Babeş-Bolyai University

Doctoral Thesis Summary

Applications of the Local Growth and Global Reset (LGGR) model for socio-economic and biological problems

Author:
Kelemen Szabolcs

Supervisors:
Prof. Dr. Néda Zoltán
Prof. Dr. Nagy Ladislau

A Thesis Submitted in Fulfilment
of the Requirements for the Degree of
Doctor of Philosophy
in the
Domain of Physics

at

Department of Physics
Babeş-Bolyai University

Cluj-Napoca, Romania

February, 2024

Applications of the Local Growth and Global Reset (LGGR) model for socio-economic and biological problems

Abstract

The internal dynamics of complex systems emerge from the interaction of numerous components guided by characteristic rules of the system. Biological, economic, and social systems, in particular, exhibit these inherent interactions among their numerous components. A particular type of frequently observed dynamics is one that is defined by unidirectional growth and probabilistic reset processes. Due to the emerging complexity, the study of such systems necessitates the proper tools of analysis.

This thesis provides a comprehensive overview and analysis of real-world phenomena through the lens of a mean-field approach founded on a master equation, the Local Growth and Global Reset (LGGR) model. A central objective is to offer a pedagogical presentation, by providing a detailed overview of general master equations, and Markov processes, and establishing the mathematical foundation for later applications of the LGGR model. The study encompasses four modeling investigations, each applied to phenomena observed in socio-economic and biological systems. The socio-economic section of the thesis explores wealth distributions in leading countries and a small Romanian commune. The first study analyzes individual wealth distribution in the United States, Russia, and France. The model captures wealth dynamics, including negative values, and outperforms existing models. The second study applies the LGGR model to wealth distribution in a Transylvanian commune, revealing shifts across different economic systems and contrasting wealth distribution patterns during communism and the free-market era. A third study, which is still focused on socio-economic systems, explores the dynamics of lottery jackpots. Analyzing six lotteries, the LGGR model captures the emergent evolution of Jackpot values influenced by player behavior, lottery rules, and probabilistic reset rates. The final study steps into the realm of biological systems, examining diversity patterns in tree sizes within deciduous wooded environments. Confirming the Gamma distribution for specific deciduous species, the model demonstrates statistical universality across taxa and ecosystems.

To conclude, the thesis demonstrates the versatility of the LGGR model in understanding complex phenomena across socio-economic and biological systems. The success in modeling diverse phenomena suggests its potential for wide applicability in various scientific fields.

Keywords: complex systems, local growth, global reset, socio-economic phenomena, biological diversity

Contents

Chapter 1

Introduction

The landscape of physics has undergone significant evolution over the past two centuries, with physicists increasingly tackling complex systems that defy conventional intuition [1–4]. This shift toward confronting complexity has led to a strong interdisciplinary character within the field, spanning biology [5], chemistry [6], economics [7], and social sciences [8]. At the core of physics lies the mission to elucidate the fundamental principles governing matter and energy behavior, often achieved through mathematical modeling. The study of intricate systems has prompted the development of elegant mathematical models, particularly in understanding probabilistic phenomena. For instance, the discovery of Brownian motion by Robert Brown in 1827 sparked interest in probabilistic intricacies underlying seemingly chaotic phenomena [9]. Einstein, Schmoluchowski, and Langevin were among the first to successfully model Brownian motion, paving the way for the Fokker-Planck equation and subsequent evolutionary equations describing probabilistic system evolution [10]. Amidst various modeling paradigms, the phenomenological master equation approach has emerged as a central player in statistical physics, describing systems governed by probabilistic dynamic processes [10]. Einstein, Furry, and Feller, among others, applied master equations to diverse phenomena, including black-body radiation, electron passage through materials, and population dynamics [10, 11]. While the framework of master equations became widely known in statistical physics, the term "master equation" was first used in 1940 by Nordsieck et al [12] Since then, master equations have been extensively applied across research domains, offering profound insights into intricate systems' probabilistic evolution [13].

1.1 The general form of the master equation

In case when the system has a finite number of discrete inner states and the elements of the system can jump from state to state probabilistically and independently of each other in time, the change of the probability $P_n(t)$ taking the continuous limit of the time parameter ($\Delta t \to 0$) is given by Equation 1.1. That is the general form of the continuous time, discrete state-space master equation. To give some examples, these states could represent different positions, energy levels, or other characteristics of the system.

$$\frac{dP_n(t)}{dt} = \sum_{\{m\}} [w_{m,n} P_m(t) - w_{n,m} P_n(t)]. \tag{1.1}$$

where $P_n(t)$ denotes the occupation probability of state n at time moment t, while $w_{n,m}$ stands for the directed transition rate (probability over unit time) between the two states m and n. The transition rates $w_{n,m}$, in this case, are independent of time.

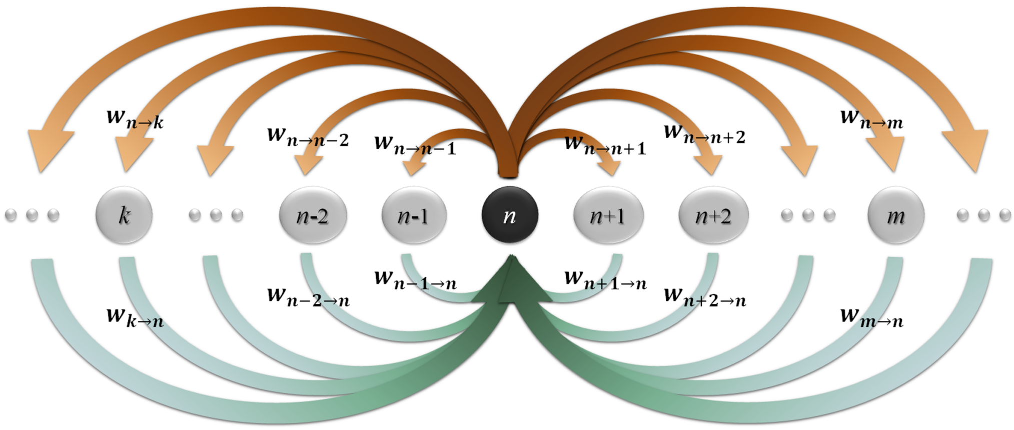

Figure 1.1: Shematic representation of the dynamics described by the discrete master equation (Equation 1.1).

The dynamics described by Equation 1.1 is illustrated in Figure 1.1. The collection of transition rates forms a transition matrix $W_{n,m}$, which encodes the dynamics of the system. The occupation probability is simply given by the fraction of the system's elements being in state n divided by the total number of elements. Therefore the probability distribution $P_n(t)$, is normalized: $\sum_{\{n\}} P_n(t) = 1$.

The master equation can be written in continuous state-space limit as well when the system has infinitely many states. In this case Equation 1.1 transforms into an integro-differencial equation written in the form of Equation 1.2 [14].

$$\frac{\partial \rho(x,t)}{\partial t} = \int_{-\infty}^{\infty} [w(x,y)\rho(y,t) - w(y,x)\rho(x,t)]dy, \tag{1.2}$$

In the continuous case, however, instead of the time evolution of the $P_n(t)$ occupation probability, the time evolution of the $\rho(x,t)$ probability density function is defined by the evolutionary Equation 1.2. The transition rates $w(x,y)$, are independent of time in this case as well. The probability density function also has to be normalized: $\int_{-\infty}^{\infty} \rho(x,t)dx = 1$.

The widely applicable nature of the master equations arises from the flexibility in defining the transition matrix $W_{n,m}$. This flexibility, however, is limited by the fact that in the case when the studied system has a large number of states (Equation 1.1), or the transition rates have complex forms (Equation 1.2), the master equation becomes analytically intractable. Hence, there are analytical solutions available in only a very limited number of instances.

1.1 Markov processes

When discussing master equations, it is crucial to also introduce Markov processes, as they are closely related. Markov processes, introduced by Andrey Markov in the early 20th century, serve as the foundation for systems employing master equations. A Markov process is characterized by its memoryless property, where the next state of the system depends solely on its current state. If one describes the time evolution of the system through discrete states $(x_0, x_1, x_2, \dots x_n, \dots)$ that can be reached sequentially, then the Markov property can be described using the conditional probability $P(x_n \| x_{n-1}, \dots, x_0)$ that the state x_n is reached throughout the states $x_0, x_1, \dots, x_{n-1}$:

$P(x_n, t \| x_{n-1}, t-\Delta t, ..., x_0, 0) = P(x_n, t \| x_{n-1}, t-\Delta t)$. The conditional probability $P(x_n, t \| x_{n-1}, t-\Delta t)$, however, corresponds to the transition probability per unit time (Δt) to jump from state x_{n-1} to x_n, which corresponds to the $w_{n,n-1}$ transition rate from state x_{n-1} to x_n appearing in the master equation, Equation 1.1. Thus, the memoryless property of transitions is inherently present within the master equation. Markov processes often assume stationarity, where the probability distribution of relevant quantities remains constant over time: $\frac{dP_n(t)}{dt} = 0$. This stationary distribution, derived from the transition rates and probabilities offers valuable insights into the system's dynamics. Reversible Markov processes represent a special category where dynamics are both stationary and satisfy the detailed balance condition ($w_{n,m} \cdot P_m = w_{m,n} \cdot P_n$). This condition ensures equilibrium dynamics, where the forward and reverse processes exhibit equivalent stationary probability distributions. Ergodicity, an essential property of Markov chains, encompasses both aperiodicity (for each state of the system it holds that there is no such finite number $k > 1$, representing the number of steps for which the probability of returning to state x_i is 1: $P^{(k)}(x_i \| x_i) = 1$) and irreducibility (the possibility of reaching any state of the system from any other state). A Markov chain is ergodic if it lacks periodic states and is irreducible, allowing for the existence of a unique stationary distribution over its possible states.

Chapter 2

The Local Growth and Global Reset (LGGR) process

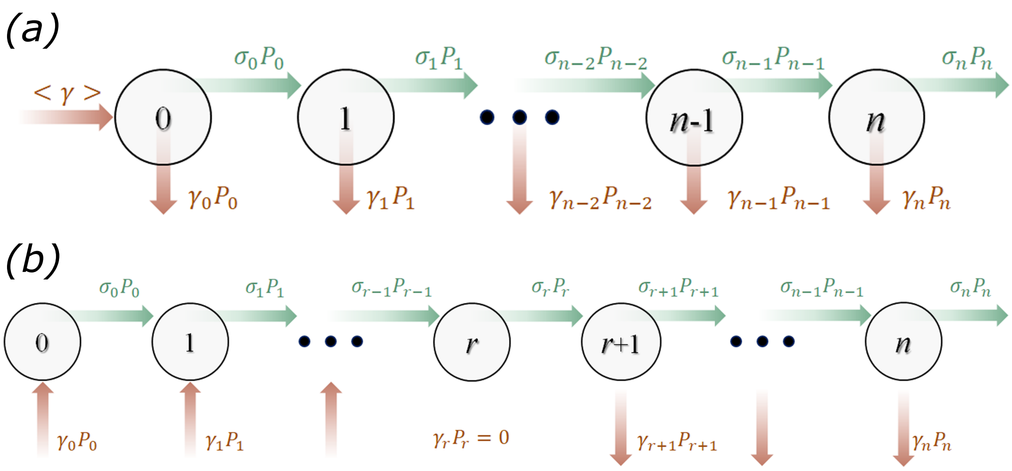

Figure 2.1: The sketch of the growth and reset process (a) for the simple mechanism with a positive reset rate; (b) when the reset rate can be both positive and negative ($\gamma_n < 0$ if $n < r$, and $\gamma_n > 0$ for $n > r$).

