Data Driven Analysis and Modelling of the Wealth Distribution

by

Samuel Forbes

Thesis

Submitted to the University of Warwick

for the degree of

Doctor of Philosophy in Mathematics of Systems

Mathematics

October 2022

THE UNIVERSITY OF WARWICK

Contents

Acknowledgments

I would like to thank my supervisor Stefan Grosskinsky for his invaluable help and support over the course of the PhD. He was always willing to answer my questions and provide very useful feedback. His probability knowledge was crucial for proceeding with the main research in this thesis. Stefan gave great help with improving the plots and finding interesting aspects in them that I would otherwise have missed. He also alerted me to many opportunities during the PhD and our discussions always gave me enthusiasm for new directions of research. I would like to thank my supervisor Alexander Karalis Isaac who provided great help in my attempt to learn more about economics. It was enjoyable to have many discussions with him on economic topics. I am very grateful to the time Paul Chleboun took to show me how the binary tree search worked and his help with the code. Paul was also kind with his time to help with other aspects of fitting power laws. I also thank Colm Connaughton as he gave me good ideas for returns and thinking about the bias question in survey data. I would like to thank all of the staff involved with MathSys and the EPSRC for the funding. It was a privilege to be involved with a centre that allowed for many different pathways into research. My fellow students at MathSys and those in the wider Warwick maths and stats community created a fun environment and I enjoyed our time together and thank them for all their help and support. Finally I would like to acknowledge the great support of my friends and family outside MathSys.

Declarations

This thesis is submitted to the University of Warwick in support of my application for the degree of Doctor of Philosophy. It has been composed by myself and has not been submitted in any previous application for any degree.

Chapter 1 and Chapter 2 contain primarily reviews of the existing literature along with some preliminary new work. Chapter 3 contains joint work with Stefan Grosskinsky with some original results involving the master equation. The code used for some simulations in Chapter 3 is explained in Appendix B and is joint work with Paul Chleboun. Chapter 4 is new work that has been published in PLOS ONE with Stefan Grosskinsky [60].

Abstract

We study the wealth distribution empirically with analysis of the UK wealth and asset survey and rich list data and focus on prominent factors of the distribution through mathematical modelling. Probability distributions for both debt and positive wealth are fitted to the wealth data concentrating on the time period 2008-2016. We fit power laws, a key property of the wealth distribution, to the upper tail and analyse the difference in power law exponents between the survey rich and the rich lists. We present an overview of potential agent based wealth models under the themes of hierarchy, exchange, feedback and multiplicative processes. Two of these models, one in each of the latter two categories, are studied in detail in the final main results chapters. Both models are characterised by a critical power γ parameter, exhibit power law tails and eventually extreme inequality. The first is the balls in bins process with feedback originally studied in the combinatorics literature and only recently applied to wealth. We analyse theoretical aspects of this model as well as some general simulations. The second model, which we call a non-linear Kesten process, has not been studied before to our knowledge and is a generalisation of the Kesten process. This model was conceived by finding a rough power law relationship between agent's wealth and their wealth returns. Agents evolve independently through time based on these returns. Due to the independence of agents we can run the model for a large number of agents and we do so for general and realistic 2008 UK initial conditions. We conclude that a non-linear rich gets richer effect may be important when modelling the wealth distribution in times of growing inequality.

Introduction

The study of wealth and its distribution has been reignited in recent years due to evidence of rising inequality since the 1980s [9, 7]. Especially since Piketty's 2014 work 'Capital in the 21st Century' there has been an increasing focus on billionaires and controversy over potential policies such as a wealth tax [121, 119]. The Covid-19 pandemic and recent record rises in inflation in 2022 have added flames to the fire over discussions around wealth and inequality [22, 15].

Although we shall consider inequality, the primary focus of this thesis is a related feature of the wealth distribution when viewed as a probability distribution. This is the presence of a power law in the upper tail of the distribution. The power law indicates that the richest in society obtain vast quantities of wealth not possible in lighter tailed distributions. It is noteworthy that a mathematical function as simple as the power law appears in the upper tail of a wide variety of quantities from financial returns to city size [63, 111].

To model the wealth distribution we want to capture the most important factor(s) determining the distribution. In particular we focus on models that generate power laws and increase inequality by a *rich gets richer* effect. We focus on two such models: a balls in bins process with feedback also referred to as a non-linear generalised Pólya urn model and a non-linear Kesten process.

The main original work of the thesis appears in Chapter 4 and is published in [60]. Chapter 3 holds promise as an intuitive model to study wealth and other quantities such as city size in the future. Chapters 1 and 2 contains preliminary fitting to wealth data and review of the existing literature with some minor extensions.

We now outline in more detail the contents of these chapters. In Chapter 1 we empirically analyse wealth by discussing the definition of wealth, the data sources we use to analyse UK wealth primarily between the years 2008-2016, the main theory and definitions of heavy tailed distributions and power laws, fitting the wealth distribution and a brief overview of inequality which we measure for simulations in the final chapter. A mixture distribution is fitted to negative and positive wealth. We find that a relatively unknown κ-generalised distribution fits the positive wealth distribution well. We also discuss the disparity of the power law exponents seen between the rich and very rich and show how survey bias may be a cause of this phenomenon.

