34–105, 1965.

[281] Eugene F Fama. Mandelbrot and the stable Paretian hypothesis. *The journal of business*, 36(4):420–429, 1963.

[282] Rosario N Mantegna and H Eugene Stanley. Scaling behaviour in the dynamics of an economic index. *Nature*, 376(6535):46–49, 1995.

[283] Vasiliki Plerou, Parameswaran Gopikrishnan, Luis A Nunes Amaral, Martin Meyer, and H Eugene Stanley. Scaling of the distribution of price fluctuations of individual companies. *Physical review e*, 60(6):6519, 1999.

[284] Ahmet Celikoglu and Ugur Tirnakli. Skewness and kurtosis analysis for non-Gaussian distributions. *Physica A: Statistical Mechanics and its Applications*, 499:325–334, 2018.

[285] Peter D Praetz. The distribution of share price changes. *Journal of business*, pages 49–55, 1972.

[286] Robert C Blattberg and Nicholas J Gonedes. A comparison of the stable and student distributions as statistical models for stock prices. In *Perspectives on promotion and database marketing: The collected works of Robert C Blattberg*, pages 25–61. World Scientific, 2010.

[287] Eckhard Platen and Renata Rendek. Empirical evidence on Student-t log-returns of diversified world stock indices. *Journal of statistical theory and practice*, 2(2):233–251, 2008.

[288] Gao-Feng Gu, Wei Chen, and Wei-Xing Zhou. Empirical distributions of Chinese stock returns at different microscopic timescales. *Physica A: Statistical Mechanics and its Applications*, 387(2-3):495–502, 2008.

[289] Austin Gerig, Javier Vicente, and Miguel A Fuentes. Model for non-Gaussian intraday stock returns. *Physical Review E—Statistical, Nonlinear, and Soft Matter Physics*, 80(6):065102, 2009.

[290] Yuri A Katz and Li Tian. q-Gaussian distributions of leverage returns, first stopping times, and default risk valuations. *Physica A: Statistical Mechanics and its Applications*, 392(20):4989–4996, 2013.

[291] Fernando Alonso-Marroquin, Karina Arias-Calluari, Michael Harré, Morteza N Najafi, and Hans J Herrmann. Q-Gaussian diffusion in stock markets. *Physical Review E*, 99(6):062313, 2019.

[292] Constantino Tsallis. Possible generalization of Boltzmann-Gibbs statistics. *Journal of statistical physics*, 52(1):479–487, 1988.

[293] Constantino Tsallis, Celia Anteneodo, Lisa Borland, and Roberto Osorio. Nonextensive statistical mechanics and economics. *Physica A: Statistical Mechanics and its Applications*, 324(1-2):89–100, 2003.

[294] R Rak, S Drożdż, and J Kwapień. Nonextensive statistical features of the Polish stock market fluctuations. *Physica A: statistical mechanics and its applications*, 374(1):315–324, 2007.

[295] Yannick Malevergne and Didier Sornette. Multivariate Weibull distributions for asset returns: I. *Available at SSRN 714161*, 2005.

[296] Yannick Malevergne, Vladilen Pisarenko, and Didier Sornette. Empirical distributions of stock returns: between the stretched exponential and the power law? *Quantitative Finance*, 5(4):379–401, 2005.

[297] Jean Laherrere and Didier Sornette. Stretched exponential distributions in nature and economy: "fat tails" with characteristic scales. *The European Physical Journal B-Condensed Matter and Complex Systems*, 2(4):525–539, 1998.

[298] Saralees Nadarajah and Samuel Kotz. The modified Weibull distribution for asset returns. *Quantitative Finance*, 6(6):449–449, 2006.

[299] Hari M Gupta and José R Campanha. The gradually truncated Lévy flight for systems with power-law distributions. *Physica A: Statistical Mechanics and its Applications*, 268(1-2):231–239, 1999.

[300] L Couto Miranda and R Riera. Truncated Lévy walks and an emerging market economic index. *Physica A: Statistical Mechanics and its Applications*, 297(3-4):509–520, 2001.

[301] Maria Christina Mariani and Y Liu. Normalized truncated Lévy walks applied to the study of financial indices. *Physica A: Statistical Mechanics and its Applications*, 377(2):590–598, 2007.

[302] Ismo Koponen. Analytic approach to the problem of convergence of truncated Lévy flights towards the Gaussian stochastic process. *Physical Review E*, 52(1):1197, 1995.