In this chapter, we present a particular class of master equations that consists of a unidirectional growth process characterized by the growth rate σ_n (or $\sigma(x)$ in the continuous case) and a resetting process, characterized by the reset rate γ_n (or $\gamma(x)$)

to the ground state denoted as $n = 0$, which ensures the possibility of stationary dynamics. As a result of the broken symmetry, the detailed balance condition is not satisfied, turning the dynamics into a non-reversible Markov process. The sketch of this dynamics is represented in Figure 2.1a. The model is known as the Local Growth and Global Reset (LGGR) model, denoted by the encapsulating processes, or it is commonly recognized as the Biró-Néda model, named after its developers [15, 16]. The dynamics involving persistent growth followed by resetting is a recurring pattern in numerous physical, biological, social, and economic phenomena. In these complex systems the intricacy of the dynamics gives rise to diverse probability distributions of the characteristic quantity. The most commonly observed distributions are characterized by power-law tails, which emerge as a consequence of a form of preferential behavior within the underlying dynamics. However, this type of dynamics is not limited to generating such fat-tailed distributions. By appropriately choosing the forms of the growth and reset kernel functions, a wide array of other distribution types can be analytically derived [16].

In most real-world systems our interest lies in demonstrating the systems in equilibrium. Therefore, we aim to model the stationary fluctuations that result in the characteristic, stationary distribution P_n^s ($\rho_s(x)$).

The transition rates ($w_{m,n}$) (meaning here n the starting state and m being the destination of the transition) corresponding to the discrete growth and reset processes are defined as: $w_{m,n} = \sigma_n \delta_{m,n+1} + \gamma_n \delta_{m,0}$. Utilizing the general form of the master equation (Equation 1.1) together with these transition rates one obtains the form of the master equation corresponding to the discrete state-space LGGR process given by Equation 2.1.

$$\frac{dP_n(t)}{dt} = \sigma_{n-1} P_{n-1}(t) + \langle\gamma\rangle(t)\delta_{n,0} - (\sigma_n + \gamma_n)P_n(t). \tag{2.1}$$

Assuming the stationary limit (the left hand size of Equation 2.1 becomes 0: $\frac{dP_n(t)}{dt} = 0$) the P_n probability can be calculated as the following formula [15]:

$$P_n = P_0 \frac{\sigma_0}{\sigma_n} e^{-\sum_{i=0}^{n} \ln(1+\frac{\sigma_i}{\gamma_i})}. \tag{2.2}$$

When the relevant quantity characterizing the elements of the system changes continuously the resulting form of the evolution equation determines now the time evolution of the probability density function $\rho(x,t)$:

$$\frac{\partial \rho(x,t)}{\partial t} = -\frac{\partial}{\partial x}\left[\sigma(x)\rho(x,t)\right] - \gamma(x)\rho(x,t) + \langle\gamma(x)\rangle(t)\delta(x). \tag{2.3}$$

The mean value of the reset rate $\langle\gamma(x)\rangle(t)$ is calculated using the probability density function $\rho(x,t)$: $\langle\gamma\rangle(t) = \int_0^\infty \gamma(x)\rho(x,t)dx$. The normalization of the probability density function $\int_0^\infty \rho(x,t)dx = 1$ is ensured by the refeeding of the system with new entities at the ground state $x = 0$ through the $\langle\gamma(x)\rangle(t)\delta(x)$ component. In the stationary condition, the solution of Equation 2.3 gives the general form of the stationary probability density function $\rho_s(x)$ (Equation 2.4).

$$\rho_s(x) = \frac{\sigma(0)\rho_s(0)}{\sigma(x)} e^{-\int\limits_0^x \frac{\gamma(u)}{\sigma(u)}du}, \tag{2.4}$$

This is the continuous correspondent of Equation 2.2. In most cases when dealing with real-world phenomena, the studied quantity is quasi-continuous (i.e. income, wealth, size of individuals in biological systems, ...). Thus in these cases, we root back to Equation 2.4 for calculating the system's characteristic distribution. The stationary solutions presented in Equations 2.2 and 2.4 exhibit a high degree of universality, as they depend exclusively on the forms of the σ_n ($\sigma(x)$) and γ_n ($\gamma(x)$) functions [16–19]. The reset rate $\gamma(x)$ allows for the differentiation of two unique dynamical scenarios. The simplest scenario occurs when the reset rate $\gamma(x)$ is positive for all values of x (and for n in case of discrete state space, γ_n). This behavior is illustrated in Figure 2.1a. A more complex dynamic scenario arises when the reset rate $\gamma(x)$ (γ_n) can be positive or negative depending on the value of x (x_n). The second scenario considering the state-dependent "smart reset" rate proved to be more flexible for applications. This behavior is illustrated in Figure 2.1b.

An alternative way to expand the evolutionary equation is by accounting for scenarios where the number of elements changes multiplicatively over time: $\frac{dN_{total}(t)}{dt} = \kappa N_{total}(t)$. By introducing such dilution into the master equation, a new reset-like term emerges.

$$\frac{dP_n(t)}{dt} = \sigma_{n-1}P_{n-1}(t) - \sigma_n P_n(t) - (\gamma_n + \kappa)P_n(t) + \delta_{n,0}\langle\gamma\rangle(t). \tag{2.5}$$

In the continuous state space, the corresponding partial differential equation to Equation 2.5 is expressed in the following form:

$$\frac{\partial\rho(x,t)}{\partial t} = -\frac{\partial}{\partial x}\left[\sigma(x)\rho(x,t)\right] - (\gamma(x) + \kappa)\rho(x,t) + \langle\gamma(x)\rangle(t)\delta(x). \tag{2.6}$$

In the stationary state, the general form of the probability density function for $x > 0$ also depends on the κ dilution rate:

$$\rho_s(x) = \frac{C}{\sigma(x)} \mathrm{e}^{-\int_{\{x\}} \frac{(\gamma(u)+\kappa)}{\sigma(u)} du}, \tag{2.7}$$

where C is the normalization constant.

Chapter 3

Distribution of wealth in the society

3.1 Introduction to social inequality based on wealth

Social inequalities refer to the widespread discrepancies in the allocation of resources, opportunities, and privileges among individuals or groups within a given society. One common and vital quantity for measuring these inequalities is wealth, which encompasses all types of property that an individual owns. Such inequalities have always been present in society, even in historical times, regardless of the type of administrative system [20]. Although quantitative evidence of wealth inequality can be found in the literature, in all cases, when the goal is to measure wealth, it is approximated through various methods that rely on different proxies [21]. Vilfredo Pareto, a renowned italian sociologist, delivered the initial empirical and qualitative description of wealth distribution in the late 19th century. His discoveries exposed a distinct heavy-tailed power-law trend in cumulative wealth distribution [22]. The well-known Pareto principle is also connected to his observations. According to the Pareto principle, a p fraction of the society owns $1-p$ fraction of the total wealth distributed among the society. In the time of Pareto, the value of p was approximately 0.2, meaning that 20% of the society owned 80% of the total wealth.

Beginning in the last decade of the 20th century, a new field of interdisciplinary science, called econo-physics, emerged, blending elements from both physics and economics. Several models were designed to describe in a unified manner the whole range of the wealth distributions [18, 23–25]. Nonetheless, even with the collaboration between physicists and economists, as far as our knowledge goes, a comprehensive model for describing social inequalities has not been developed until recent times [18, 19, 23, 26].

A general approach, for describing the dynamics of economic systems is based on master equations, which are the evolution equations of a given system in the probability space [18]. In this thesis, we present a modeling framework based on the LGGR model that proved to be suitable for filling this gap in the toolbox for modeling inequalities. For applying the LGGR model to model socio-economic systems, we consider the following assumptions to be valid for the society as well: (1) we assume that the direct interactions on the network can be approximated by a mean-field interaction captured within the growth ($\sigma(x)$) and reset ($\gamma(x)$) rates; (2)e assume that all individuals are identical, and no differentiation can be made.

3.1 Experimental study and challenges

Good quality data is necessary for designing mathematical models that would be able to explain the causality behind the phenomenon. Nowadays, due to easy access to the internet, an immeasurable amount of digitized financial data for different countries is available [27, 28]. The monitored socio-economic measures are in general considered both for individuals and groups of individuals (families, organizations, settlements, etc.) as well [29–31]. The distribution of these may vary depending on whether it is considered for individuals or groups of people. So it is in the case of wealth data as well. Wealth cannot be measured using a simple quantity, since it might be composed of multiple components. With the increasing complexity of the society, it became nearly impossible to do an exhaustive mapping of it. Sampling procedures for wealth tend to be inaccurate, the available wealth data commonly involves estimations and annual surveys [28]. Sampling can be very dangerous because the chosen samples do not necessarily represent properly the entire group they belong to. An illustrative representation of sampling and its deficiency is illustrated in Figure 3.1. So it is quite a challenge to find representative data characterizing ensembles of individuals.

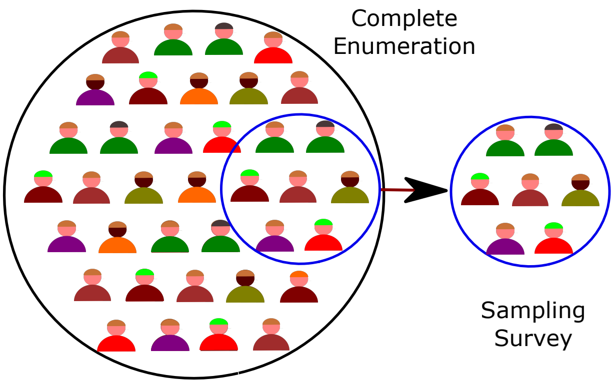

Figure 3.1: Illustration of ensemble sampling deficiency.

Chapter 4

Wealth in modern societies

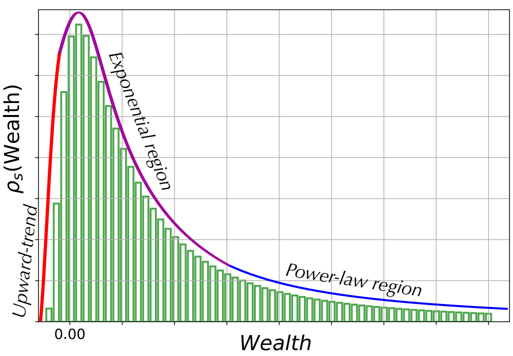

Figure 4.1: Ilustration of the three regions of the wealth distribution.

Nowadays the distribution of wealth has been closely monitored by governmental statistical offices and numerous organizations [27, 32], providing accurate wealth data. It is important to note that in modern societies, in the case of wealth, negative values are possible as well, meaning depts. We differentiate three regions based on the wealth distribution's shape, as illustrated in Figure 4.1. The first region is an increasing exponential region at negative and very small wealth values. The middle part of the wealth distribution is an exponential region with a negative trend. The uppermost part of the wealth distribution follows a power law-like decaying trend. In the relevant literature however, the first region is not identified as a separate region, since dept is not considered in the distribution [23, 26]. In the literature, instead of the part that we consider the middle, exponentially decaying region, a Boltzmann-Gibbs-like distribution is considered [23, 26]. The other shortcoming of the existing models is that they demonstrate the two regions separately. The main ambition of our work was to design an analytically tractable, dynamical model that is suitable for demonstrating the observed wealth distribution in its integrity (negative and positive regions of the experimental wealth data).

The focus of this section is on applying the LGGR model to depict the stationary wealth distribution in the USA, Russia, and France at the beginning of the 21^{st} century. We assume that the dynamics of wealth is in a quasi-equilibrium condition based on the following two circumstances: (1) the long enough existence of free-market economy in these countries for the past few decades; and (2) the substantial size of these economies, ensuring their resilience to minor fluctuations in the global economy.

The linearly increasing growth rate ($\sigma(x)$) can be considered for wealth, consistent with the preferential wealth accumulation phenomenon. This growth rate, expressed by the first line of Equation 4.1, indicates positive growth within the $[-g, \infty)$ interval (g being a positive constant). Financial growth begins primarily at low wealth values and concludes at higher wealth levels. Hence, the reset rate ($\gamma(x)$) should be negative for low and for negative wealth and saturate at a positive value for high wealth values, preventing the possibility of infinite wealth. Such a reset rate can be formulated as the second line of Equation 4.1.

$$\begin{aligned} \sigma(x) &= l \cdot (x+g) \\ \gamma(x) &= l \cdot \left(s - \frac{v}{x+g}\right), \ (s, v \in \mathbb{R}_+). \end{aligned} \tag{4.1}$$

Considering these forms of growth and reset rates, we calculate the form of the stationary probability density function given by the LGGR model (Equation 2.4). We rescaled the x variable with its expected value ($y = x/\langle x \rangle$) calculated analytically from the stationary probability distribution function $\rho_s(x)$. The obtained form of the renormalized probability distribution function $\rho_s(y)$:

$$\rho(y)_s = \frac{c\,(c+1)^s\,(s-1)^s}{\Gamma(s)} e^{-\frac{(c+1)(s-1)}{1+c\,y}}\,(1+c\,y)^{-1-s}. \tag{4.2}$$

where $c = v/[g(s-1)] - 1$.

The rescaling of the probability density function to be normalized when the expected value $\langle y \rangle = 1$ is practical not only because it allows us to cancel out a parameter but also because it facilitates visual comparison of probability distribution functions obtained for different countries (refer to Figure 4.2).

4.1 Validating the model for the USA, Russia, and France

Wealth-related data was obtained from the World Inequality Database (WID) [28]. Our study utilized data from multiple years for the United States of America, Russia, and France. The data supplied is presented as cumulative percentile fractions representing the population and their corresponding wealth in the local currency of the countries. From these cumulative distributions, we derived the normalized wealth distributions. We computed the Probability Density Function (PDF) for normalized wealth, where wealth values were rescaled with the average wealth $\langle Z \rangle$ of each respective year, expressed as $z = Z/\langle Z \rangle$. Employing this approach resulted in the collapse of wealth distributions for both countries, the USA and Russia, across various years onto a unified curve. This collapse of the PDFs allowed the calculation of the average wealth distribution for both countries.

Both distributions (USA and Russia) exhibit strikingly similar trends, allowing for a more comprehensive comparison when their averaged patterns are overlaid on a shared graph, Figure 4.2. Figure 4.2 indicates a reasonably accurate fitting for the en-

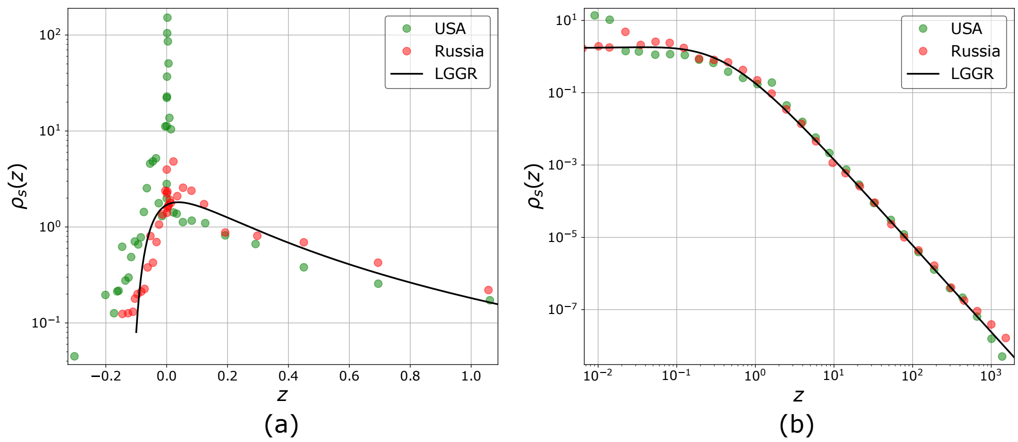

Figure 4.2: Comparison of the distribution function of the normalized relative wealth ($z = Z/\langle Z \rangle$) from the USA and Russia (represented with dots) and the probability density function given by the LGGR model in the form of Equation 4.2 (continuous black line). The experimental distributions are the averaged distributions over the considered years. The negative region of the distributions is represented on log-normal scale (a), while the positive region is represented on log-log scale (b).

tire wealth spectrum using parameters $s = 1.4$ and $c = 6.5$ within Equation 4.2, the analytical form of the probability density function. Interestingly, despite the significantly different economic histories of the USA and Russia, the distribution of relative personal wealth exhibits a universal trend.

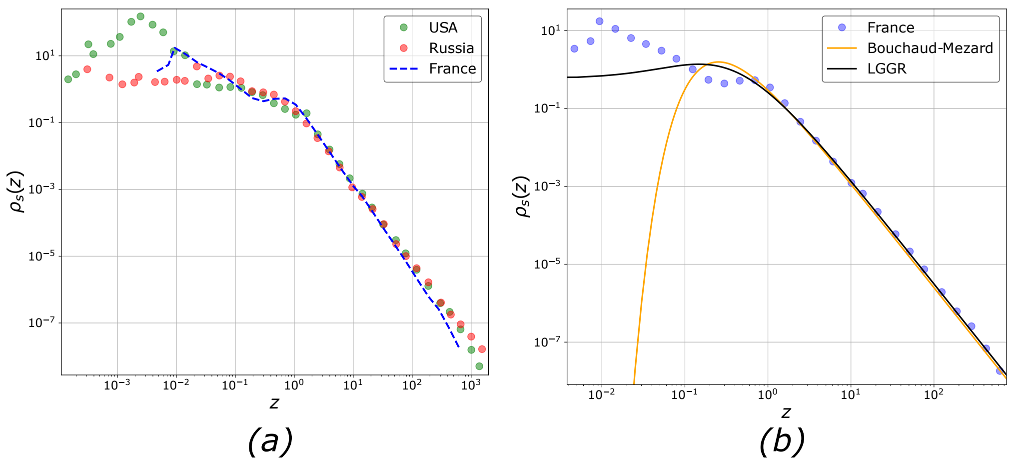

Figure 4.3: (a) Comparison of the experimental probability density functions (averaged over multiple years) of the relative wealth values ($z = Z/\langle Z \rangle$) from the USA, Russia, and France. (b) Comparisons between the best fits generated by the LGGR model (Equations 4.2) and the Bouchaud and Mezard model [24] to the experimental wealth distribution for France.

We extended the analysis to the wealth distribution of France [28], it became evident that this trend does not universally apply, as shown in Figure 4.3a. The wealth distribution for France suggests a higher equality among the population compared to the USA and Russia, indicated by the faster decay of the wealth probability distribution function presented in Figure 4.3a. For France, similarly to the USA and Russia, the distributions calculated for different years collapse. In the French dataset, there are no recorded instances of negative wealth, so only the positive wealth region ($z > 0$) is plotted on a log-log scale in Figure 4.3.

We also compared our modeling results with a generally accepted model for wealth distributions, the Mezard-Bouchaud model [24]. Our model shows better alignment with experimental data in the region of small wealth regions. This is alos illustrated in Figure 4.3b.

4.2 Remarks on the kernel functions

To test the preferential growth rate, we examined the wealth growth of the top 15 wealthiest individuals consistently featured in the Forbes list from 2001 to 2019. We tracked the wealth $Z_i(t)$ of each individual in each year t, relative to their wealth in 2001, denoted as ${}^rZ_i(t) = Z_i(t)/Z_i(2001)$. The average increase ${}^rZ(t) = \langle {}^rZ_i(t) \rangle_{\{i\}}$ over time is illustrated in Figure 4.4a using log-normal scales. The seemingly linear pattern (black curve) observed in Figure 4.4a supports the exponential trend in wealth growth: ${}^rZ(t) \approx \exp[-0.075 \cdot (t - 2001)]$.

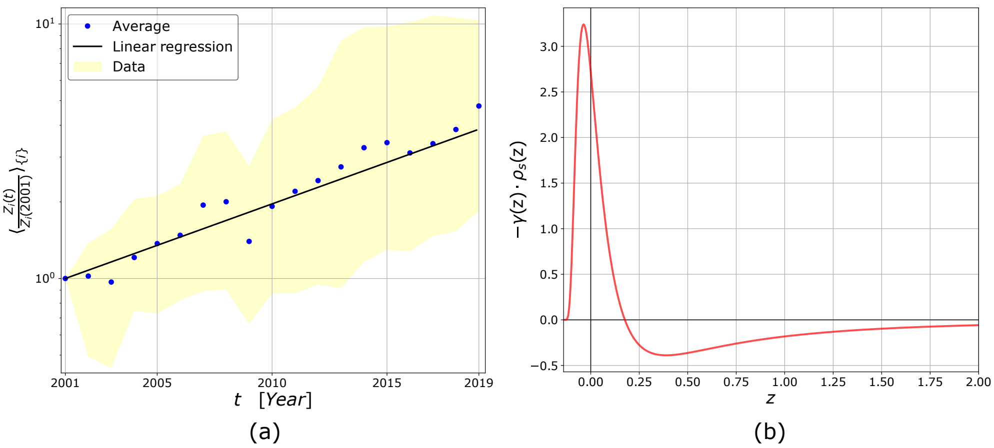

Figure 4.4: (a) The logarithmic y-axis depicts the exponential growth trend of the averaged rescaled wealth, ${}^rZ(t)$, among the world's wealthiest individuals from 2001 to 2020. The transparent curves in the background represent the wealth growth of the 15 richest individuals separately. (b) Changes per unit time in the population fraction within a unit wealth interval ($n(z) = N(z, z + dz)/dz$) around z as a consequence of the smart reset process (see Equation 4.2). The model parameters correspond to those obtained from the fitting of the experimental density functions in Figure 4.3 ($s = 1.4$ and $c = 6.5$).

We lack experimental data to substantiate the chosen kernel function for the reset rate. Nevertheless, based on the reset kernel function and the form of the PDF from Equation 4.2, we can calculate the changes per unit time in the population fraction within a unit wealth interval ($n(z) = N(z, z+dz)/dz$) using the $\frac{dn(z)}{dt} \propto -\gamma(z)\,\rho_s(z)$ product.

We used the fit parameters appropriate for fitting the experimental wealth distributions in Figure 4.3 ($s = 1.4$ and $c = 6.5$), and plot the product of the reset rate and the probability density function multiplied by -1 ($-\gamma(z) \cdot \rho(z)$) in Figure 4.4b. We observe that most individuals initiate their wealth dynamics within the range of small and negative values ($\gamma(z)\,\rho(z) > 0$), significantly lower than the average wealth in society ($z < 0.2 \cdot \langle z \rangle$). By the time when they exit the system (positive reset rate), their wealth is in general higher. This aligns well with our common experiences in everyday life.

4.3 General discussions and conclusions

Our aim was to create a comprehensive mathematical formula that encapsulates wealth distribution across all categories, including negative values representing debts. We observed that preferential growth (Equation 4.1) is substantiated by the wealth dynamics of top billionaires. Our assumed smart reset rate (Equation 4.1) aligns with real-world observations where younger individuals typically start with debts or low wealth (negative reset rate), progressing to higher wealth categories over time. This perspective reinforces our belief that the rates we have employed suitably address the dynamics of wealth within the society using a mean-field approach. The good agreement between the collected data with the probability density function given by Equation 4.2 shows that this model provides a strong representation of wealth distributions observed in contemporary societies. The probability density function resulting from our model possesses two main advantages relative to the Bouchaud and Mezard approach: (1) it accommodates negative wealth values (debts); (2) with its two free parameters (Equation 4.2), allows for a refined fit for the low wealth region. Our analysis unveiled unexpected similarities between wealth distributions in the historically distinct economies of the USA and Russia. The observation that the rescaled wealth distribution collapsed for the USA and Russia but not for France suggests that there is no singular universal trend in wealth distributions across different countries, despite apparent similarities.