[303] Carl De Boor and Carl De Boor. *A practical guide to splines*, volume 27. springer New York, 1978.

[304] Charles R Harris, K Jarrod Millman, Stéfan J Van Der Walt, Ralf Gommers, Pauli Virtanen, David Cournapeau, Eric Wieser, Julian Taylor, Sebastian Berg, Nathaniel J Smith, et al. Array programming with NumPy. *nature*, 585(7825):357–362, 2020.

[305] Pauli Virtanen, Ralf Gommers, Travis E Oliphant, Matt Haberland, Tyler Reddy, David Cournapeau, Evgeni Burovski, Pearu Peterson, Warren Weckesser, Jonathan Bright, et al. SciPy 1.0: fundamental algorithms for scientific computing in Python. *Nature methods*, 17(3):261–272, 2020.

[306] Luc Devroye. Sample-based non-uniform random variate generation. In *Proceedings of the 18th conference on Winter simulation*, pages 260–265, 1986.

[307] Christian P Robert, George Casella, and George Casella. *Monte Carlo statistical methods*, volume 2. Springer, 1999.

[308] William J Thistleton, John A Marsh, Kenric Nelson, and Constantino Tsallis. Generalized Box–Müller method for generating q-Gaussian random deviates. *IEEE transactions on information theory*, 53(12):4805–4810, 2007.

[309] Kenric P Nelson and William J Thistleton. Comments on "Generalized Box-Müller Method for Generating q-Gaussian Random Deviates". *IEEE Transactions on Information Theory*, 67(10):6785–6789, 2021.

[310] George EP Box and Mervin E Muller. A note on the generation of random normal deviates. *The annals of mathematical statistics*, 29(2):610–611, 1958.

[311] John M Chambers, Colin L Mallows, and BW4159820341 Stuck. A method for simulating stable random variables. *Journal of the american statistical association*, 71(354):340–344, 1976.

[312] Rosario Nunzio Mantegna. Fast, accurate algorithm for numerical simulation of Lévy stable stochastic processes. *Physical Review E*, 49(5):4677, 1994.

[313] Rafał Weron. Lévy-stable distributions revisited: tail index¿ 2 does not exclude the Lévy-stable regime. *International Journal of Modern Physics C*, 12(02):209–223, 2001.

[314] S Drożdż, M Forczek, J Kwapień, P Oświe, R Rak, et al. Stock market return distributions: From past to present. *Physica A: Statistical Mechanics and its Applications*, 383(1):59–64, 2007.

[315] Ruey S Tsay. *Analysis of financial time series*. John wiley & sons, 2005.

[316] María del Mar López Martín, Catalina García García, and José García Pérez. Treatment of kurtosis in financial markets. *Physica A: Statistical Mechanics and its Applications*, 391(5):2032–2045, 2012.

[317] Michael Grabchak and Gennady Samorodnitsky. Do financial returns have finite or infinite variance? A paradox and an explanation. *Quantitative Finance*, 10(8):883–893, 2010.

[318] Christophe Schinckus. How physicists made stable Lévy processes physically plausible. *Brazilian Journal of Physics*, 43(4):281–293, 2013.

[319] Xuerong Mao. The truncated Euler–Maruyama method for stochastic differential equations. *Journal of Computational and Applied Mathematics*, 290:370–384, 2015.

[320] Riccardo Mannella. A gentle introduction to the integration of stochastic differential equations. In *Stochastic processes in physics, chemistry, and biology*, pages 353–364. Springer, 2001.

[321] John Y Campbell, Andrew W Lo, A Craig MacKinlay, and Robert F Whitelaw. The econometrics of financial markets. *Macroeconomic Dynamics*, 2(4):559–562, 1998.

[322] Zhuanxin Ding, Clive WJ Granger, and Robert F Engle. A long memory property of stock market returns and a new model. *Journal of empirical finance*, 1(1):83–106, 1993.

[323] Vasiliki Plerou, Parameswaran Gopikrishnan, Luís A Nunes Amaral, Xavier Gabaix, and H Eugene Stanley. Economic fluctuations and anomalous diffusion. *Physical Review E*, 62(3):R3023, 2000.

[324] Sidney Redner. Random multiplicative processes: An elementary tutorial. *Am. J. Phys*, 58(3):267–273, 1990.