Chapter 5

Wealth distribution within a small community

5.1 Owerview and motivation of the study

In the preceding section, our focus was on examining social inequalities by analyzing wealth distribution at the scale of large countries (USA, Russia, France). In this section, we shift our focus backward, delving into a smaller human society comprising around a thousand households. Here, our aim is to once again model social inequalities by measuring wealth. For the work presented in this section, we accessed an exhaustive wealth database from a small Romanian commune, *Comuna Sâncraiu* (*Kalotaszentkirály*). The database has been digitized from ownership and taxation records stored at the local authorities of Sâncraiu commune, Romania. As an extension of our prior work, we explore various time periods that represent distinct systems of authority. The database contains information from three distinct years: 1961, 1989, and 2021. In 1961, it captures the period just before collectivization, marking the time when lands previously allocated to peasants by the communists post-1947 were once more collectivized. The year 1989 marked the culmination of the communist regime in Romania, while 2021 represents the present, encapsulating over three decades since the fall of communism and reflecting the effects of the transition to a free-market economy in Romanian communes.

In this work, we demonstrate the applicability of the LGGR model for a relatively small society. Furthermore, by having data from three different economic ages of Romania, we are able to compare the population of this commune from the point of view of social inequality. In order to do this we compare the applied growth and reset rates in the LGGR framework and also compute the Lorenz curves, Pareto points, and the Gini indices for all three years [26].

5.2 Collected wealth data

Anonymized records on household wealth were obtained, from the local authorities of Sancraiu (in Hungarian: Kalotaszentkirály-Zentelke) commune, through an agreement between Babeş-Bolyai University of Cluj-Napoca. The population of the commune shows a decreasing trend over the approximately last 70 years. The 1956 census counted 3557 inhabitants [33]. 36 years later, the 1992 census listed 2053 individuals [33], while the commune's current population stands at 1628 inhabitants according to the 2011 census survay. In terms of households in 1961, there were 1133 households, for 1989, this number decreased to 921. While in 2021, the number of taxpayers was 1595. The information regarding wealth components for 1961 and 1989 was extracted from anonymous agricultural registers maintained by the mayor's office. These registers contain comprehensive information about every household, documenting the land owned, the size of the house and auxiliary buildings, as well as the count of livestock owned by each household. Given the commune's historical reliance on agriculture, it can be assumed that these records held the majority of a household's assets before 1990. The data for 2021 were sourced from an anonymized taxation database specifically detailing land and building information.

To estimate the total wealth of a single household i, we computed a linear combination of all recorded valuables as defined by Equation 5.1. To create an unbiased proxy, we employed ten sets of weighting factors to estimate total wealth, assuming the actual value lies between the minimum and maximum estimated values. This estimation expresses wealth in an arbitrary quantity (in the figures we denote it as $[a.u.]$), rendering it incomparable across different studied years.

$$W_i = \sum_{\{j\}} M_{j,i} \cdot P_j, \tag{5.1}$$

where j represents the valuable categories, and P_j denotes the corresponding weighting coefficient. The P_j weighting factors fulfill the normalization condition: $\sum_{\{j\}} P_j = 1$.

To provide a better overview of the applied estimation method and support the assumed realistic nature of the weighting parameters, we calculated the percentile composition of the total wealth of the commune. This calculation considered the valuable components included in the households' wealth estimation. The percentile share of a single valuable category from the total wealth possessed by the entire commune can be calculated as follows:

$$S(j) = \frac{P_j \sum_{\{i\}} M_{j,i}}{\sum_{\{i\}} W_i} \cdot 100 \quad [\%], \tag{5.2}$$

where i represents the summing over the households, while j indexes the different

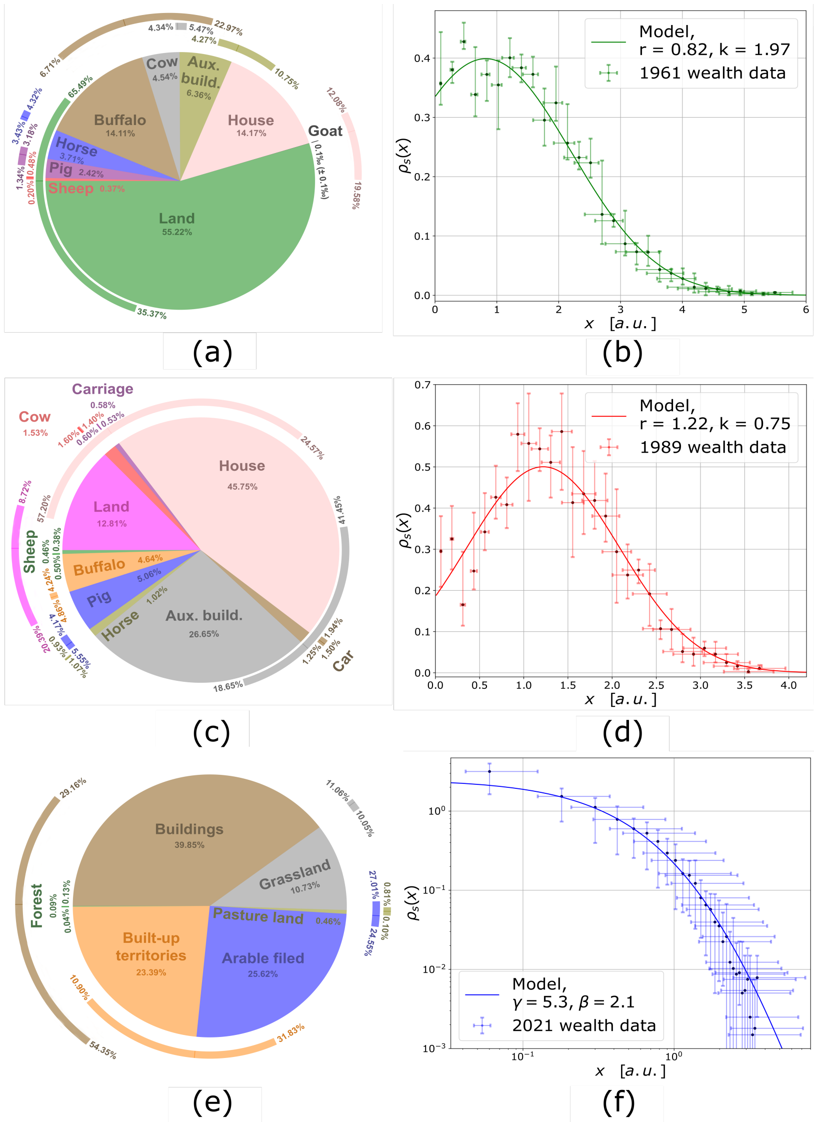

Figure 5.1: The panels from the left column illustrate the composition of the commune's total wealth in (a) 1961, (c) 1989, and (e) 2021. The shares of the different categories ($S(X)$) are calculated based on Equation 5.2. The panels from the left column illustrate experimental probability density for wealth in (b) 1961, (d) 1989, and (f) 2021. Error bars are obtained by combining the results of the different weight parameter sets. Wealth values are given in arbitrary units (a.u.) as it is explained in Section 5.2. The theoretical distributions (continuous lines) defined by Equations 5.4 (b, d) and 5.4 (f) are fitted to the averaged experimental distributions (black dots). The fitting parameters appear in the legend of the figures.

valuable categories. For each valuable j and every set of considered weighting factors P_j, we calculate the quantities $S(j)$ and visually represent them for each year under study. In Figures 5.1a,c,e, we represent the percentile composition of the communes total wealth (calculated based on Equation 5.2) in the tree analyzed years. The different wealth components are listed in the figures. The shaded areas at the borders of different colors represent the uncertainty induced by the differences between the weighting sets. The relatively large thickness of these diffuse transition areas in the figure suggests that the chosen weighting sets cover a considerably different set of estimation methods ensuring the robustness of the estimation. Additionally, we compute the experimental wealth distributions from the estimated household wealth for each set of weighting parameters. In Figures 5.1b,d,f, we represent the bin means and extreme points (on both axes) with dots and error bars.

5.3 Wealth distributions through the LGGR model

Considering that the initial two years (1961 and 1989) mark the end of distinct socio-economic conditions, and the most recent year, 2021, signifies a situation that emerged as a result of 32 years of a free-market economy after the collapse of communism, we confidently infer that the distribution of household wealth has attained stationarity. In the data processed in this chapter, there is no negative wealth. This somewhat simplifies the modeling of the estimated wealth distributions. By the definition of the realistic kernel functions for the growth and the reset the computation of the stationary probability density functions based on Equation 2.4 is possible. Subsequently, we fine-tune the model parameters to achieve a satisfactory alignment between the experimental data and the model.

Wealth distribution in controlled economic system: 1961 and 1989

In communism, economic growth was facilitated, but it was also moderated by the government. Taking into account these facts we approximated the economic growth of the households with a constant value ($\sigma(x) = k$) meaning that the evolution of wealth is not dependent on the actual wealth of the household.

Regarding the reset rate, we consider again the fact that the economic wellness of the population was controlled by the government. Based on the ideology of communism, there was an ideal average of wealth characterizing the society and the goal of the government was, to maintain the financial equilibrium within the society. In order to achieve such a scenario in the LGGR framework, we opted for a linear reset rate that is shifted relative to the 0 wealth value with a constant: $\gamma(x) = x - r$. This simple function takes negative values for wealth (x) below r and it becomes positive for $x > r$, ensuring a smart reset scenario, illustrated in Figure 2.1a. This ensures that new families are appearing with less wealth than r and when they become richer than this optimal wellness it becomes probable for them to lose their wealth or exit the system.

Considering these rates, the stationary solution of the LGGR model (Equation 2.4) leads to a normal distribution that is normalized on the $[0, \infty)$ interval (Equation 5.3), since negative wealth is not present in the data.

$$\rho_{s1}(x) = \sqrt{\frac{2}{k\pi}}\, e^{-\frac{r^2}{2k}}\, e^{\frac{-x(x-2r)}{2k}} \left[\mathrm{erf}\left(\frac{r}{\sqrt{2k}}\right) + 1\right]^{-1}, \tag{5.3}$$

where erf() denotes the error-function ($\mathrm{erf}(x) = \frac{2}{\sqrt{\pi}} \int_0^x e^{-t^2} dt$). The maxima of this normal distribution is at the value of the parameter r, which is in agreement with the explanation given before regarding the existence of an ideal wealthiness in communism. The probability density function defined by Equation 5.3 is appropriate for fitting both experimental wealth distributions from 1961 and 1989. For 196, the experiment and theoretical wealth distributions are presented in Figure 5.1b, while for 1989 in Figure 5.1d. In both cases, the fit parameters are listed in the legends of the figure.

Wealth distribution in the free market: 2021

In the free-market economy, financial growth can also be presumed preferential: $\sigma(x) = x+\beta$. Considering only positive wealth within the system, the simplest option for the reset rate is to assume it is constant: $\gamma(x) = \gamma$. This implies that regardless of a taxpayer's wealth, the probability of being reset within a given time window remains the same. These forms of the rates (using here again Equation 2.4) lead to a Tsallis-Pareto type stationary probability density function defined by Equation 5.4.

$$\rho_{s2}(x) = \frac{\gamma}{\beta}(1+\frac{x}{\beta})^{-1-\gamma}, \tag{5.4}$$

The visual comparison of the theoretical and experimentally estimated wealth distributions is presented in Figure 5.1f.

5.4 Social inequality based on known measures

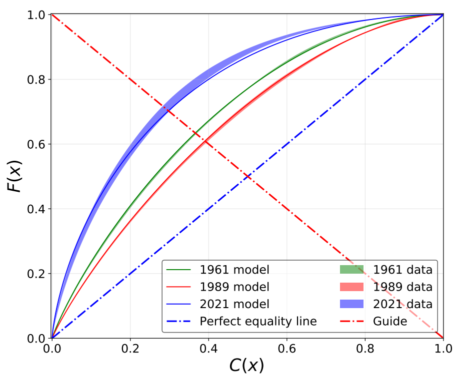

Figure 5.2: Lorenz curves, both experimental and theoretical, for the years 1961, 1989, and 2021. The shaded area represents the range covered by the experimental outcomes across various weight parameter sets. The theoretical curves, derived from the fitted distributions, are depicted by solid bold lines.

First, we computed the Lorenz curves [34] for each studied year based on both the theoretical curves and the experimental wealth distributions. Figure 5.2 depicts the Lorentz curves corresponding to the analyzed years. The Lorenz curves built from the modeled probability density functions considering the established optimal fitting parameters are indicated with solid lines. The regions bounded by the extremes of the experimentally observed Lorenz curves, corresponding to various weight parameters, are shaded with the respective colors. The theoretical Lorenz curves align well with the experimental ones, just like the wealth distribution functions. We calculated the Pareto point values and the Gini indices from both the experimental wealth data and the theoretical distributions. We established intervals for the experimentally obtained values for each of the three years, finding that the values derived from the modeled curves fell within the defined intervals. The values of the experimental and theoretical Pareto points and Gini indices are listed in Table 5.1.

Year	G		P	
	Data	Model	Data	Model
1961	[0.377;0.379]	0.378	[0.366;0.368]	0.367
1989	[0.304;0.315]	0.312	[0.390;0.395]	0.391
2021	[0.543;0.579]	0.552	[0.282;0.299]	0.298

Table 5.1: Inequality metrics across the examined years: the Gini coefficient (G) and Pareto point (P). Boundaries derived from the data using varying weight parameters and the value extracted from the fitted probability density function.

5.5 Chapter summary

This study provides a second confirmation of the applicability of LGGR dynamics in modeling wealth distribution within socio-economic systems. The distinctive aspect of this research lies in applying this dynamics with different growth and reset rates to different economic scenarios, proving the model's effectiveness in both cases.

The experimentally and theoretically determined inequality indices exhibit substantial agreement (Figure 5.2), providing additional confirmation of the validity of the proposed probability density functions from a different standpoint. The Gini and Pareto measures' values indicate minimal social inequalities during the communist regime. In 1961, the Gini stood at approximately 0.38, dipping to 0.31 by 1989. However, with 32 years of a market economy in this region, the Gini rose significantly to about 0.55. This suggests the current intensification of social inequalities. Besides demonstrating the modeling capability of the LGGR model, we also provided valuable inequality data for a well-defined geographic region spanning different historical periods of Romania. Conducting similar studies in diverse regions worldwide within well-delimited communities might be also interesting.

Chapter 6

Stationary LGGR dynamics of Lotteries

6.1 Introduction

In this chapter, we present a study that is also strongly related to the field of social and economic sciences, yet distinctly different from the preceding two topics. Global forms of gambling, have reached such a high level of complexity that it is appropriate to investigate them as complex systems. Lotteries, in general, are widely studied from numerous viewpoints like mathematical, dealing with winning chances [35]; psychological, trying to explain the psychological motivation of playing [36]; economical, tailoring their regulations to enhance state revenue [37]. However, it has not been the focus of statistical physics, therefore this simple approach possesses novelty.

In lotteries, participants gain the chance to win a Jackpot prize by selecting a set of numbers within a specified range. The selection of winners is based on periodic random number draws, usually taking place on a weekly or bi-weekly basis. In the case where a ticket bears the same numbers as those drawn, the possessor becomes the winner. In the event of no winner, the prize pool continues to grow through successive draws until a winner is ultimately announced [35]. The Jackpot's escalation following successive unsuccessful draws contributes to higher Jackpot values, that naturally attract more players to participate [38]. This leads to accelerated growth until the Jackpot is won. Participants may qualify for smaller prizes if they match fewer numbers than needed to win the Jackpot but still exceed a minimum threshold. Funding for these smaller prizes, depending on the specific lottery rules, may originate from either a portion of the Jackpot prize pool or be covered by a designated percentage of the weekly sales.

Although the rules may seem simple, the combination of player behavior, the random nature, and additional rules of the game emerge into complex dynamics. The growth and reset subprocesses can be clearly identified in Figure 6.1, which displays a segment of the Jackpot value time-series for the Powerball lottery [39, 40]. The depicted time series of Jackpot values entirely characterizes the dynamics. The dynamics governing lottery Jackpot values, align with the principles underpinning the LGGR mean-field type dynamical model.

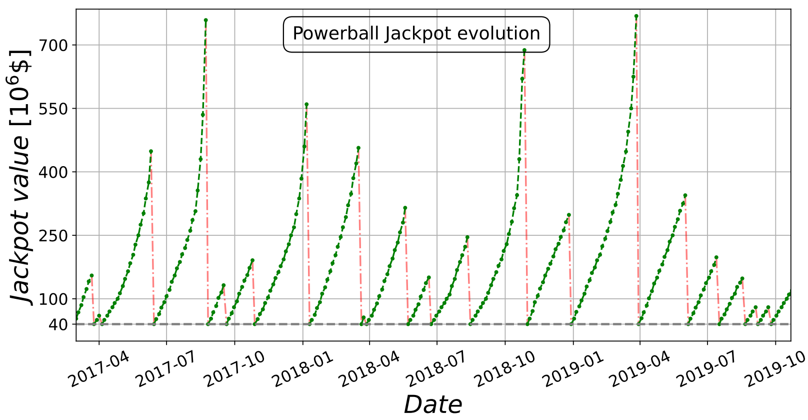

Figure 6.1: Time-series data depicting the Jackpot prize for Powerball spanning from March 2017 to July 2019 [40].

6.2 Model applicability and Jackpot value time-series data

Jackpot time-series data for six lotteries, each with distinct rule sets and player-pool sizes, was collected and processed [39, 41–45]. We collected the complete time series for each of the six lotteries from the launch of the game until the present. However, we focused only on shorter time periods during which neither the rules nor the territory had been altered.

We have listed in Table 6.1 the selected time periods for analysis and modeling, for each of the six lotteries. We have also written in separate columns the chance of winning the Jackpot and the rules that define each lottery: the number of numbers that need to be chosen and the size of the set of selectable numbers. We also present the size of the player pool and some statistical data regarding the mean Jackpot value, and average time between consecutively winning the Jackpot. Additionally, alongside the above-mentioned characteristics of the lotteries, we also list the value of the product

$w = P_J \cdot Pop$ for each lottery in a separate column. This simple parameter might provide valuable information about the lottery. Some of the data sources for each lottery are also listed in Table 6.1. For the Powerball and UK lotteries, the visual representation of the time series considered for modeling can be seen in the (A) sections of Figure 6.2.

While modeling the distribution of Jackpot values in this study, we exclusively refer to the distribution of the rescaled Jackpot values by the average value of the Jackpot over time ($J \to x = J/\langle J \rangle_{\{t\}}$). The LGGR model is applicable only to ergodic Markov processes. Therefore, first, we demonstrate the ergodicity of the six introduced time series. The stationarity of the studied time series, as manifested through the convergence of the mean Jackpot value across an extending time window $\langle x(t) \rangle_t \approx c$, is illustrated in sections (B) of Figure 6.2. Their aperiodic nature is demonstrated by the absence of autocorrelation ($ACF(s) = Corr(x(t), x(t+s))$), as depicted in section (C) of Figure 6.2. The same holds for the other four lotteries as well. Based on this argumentation we consider the lotteries under investigation as ergodic.

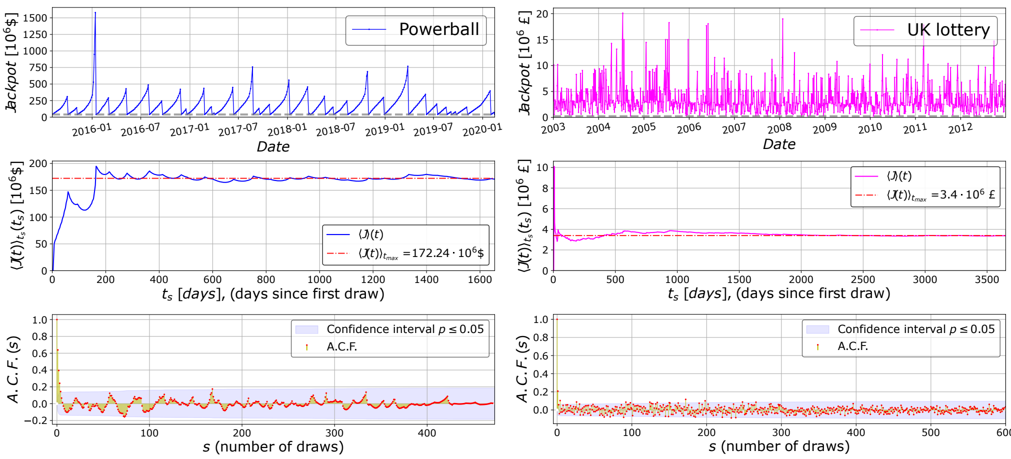

Figure 6.2: The time series plots of the studied lotteries are presented. Each row corresponds to the following: (A) The time series plot of the lotteries within a stable period, where rules and player pools remain constant. (B) The mean of the time series $\langle x(t) \rangle_t$, accompanied by the convergence value. (C) The autocorrelation function of the time series $ACF(s)$ up to s days, along with a confidence interval with $p \leq 0.05$.

6.3 Modeling the growth and reset dynamics

We established the mathematical form of the growth and reset kernel functions through analysis of the experimental data. This approach confirms the internal coherence of the proposed model.

The growth rate

The Jackpot's value accumulates from a fraction of the sales (which is determined by the number of sold tickets) between two consecutive draws. Additionally, as a natural consequence of the Jackpot value on players, the higher the Jackpot, the more tickets are bought by players. Hence, a preferentially increasing form of the growth rate would be natural to assume: $\sigma(x) = a(1 + b \cdot x)$.

To ensure the validity of such a multiplicative Jackpot growth, we have computed the actual growth function (the average relative increase in Jackpot values between

Lottery	Period	Format	P_J (≈)	Pop. (≈)	⟨J⟩ (≈)	⟨t_J⟩ [draws] (≈)	w (≈)	Source
Powerball	08.05.2015-02.15.2020	5/69 + 1/26	$1/292 \cdot 10^{-6}$	$292.2 \cdot 10^6$	$172.2M.\$$	14	1	[39, 46]
Megamillions	06.22.2005-01.31.2010	5/56 + 1/46	$1/258 \cdot 10^{-6}$	$152.4 \cdot 10^6$	$66.8M.\$$	9	0.58	[42, 47]
Euromillions	01.12.2012-09.01.2016	5/50 + 2/9	$1/116 \cdot 10^{-6}$	$272.0 \cdot 10^6$	$46.4M.€$	5	2.34	[41, 48]
Canada lotto	03.03.2007-02.27.2019	6/49	$1/14 \cdot 10^{-6}$	$35.2 \cdot 10^6$	$9.8M.\$$	3	2.70	[44]
UK lotto	01.07.1995-09.09.2023	6/49	$1/14 \cdot 10^{-6}$	$62.5 \cdot 10^6$	$3.4M.£$	2	4.80	[45]
Texas lotto	07.14.1994-07.18.2000	6/50	$1/17 \cdot 10^{-6}$	$18.9 \cdot 10^6$	$12.6M.\$$	4	1.12	[43, 49]

Table 6.1: The lotteries examined in this article, within timeframes of consistent rules. The table includes information on the lottery format, the likelihood of winning with a single ticket (P_J), the potential maximum player population (Pop.), and some statistical properties such as the mean value of the Jackpot ($\langle J \rangle$) and the average time between consecutive Jackpot wins ($\langle t_J \rangle$). Additionally, the lottery parameter w, as discussed in Section 6.2, is also provided. In the header, the symbol "≈" suggests that the values in the column are only approximately accurate.

successive draws) based on the time-series data for each lottery. In Figure 6.3a, we present the experimentally derived growth rate for the Powerball lottery. The uncertainty associated with each data point's position is indicated in the figure, demonstrating the standard deviation of the data in the respective bins along both axis. For each lottery, we exclusively determine parameter values using data from the lower region depicted in the graphs (experimental growth rates of other lotteries are included in the thesis). The experimentally determined growth rate in Figure 6.3 supports the chosen linear form for the growth kernel function.