[325] Rosario N Mantegna and H Eugene Stanley. Stochastic process with ultraslow convergence to a Gaussian: the truncated Lévy flight. *Physical Review Letters*, 73(22):2946, 1994.

[326] Christopher Jarzynski. Rare events and the convergence of exponentially averaged work values. *Physical Review E—Statistical, Nonlinear, and Soft Matter Physics*, 73(4):046105, 2006.

[327] Liuren Wu. Dampened power law: reconciling the tail behavior of financial security returns. *The Journal of Business*, 79(3):1445–1473, 2006.

[328] Erich Leo Lehmann and Joseph P Romano. *Testing statistical hypotheses*. Springer, 2005.

[329] Rick Durrett. *Probability: theory and examples*, volume 49. Cambridge university press, 2019.

[330] Herman Callaert and Paul Janssen. The Berry-Esseen theorem for U-statistics. *The Annals of Statistics*, pages 417–421, 1978.

[331] Andrew C Berry. The accuracy of the Gaussian approximation to the sum of independent variates. *Transactions of the american mathematical society*, 49(1):122–136, 1941.

[332] Carl-Gustav Esseen. On the Liapounoff limit of error in the theory of probability. *Arkiv för matematik, astronomi och fysik*, 28A(9):1–19, 1942.

[333] Irina Shevtsova. On the absolute constants in the Berry-Esseen-type inequalities. *Doklady Mathematics*, 89(3), 2014.

[334] Sidney Redner. *A guide to first-passage processes*. Cambridge university press, 2001.

[335] Jaume Masoliver and Josep Perelló. First-passage and risk evaluation under stochastic volatility. *Physical Review E—Statistical, Nonlinear, and Soft Matter Physics*, 80(1):016108, 2009.

[336] Bikas K Chakrabarti, Anirban Chakraborti, and Arnab Chatterjee. *Econophysics and sociophysics: trends and perspectives*. John Wiley & Sons, 2006.

[337] John C Hull and Sankarshan Basu. *Options, futures, and other derivatives*. Pearson Education India, 2016.

[338] Fischer Black and Myron Scholes. The pricing of options and corporate liabilities. *Journal of political economy*, 81(3):637–654, 1973.

[339] Robert C Merton et al. *Theory of rational option pricing*. World Scientific, 1971.

[340] NobelPrize.org. The Nobel Prize in Economics sciences, 1997. Accessed: 2025-06-30.

[341] Paul Glasserman. *Monte Carlo methods in financial engineering*, volume 53. Springer, 2004.

[342] Les Clewlow and Chris Strickland. *Implementing derivatives models*. Wiley series in financial engineering. Wiley, 1998.

[343] Quasi-Monte Carlo. Monte carlo methods in financial engineering, 2001.

[344] J Michael Harrison and Stanley R Pliska. Martingales and stochastic integrals in the theory of continuous trading. *Stochastic processes and their applications*, 11(3):215–260, 1981.

[345] Andrew Matacz. Financial modeling and option theory with the truncated Lévy process. *International Journal of Theoretical and Applied Finance*, 3(01):143–160, 2000.

[346] Tamirat Temesgen Dufera. Fractional Brownian motion in option pricing and dynamic delta hedging: Experimental simulations. *The North American Journal of Economics and Finance*, 69:102017, 2024.

[347] Steven L Heston. A closed-form solution for options with stochastic volatility with applications to bond and currency options. *The review of financial studies*, 6(2):327–343, 1993.

[348] Svetlana I Boyarchenko and Sergei Z Levendorskiĭ. Option pricing for truncated Lévy processes. *International journal of theoretical and applied finance*, 3(03):549–552, 2000.

[349] Lisa Borland. A theory of non-Gaussian option pricing. *Quantitative Finance*, 2(6):415, 2002.

[350] Salvatore Micciche, Giovanni Bonanno, Fabrizio Lillo, and Rosario N Mantegna. Volatility in financial markets: stochastic models and empirical results. *Physica A: Statistical Mechanics and its Applications*, 314(1-4):756–761, 2002.

[351] Veli-Pekka Heikkinen and Antti Kanto. Value-at-risk estimation using non-integer degrees of freedom of Student's distribution. *Journal of Risk*, 2002.

[352] Giacomo Bormetti, Enrica Cisana, Guido Montagna, and Oreste Nicrosini. A non-Gaussian approach to risk measures. *Physica A: Statistical Mechanics and its Applications*, 376:532–542, 2007.