The reset rate

There is a connection between the growth and reset rates through the number of sold tickets. Calculating the reset probability $\gamma(y)$ for a specific number of sold tickets, denoted as y, requires accounting for the number of distinctively completed lottery tickets. The number of sold tickets (y) is proportional to the increment of the Jackpot value between consecutive draws, which is the growth rate itself: $y = l \cdot \sigma(x)$, where l is a positive parameter mapping the growth to the number of thickets. The reset rate as a function of the Jackpot value is given by Equation 6.1.

$$\gamma(x) = \left(1 - \beta e^{-\kappa(1+b\,x)}\right), \tag{6.1}$$

with $\kappa = l \cdot a \cdot \alpha$. The parameter l is hidden in κ. The connection between growth and reset brings in two more model parameters: α (or $\kappa = l \cdot a \cdot \alpha$) and β. The nature of the reset process (conventional or smart) is controlled by the parameter β.

To experimentally calculate the reset rate we had enough data only for the Canadian and UK lotteries. Figure 6.3b illustrates the experimentally determined reset rates for the UK lotteries, plotted against the rescaled Jackpot x. We emphasize again that the final few bins in these plots contain a small number of data entries, making this section of the plot less reliable. To demonstrate the consistency, in Figure 6.3b we visually compare the experimental reset rates with Equation 6.1. The limited qualitative agreement can be attributed to the poor quality of the data.

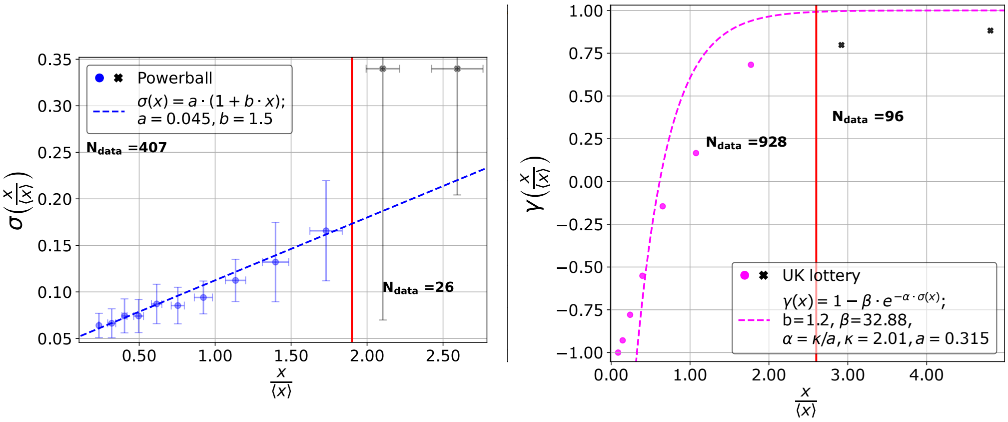

Figure 6.3: (a) The experimental growth rate, denoted as $\sigma(x)$ and expressed as a function of $x = J/\langle J \rangle_{\{t\}}$. The data averages are accompanied by error bars representing the ensemble's standard deviation. The chosen growth kernels ($\sigma(x) = a(1 + b \cdot x)$) with the a and b parameter values used to fit the relative Jackpot distributions is presented alongside the experimental results. (b) A comparison between the experimental reset rate and its theoretical counterpart for the UK lotteries. The theoretical reset rate is based on parameters obtained from fitting probability density functions and the growth rates. The limited qualitative agreement can be attributed to the poor quality of the data. We provide the number of data points, N_{data}, utilized for calculating the averages. This presentation aims to emphasize that the regions to the right of the red vertical lines should not be considered reliable for statistical inference.

6.3 Statinary probability density function

Relying on the experimentally supported linear form of the growth rate and the derived relation between the reset and growth rates, the analytical form of the stationary probability density function:

$$\rho_s(x) = C\,(1 + b\,x)^{-\lambda-1} e^{\nu\,\mathrm{Ei}(-\kappa(1+b\,x))}, \tag{6.2}$$

where C is the normalization constant, and the parameters λ and ν are derived from the parameters appearing in the growth and reset rates: $\lambda = 1/(a \cdot b)$, $\nu = \beta \cdot \lambda$. $\mathrm{Ei}(x)$ in the exponent is the Exponential Integral function, defined as $\mathrm{Ei}(x) = -\int_{-x}^{\infty} \frac{e^{-t}}{t}\,dt$. When $\beta = 1$, the reset scenario is conventional (the reset rate is positive), happening always to s minimum Jackpot value $x = J_0$. When $\beta > 1$ the reset rate can be negative (for smaller x values) implying that, following the reset process, the growth dynamics restarts with a Jackpot value $x > J_0$. Such dynamics can be identified in the case of the Canadian and UK lotteries. The shortcoming of the obtained form of $\rho_s(x)$ is that neither the normalization constant C nor the expected value of x can be analytically calculated. Therefore, when fitting the experimental data, the normalization and rescaling to $\langle x \rangle = 1$ of the probability density function need to be numerically performed for fixed parameter values (a, b, κ, and β).

In Figure 6.4, we present the experimental distributions of the relative Jackpot values ($x/\langle x \rangle$) alongside the probability density function described by Equation 6.2. The comparison between the experimental data and the stationary probability density function obtained from the model reveals good qualitative agreement.

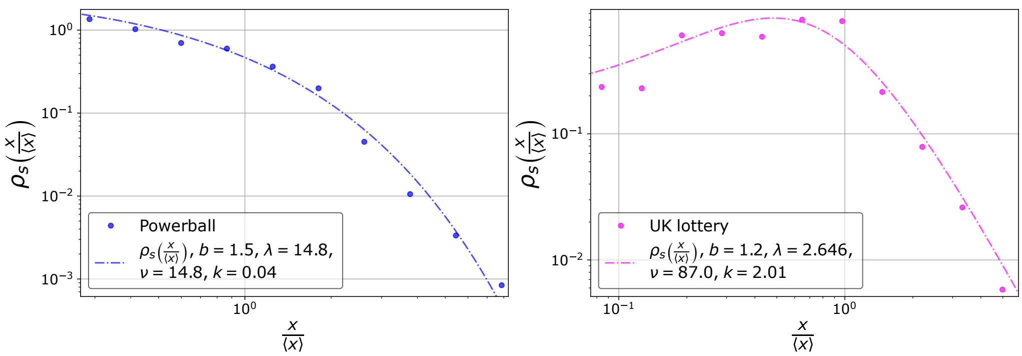

Figure 6.4: The probability distribution at equilibrium for the Jackpot prizes $\rho_s(x)$ as a function of the relative Jackpot for the six examined lotteries.

6.4 Discussion and summary

Despite variations in lottery formats, the resulting dynamics of the collective system involving players and the rules set by lottery associations are similar. The resulting complex dynamics is the same as the one mathematically modeled by the LGGR approach. To select a linear growth rate, we validated our intuition using real-world data, as illustrated in Figure 6.3. Additionally, the relation described in Equation 6.1 linking growth and reset rates was qualitatively confirmed through the experimentally computed reset rate (Figure 6.3b). Utilizing these rates, in the stationary limit, the LGGR model generated a probability distribution function (Equation 6.2) that, with appropriate parameters, aligns consistently with the computed probability density functions of Jackpot values in all six lotteries under study. The fitting process of the data was consistently conducted, taking into account both the experimental growth curves and the obtained Jackpot distributions for each lottery separately. The goodness of fit for the reset rates was not considered due to the limited amount of data available. The objective was not to achieve a perfect fit for the experimental Jackpot distributions and the supporting data. Designing a perfect fit for the experimentally obtained probability density functions is unfeasible given the limited availability of experimental data and the dynamic nature of lottery rules, which change occasionally. Our objective was to develop a statistical mean-field model that offers a consistent demonstration of the stationary fluctuations of Jackpot values. A noteworthy aspect of this model is its capacity to illustrate two distinct lottery dynamics: one characterized by a complete reset (Euromillions, Powerball, Megamillions, Texas lottery) and the other by a "smart-reset" (UK Lotto and Canadian Lotto) process.

Chapter 7

Stationary LGGR dynamics in biological systems

7.1 Introduction and motivation of the study

Biological systems can be interpreted as complex systems, because of their numerous components and the complex interactions between these. In the case of wood ecosystems, the relevant quantity whose distribution is frequently studied is the size of the trees, quantified as the Diameter at Breast Height (DBH). The fundamental ecological processes present in all biological systems, that are shaping the macroscopic view of woodland ecosystems, include tree growth, mortality, tree recruitment, and the diversification of species [50, 51]. Most of the existing models in the literature consider only the first two processes [50–52]. Although the time scale of the dynamics of wood ecosystems is much slower than the one of financial systems, for example, this dynamics is still the string that leaves its fingerprint on the view of such ecosystem. Thus, even if equilibrium conditions (stationarity) take much more time to be reached, the same modeling methodologies can be applied.

Considering the first three processes the analogy between the LGGR dynamics and these is straightforward. The last process might be less obvious to understand in the context of the LGGR dynamics, but the demonstration of its effect as a size-independent reset process is deduced in Chapter 2. By considering old enough woodland ecosystems, we assume the applicability of the stationary limit of the LGGR model. We considered modeling deciduous tree ecosystems of two different types: semi-natural forests with mature trees and ancient wood-pastures. Both of the ecosystems are geographically located in central Romania. We focused on three tree species that are present in both of these ecosystems: Oak, Hornbeam, and Beech. Besides the two ecosystems in Romania, to confirm the universal nature of our findings, we analyzed data from comparable ecosystems in Hungary [53].

According to the literature, the distribution of tree sizes in deciduous forests is primarily fitted by Gamma or Weibull distributions. Here, we introduce a mathematically straightforward and coherent modeling method that offers robust theoretical support for the validity of the Gamma distribution concerning tree stem diameters within these woodland ecosystems. Our approach comprehensively integrates the four ecological processes mentioned earlier. To highlight the contrast between size distributions of trees in quasi-natural, and mature environments versus those in artificial wood ecosystems where human influence is more pronounced and environmental conditions are more homogeneous, we also analyzed data from Poplar tree plantations in eastern Hungary. These measurements served a dual purpose. The first purpose is to illustrate the contrast in tree size distribution between these controlled ecosystems, which had not yet attained equilibrium state, and mature natural woodland environments where a stationary tree-size distribution is presumed. The second purpose is to glean insights into the growth dynamics of genetically identical trees developing in controlled settings.

In addition to the DBH datasets that we described above, we also had access to auxiliary datasets from the United States National Park Service (NPS) [54], that we utilized for revealing the form of the growth and reset functions.

7.2 Experimental tree-size distributions

Our databases contain exhaustive measurements of tree diameters at breast height, forming eight tree size datasets [55]. Based on these datasets we computed the DBH distributions for each species and ecosystems apart. In figure Figure 7.1, these DBH distributions are shown. Based on this figure one can already notice the significant difference between the DBH distributions in wood-pastures, forests, and plantations.

Scaling the DBH values with their respective average DBH ($x \to y = \frac{x}{<x>}$) led to the collapse of the resulting distribution functions (lower panels of Figure 7.1). In natural ecosystems, this converging distribution follows the Gamma distribution, while in plantations, it corresponds to the Normal distribution. This collapse of the DBH distributions, unveils an intriguing universality across these environments, providing valuable insights into their comparative dynamics.

7.3 Modeling through the LGGR framework

For modeling, we once again apply the LGGR modeling framework. As anticipated in the section describing the experimental data, the studied semi-natural forests and ancient wood-pasture environments, due to their long existence, are already in an equilibrium state.

The growth rate

To determine the mathematical form of the growth rate, we utilized data from the United States National Park Service [54]. The annual growth rate, identified from tree ring diameters, was analyzed for three tree genera as depicted in Figure7.2a. The averaged annual growth rate is plotted against $DBH/\langle DBH\rangle$ (DBH measured approximately 1 m above the ground). A suitable mathematical representation of the growth rate that aligns with the data:

$$\sigma(x) = d_1 \frac{x}{x+b}, \quad b \geq 0, \tag{7.1}$$

where d_1 and b are positive constants. In Figure 7.2, the trend fitting the experimental growth rate is defined by Equation 7.1.

The reset and diversification rates

Mortality and recruitment: Considering that the recruitment rate displays a substantial negative reset for smaller tree sizes and the mortality rate is a converging function, taking positive values for greater tree sizes [56, 57], the combined impact of these functions approximates an increasing but converging function.

Diversification: Through diversification, new young trees of various species appear to replace the dying old trees. Consequently, the number of trees belonging to a species decreases multiplicatively. This multiplicative decrease in trees of a specific

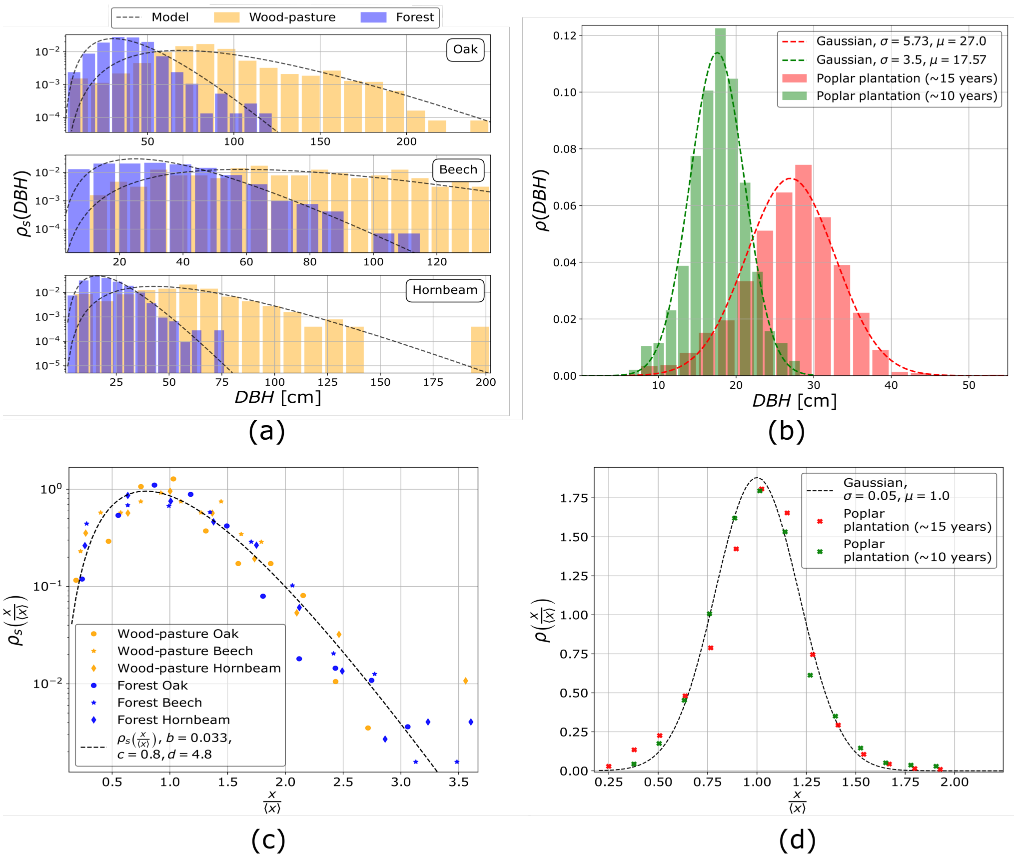

Figure 7.1: Empirical probability density functions representing DBH distributions are illustrated. In the (a) panel, the distributions for natural forests (blue) and wood-pastures (orange) are presented alongside the Gamma fit obtained from the LGGR model (refer to Equation 7.4). The (b) panel displays the DBH distribution within a 10-year-old (green) and a 15-year-old (red) Poplar plantation, both with closely aligned ecological backgrounds. The experimental distributions are fitted with Gaussian distributions. The accompanying histogram-type plot serves as an additional visual aid, emphasizing differences in mean DBH values and showcasing the presence of empty bins. The (c) and (d) panels illustrate the distributions of the relative DBH ($x \to y = \frac{x}{<x>}$) values. The (c) panel shows the DBH distribution in semi-natural woodlands (same data as in panel (a)), while the (d) panel corresponds to the two plantations (same data as in panel (b)).

species results in an exponential decline in its abundance over time. This exponential decrease in the number of trees introduces a state-independent reset-like term ($\kappa < 0$) in the evolutionary equation.

By combining the reset kernel motivated by the mortality and recruitment processes with the negative diversification rate (κ), the resulting form of the reset rate will be as follows:

$$\gamma(x) = f_1 \frac{x - r}{x + b} + \kappa \equiv d_2 \frac{x - c}{x + b}, \tag{7.2}$$

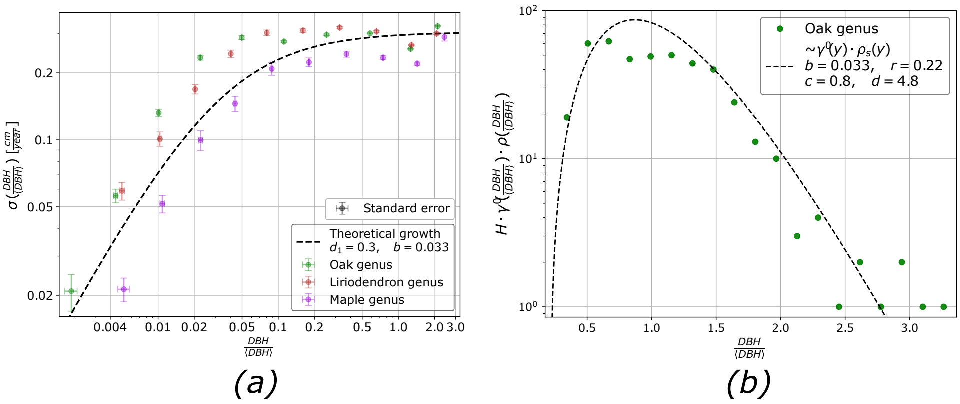

Figure 7.2: (a) The growth rate was inferred from tree ring width data, plotted on a log-log scale against relative stem diameter at one meter above the ground for three tree genera (specified in the legend). The dashed line represents the trend defined by Equation 7.1, with the associated parameters displayed in the figure. Error bars depict the standard error around the data points. (b) Alignment among the relative size distribution of dead trees, and the product of the selected reset rate and the fitted tree-size distribution. The histogram illustrates the size distribution of dead Oak trees in surveyed national parks in the USA, with sizes rescaled by the mean. The fit is defined as $H \cdot \rho_s(y)\gamma'(y)$, incorporating a constant H necessary for aligning with the experimental histogram. The fit is determined by $r = 0.22$ and parameters estimated from the experimental probability density function (Figure 7.1a). The data used here was sourced from [54].

where:

$$\begin{aligned} c &= \frac{f_1 r - b\kappa}{f_1 + \kappa} > r > 0, \\ d_2 &= (f_1 + \kappa) < f_1 \quad \text{and} \quad d_2 > 0. \end{aligned} \tag{7.3}$$

In Equation 7.3 r, b and f_1, are positive constants.

Stationary size distribution

With the consideration of the growth (Equation 7.1) and reset kernel (Equation 7.3) functions, the form of the stationary probability density function already renormalized to the mean DBH value ($\rho_s(y)$) is:

$$\rho_s(y) = \frac{d^{c\,d}}{(\frac{c}{(1-c)d})\Gamma[c\,d]} \mathrm{e}^{-dy} y^{dc-1} \left(y + \frac{c}{(1-c)d} - c \right), \tag{7.4}$$

where $d = d_2/d_1$. As depicted in Figure 7.1c, the probability density of the $x/\langle x \rangle$ distribution in forests and wood-pastures aligns, displaying a close approximation to the form given by Equation 7.4. The fitting parameters are listed in the legend of the figure.

7.4 Disscussions and summary

With data concerning the size distribution of dead trees in various mature deciduous forests, we can evaluate the adequacy of the proposed **reset rate**. Utilizing $\rho_s(y)$ (Equation 7.4), we anticipate that the shape of the distribution of dead trees' sizes will resemble the product $\rho_s(x) \cdot \gamma'(x)$, where $\gamma'(x)$ is defined by Equation 7.3 when $\kappa = 0$. To assess the validity of this reset rate, we utilized again data from [54], specifically focusing on the diameter of dead trees ($\rho_s(y) \cdot \gamma'(y)$). Figure 7.2b illustrates the Oak species' dead tree histogram. The dashed line represents a fit defined by $\rho_s(y) \cdot \gamma'(y)$, with the parameters indicated in the legend. The best-fit parameters were determined by minimizing the Root Mean Squared Logarithmic Error through iteration across a fine grid within the parameter space. Equal consideration was given to fitting the growth data (Figure 7.2a), the reset rate data (Figure 7.2b), and the DBH distribution data (Figure 7.1a) when selecting the parameters. Simultaneously, we minimized the Root Mean Squared Logarithmic Error for all three quantities.

Understanding the diversity patterns of tree sizes in natural deciduous wooded environments is a complex challenge, requiring new data and realistic mathematical models. Our contribution involves presenting fresh evidence to support the validity of the Gamma distribution [58] in deciduous forests. We have analyzed comprehensive measurement data for three tree taxa in distinct ecosystems. We have shown an intriguing statistical universality. Once tree diameters are rescaled using the average diameter for each species within the ecosystem, all datasets merge into a unified distribution. Observations from young poplar plantations indicate distinct size diversity patterns, resembling a Normal distribution. Consequently, these results imply the utilization of a Gamma-type fit of tree-size distribution as a potential indicator of woodland naturalness and maturity. We chose a simpler analytically solvable model with only two free parameters instead of a complex quantitative model with numerous unknowns. Since our primary focus was not on delivering robust statistical analyses but rather on crafting a methodically grounded analytical approach, our parameter optimizations were directed toward ensuring the consistent determination of model parameters within the framework. To strengthen confidence in the model, further high-quality data collection is crucial.

Chapter 8

General summary

As a general summary let us repeat the main presented ideas and results of this thesis. The present thesis contains four modeling studies of real-world phenomena observed in socio-economic and biological systems. The connecting pillar of these four studies is that for each of them, the Local Growth and Global reset model was applied. The thesis also provides a pedagogical presentation of the LGGR model.

The second part of the thesis contains the modeling of phenomena from socio-economic systems. Within this category, we presented the modeling of wealth distributions in leading countries with relatively stable economies and also in a small commune from Romania. The first study (Chapter 3) belonging to this topic deals with the distribution of individual wealth in the United States of America, Russia, and France. In the case of the USA and Russia, the presence of negative wealth in the data suggests the existence of debt in these societies and also the precise nature of the dataset. Our model fits the entire wealth spectrum, outperforming empirical data against the

Bouchaud and Mezard model, especially for small wealth values.

In the second study of this part of the thesis (Chapter 5) we applied the LGGR model for modeling the wealth distribution in a small Transylvanian community. The comprehensive analysis of wealth distribution patterns provides a revealing look at economic dynamics over various periods: 1961 - beginning of the communism; 1989 - the final year of communism; and 2021 - the present economic picture of the commune. The study reveals the complexities of wealth distribution dynamics across different economic systems. The collected empirical data validates the theoretical model and provides concrete examples of economic shifts.

The third chapter, although still strongly related to socio-economic systems, differs significantly from the preceding two. In Chapter 6 we used the LGGR model for describing the dynamics of lottery Jackpot values for six lotteries: Powerball, Megamillions, Euromillions, Canadian lotto, UK lotto, and Texas lotto. The emergent evolution of the Jackpot value is influenced by straightforward rules, player behavior that influences ticket sales, and a probabilistic reset rate triggered by occasional winnings. Considering the linear growth and an inherent growth-reset relation, the LGGR model generates a probability density function consistent with observed Jackpot distributions.

The last study (Chapter 7) steps out of the frame of socio-economics and targets biological systems. It delves into the complex challenge of understanding diversity patterns in tree sizes within natural deciduous wooded environments. Building on existing literature, we confirm that the diameter distribution of trees from specific deciduous species aligns with a Gamma distribution in mature natural forests. Incorporating mechanisms guiding the evolution of tree populations, in simple mathematical forms of growth and reset rates, the proposed model successfully captures this universality, suggesting consistent parameters across taxa and environments. Additionally, we study young poplar plantations, noting distinct patterns resembling a Normal distribution. This difference is attributed to the artificially created, quasi-identic conditions of each tree and the immaturity of these plantations. We suggest future investigations into growth, reset, and diversification dynamics in similar woodland environments.

The results presented here emphasize that the Local Growth and Global Reset model provides a comprehensive framework for understanding phenomena observed in socio-economic and also in biological systems.

List of publications

ISI publications:

- István Gere, **Szabolcs Kelemen**, Géza Tóth, Tamás S. Biró, and Zoltán Néda. *Wealth distribution in modern societies: Collected data and a master equation approach.* Physica A: Statistical Mechanics and its Applications, Vol. 581, 126194, (2021)

  Impact factor: 3.778

  AIS: 0.530

- István Gere, **Szabolcs Kelemen**, Tamás S. Biró and Zoltán Néda. *Wealth Distribution in Villages. Transition From Socialism to Capitalism in View of Exhaustive Wealth Data and a Master Equation Approach.* Frontiers in Physics, Vol. 10, 827143, (2022)

  Impact factor: 3.718

  AIS: 0.815

- **Szabolcs Kelemen**, Máté Józsa, Tibor Hartel, György Csóka and Zoltán Néda. *Tree size distribution as the stationary limit of an evolutionary master equation.* Scientific Reports, Vol. 14, 1168, (2024)

  Impact factor: 4.997

  AIS: 1.132

- István Gere, **Szabolcs Kelemen**, Zoltán Néda and Tamás S. Biró. *Jackpot statistics, a physicist's approach.* Physica A: Statistical Mechanics and its Applications, Vol. 637, 129605, (2024)

  Impact factor: 3.778

  AIS: 0.530

Other publications:

- **Szabolcs Kelemen**, Levente Varga, and Zoltán Néda. *Cross-correlations in the Brownian motion of colloidal nanoparticles.* Studia Universitatis Babes-Bolyai, Physica Vol. 65, 27-34 (2020)

Conference contributed talks

- István Gere, **Szabolcs Kelemen**, Tamás S. Biró and Zoltán Néda. (24-26 August, 2022) *Wealth inequality patterns based on exhaustive sampling. Data mining and modeling.* Econophysics Colloquium 2022

- **Szabolcs Kelemen**, István Gere, Tamás S. Biró and Zoltán Néda. (17-18 November, 2023) *A brief review of the LGGR model and its possible applications.* ICAS 2023: The 16th International Conference on Applied Statistics

Conference posters

- **Szabolcs Kelemen**, Levente Varga, and Zoltán Néda. (11-13 May, 2021) *Cross-correlations in the Brownian motion of colloidal nanoparticles* MECO46: 46th Conference of the Middle European Cooperation in Statistical Physics
- István Gere, **Szabolcs Kelemen**, Tamás S. Biró and Zoltán Néda. (12-16 June, 2022) *Wealth inequalities in different socio-economical situations. Exhaustive data and a general modeling framework.* MECO47: 47th Conference of the Middle European Cooperation in Statistical Physics
- **Szabolcs Kelemen**, Máté Józsa, Tibor Hartel, György Csóka and Zoltán Néda. (22-26 May, 2023) *Tree size distribution in the perspective of the Local Growth and Global Reset (LGGR) model.* MECO48: 48th Conference of the Middle European Cooperation in Statistical Physics
- **Szabolcs Kelemen**, Máté Józsa, József Benedek Cristian Litan and Zoltán Néda. (10-14 July, 2023) *Handling incomplete information: Gini coefficient from coarse-grained data.* SigmaPhi 2023

Selected Bibliography

[1] S.G. Brush et al., *Statistical Physics and the Atomic Theory of Matter: From Boyle and Newton to Landau and Onsager*, 1983.

[2] Per Bak, *How Nature Works*, 1996.

[3] Nino Boccara, *Modeling Complex Systems*, 2010.

[4] J. van Brakel, *Archive for History of Exact Sciences*, 31, 1985, pp. 369–385.

[5] Vitaly Vanchurin et al., *Proceedings of the National Academy of Sciences*, 119, 2022, e2120042119.

[6] Hong Qian and Lisa M. Bishop, *International Journal of Molecular Sciences*, 11, 2010, pp. 3472–3500.

[7] Wolfgang Weidlich and Martin Braun, *Journal of Evolutionary Economics*, 2, 1992, pp. 233–265.

[8] Zoltán Néda et al., *PLOS ONE*, 12, 2017, e0179656.

[9] Robert Brown, *A Brief Account of Microscopical Observations... on the Particles Contained in the Pollen of Plants; and on the General Existence of Active Molecules in Organic and Inorganic Bodies*, 1828.

[10] W. Ebeling et al., *Acta Physica Polonica B*, B39, 2008.

[11] Willy Feller, *Acta Biotheoretica*, 5, 1939, pp. 11–40.

[12] A. Nordsieck et al., *Physica*, 7, 1940, pp. 344–360.

[13] Markus F Weber and Erwin Frey, *Reports on Progress in Physics*, 80, 2017, p. 046601.

[14] N. G. van Kampen, *Canadian Journal of Physics*, 39, 1961, pp. 551–567.

[15] Tamás S. Biró and Zoltán Néda, *Physical Review E*, 95, 2017.

[16] T.S. Biró and Z. Néda, *Physica A: Statistical Mechanics and its Applications*, 499, 2018, pp. 335–361.

[17] Tamás S. Biró and Zoltán Néda, *Entropy*, 21, 2019, p. 993.

[18] Zoltán Néda et al., *Physica A: Statistical Mechanics and its Applications*, 549, 2020, p. 124491.

[19] István Gere et al., *Physica A: Statistical Mechanics and its Applications*, 581, 2021, p. 126194.

[20] Branko Milanovic, *Global Inequality*, 2016.

[21] Géza Hegyi et al., *Physica A: Statistical Mechanics and its Applications*, 380, 2007, pp. 271–277.

[22] Vilfredo Pareto, *Journal of political economy*, 5, 1897, pp. 485–502.

[23] A. Drăgulescu and V.M. Yakovenko, *The European Physical Journal B*, 20, 2001, pp. 585–589.

[24] Jean-Philippe Bouchaud and Marc Mézard, *Physica A: Statistical Mechanics and its Applications*, 282, 2000, pp. 536–545.

[25] Zoltán Néda et al., *Physica A: Statistical Mechanics and its Applications*, 468, 2017, pp. 147–157.

[26] Victor M. Yakovenko and J. Barkley Rosser, *Reviews of Modern Physics*, 81, 2009, pp. 1703–1725.

[27] Dan Cao and Wenlan Luo, *Review of Economic Dynamics*, 26, 2017, pp. 301–326.

[28] World Inequality Database, *Data - WID - World Inequality Database*, https://wid.world/data/, [Accessed 20-11-2023].

[29] James B. Davies et al., *The Economic Journal*, 121, 2010, pp. 223–254.

[30] F Clementi et al., *Journal of Statistical Mechanics: Theory and Experiment*, 2012, 2012, P12006.

[31] Rosie Dunford et al., *The Race*, 2021.

[32] Oren S. Klass et al., *Economics Letters*, 90, 2006, pp. 290–295.

[33] Ernszt Árpád Varga, *Erdély etnikai és felekezeti statisztikája. IV. Fehér, Beszterce-Naszód és Kolozs megye. Népszámlálási adatok 1850-1992 között*, 2001.

[34] Max O. Lorenz, *Publications of the American Statistical Association*, 9, 1905, pp. 209–219.

[35] Catalin Barboianu, *The Mathematics of Lottery: Odds, Combinations, Systems*, 2009.

[36] V. Ariyabuddhiphongs, *Journal of Gambling Studies*, 27, 2010, pp. 15–33.

[37] Ian Walker, *Economic Policy*, 13, 1998, pp. 357–402.

[38] Lisa Farrell et al., *Oxford Bulletin of Economics and Statistics*, 61, 1999, pp. 513–526.

[39] *Aug 25, 2015 - Changing Powerball to make it obnoxiously harder to win. See why they have to change it.* https://lottoreport.com/PB2015Rule.htm, [Accessed 08-10-2023].

[40] *2017, 2016, 2015, 2014 Powerball and Powerplay draw sales*, https://lottoreport.com/powerballsales4.htm, [Accessed 08-10-2023].

[41] Euro-Millions.com, *Changes to EuroMillions*, https://www.euro-millions.com/changes, [Accessed 08-10-2023].

[42] *LottoTexas.com - 2017 Mega Millions Rule Change. Obviously, the game is not intended to be won very often!*, https://lottoreport.com/MMRules.htm, [Accessed 08-10-2023].

[43] *Texas Lotto Report - How Lotto Texas monies are distributed.* https://lottoreport.com/ThePayouts.htm, The Powerball, Megamillions, and Texas lotto rules and datasets were scraped from the https://lottoreport.com website, which is an open archive of US lotteries. [Accessed 08-10-2023].

[44] Take That Ltd National-Lottery.com, *Canada 6/49 Draw Results Archive: 2012*, https://www.national-lottery.com/canada-6-49/results/2012-archive, [Accessed 08-10-2023].

[45] BeatLottery.co.uk, *Lotto Jackpots History : Lotto Statistics : Beat Lottery*, https://www.beatlottery.co.uk/lotto/statistics, [Accessed 08-10-2023].

[46] *2018, 2019, 2020, 2021 Powerball and Powerplay draw sales*, https://lottoreport.com/powerballsales5.htm, [Accessed 08-10-2023].

[47] *Mega Millions - 2012, 2011, 2010, 2009, 2008, 2007 Draw sales and jackpot amounts by date.* https://lottoreport.com/mmsales3.htm, [Accessed 08-10-2023].

[48] *Téléchargement des résultats Euro Millions*, https://www.loterieplus.com/euromillions/services/telechargement-resultat.php, [Accessed 08-10-2023].

[49] *Total sales and jackpot amounts for each of the Lotto Texas drawings.* https://lottoreport.com/lottosales1999_2001.htm, [Accessed 08-10-2023].

[50] Shin-Ichi Yamamoto, *Journal of Forest Research*, 5, 2000, pp. 223–229.

[51] Biman Chakraborty et al., *Journal of Biological Physics*, 48, 2022, pp. 295–319.

[52] Takashi Kohyama et al., *Journal of Ecology*, 91, 2003, pp. 797–806.

[53] Zsófia Szegleti et al., *Data in Brief*, 47, 2023, p. 108929.

[54] John P. Schmit et al., *National Capital Region Network Long-Term Forest Vegetation Monitoring Protocol: Version 2.1 (March, 2014)*.

[55] Szabolcs Kelemen et al., *Diameter at Breast Height (DBH) data of temperate zone trees from different woodland types. Figshare* DOI: https://doi.org/10.6084/M9.FIGSHARE.24039429.

[56] David A. Coomes and Robert B. Allen, *Journal of Ecology*, 95, 2006, pp. 27–40.

[57] Evan M. Gora and Adriane Esquivel-Muelbert, *Nature Plants*, 7, 2021, pp. 384–391.

[58] Renato Augusto Ferreira de Lima et al., *Forest Science*, 61, 2015, pp. 320–327.