In Chapter 2 we summarise a selection of potential agent based wealth models from the literature as well as presenting some variations and extensions. In line with the data we view each agent as a household. We focus on modelling positive wealth for a fixed number of agents. Many of these models have not been applied directly to wealth data but as they are characterised by their ability to produce heavy-tailed distributions and power laws they are potential candidates. The models we analyse come under the broad categories of hierarchy, exchange, feedback and multiplicative processes. The hierarchy model is a static model showing how a hierarchical arrangement of agents in terms of their wealth leads to power law distributions. The exchange models are based on work from econophysics where agents are viewed as particles and money, a subset of wealth, as energy. Each monetary transaction is seen as a transfer of energy. The simplest exchange model of repeated fixed transactions with no debt produces the Boltzmann distribution [146]. Adjusting this model by making the exchange amount dependent on the agent or having agents transact at a power law rate dependent on their wealth can produce power laws. Finally we overview a balls in bins process with feedback and study a number of multiplicative process models in both discrete and continuous time. In the balls in bins process wealth entering the system goes to an agent with probability proportional to a feedback function of their current wealth. In the multiplicative process models the wealth of agents is determined primarily by random multiplicative factors. We study the multiplicative process models in both discrete time as difference equations and continuous time as stochastic differential equations (SDE). The particular method of finding a stationary density of an SDE using the Fokker-Planck equation is used to find SDEs with power law tailed solutions [23].

In Chapter 3 we analyse further the balls in bins process with particular power law feedback functions studied originally in the combinatorics literature [114, 118]. Many results are known for $N = 2$ agents however we study the general $N \in \mathbb{N}$ case. In this model, which can be viewed as a Markov process, wealth is repeatedly added to agents with probability proportional to the agents current wealth raised to a power γ. Thus for $\gamma > 0$ those with higher wealth are more likely to gain more wealth. For $\gamma > 1$ a power law emerges in the distribution as well as a monopolising agent who gains almost the entire proportion of total wealth. An approximation to the expected value of wealth is found at time t. We solve for the probability mass function at time t of the balls in bins process via the master equation for equal initial conditions of one unit of wealth. Additionally we add a fitness attached to each agent and show this scales t in the resulting probability mass function and expected wealth meaning that higher fitness predicts higher wealth. The fitness can be thought of as qualities that give agents more success in gaining wealth such as productivity or intelligence. We run some general simulations but note that even though we use a binary tree search, see Appendix B, to speed up the process we cannot simulate a realistic number of households in an appropriate amount of time. The draw back of this model is relating wealth received by agents to a real world context. Likewise it is unclear how to fit a fitness distribution for agents. However the rich gets richer preferential attachment effect in this model is a natural concept that is reasonable to think happens in the economy [10].

Finally, Chapter 4 more naturally models the reality as we base it on an empirical relationship we see in wealth returns. Our main proposition is that an agent with higher wealth is more likely to have higher wealth returns that are proportional to current wealth raised to the power $\gamma - 1$. This model is a discrete Markov multiplicative process identical to the Kesten process [88] other than the power γ on a wealth component. This γ is found by fitting the returns relationship with the data. The additive part of the model is fixed and viewed as wealth independent savings correlated to initial wealth. The randomness in the model is characterised by a pre-factor parameter fitted to the data with a heavy-tailed non-central t distribution. Like with the balls in bins process when $\gamma > 1$ we produce power law tails and extreme inequality in the long time limit. Inequality increases rapidly until a critical region beyond which super-exponential growth sets in. We run the model for both general and realistic initial conditions and as the agents are independent it is feasible to run the model for approximately the number of households in the UK over a reasonable time period. Although the UK real world data shows fairly flat overall inequality during 2008-2016 this may not reflect the reality of potentially growing inequality over the same period, see for example [7]. We show that over periods of increasing inequality the model can produce reasonable results.

Chapter 1

Empirical Analysis of Wealth

1.1 Introduction

Wealth, specifically net wealth, can be defined as the total of assets net liabilities (or debts) [13, 9, 127]. Unlike GDP and income which are flows measured over a period of time, wealth is a stock that can conceivably be measured at any point in time with a certain value. Wealth can be both positive and negative and is measured for different entities increasing in scale from, for example, individuals to households to companies to nations. There are different types of wealth. In the wealth and assets survey the types of household wealth are divided into four components: physical, property, financial and pension [40]. However these wealth categories are not standard and can be classified in different ways, see for example [82]. We can measure wealth in nominal or real values. Nominal is the value at the particular time measured, whereas real is the value adjusted to a particular time. The difference between the nominal and real value is due to inflation. We shall not discount inflation and thus will be analysing primarily nominal wealth. Wealth at all scales has been increasing rapidly since at least the industrial revolution due to many interlinked factors such as resources, population, technology and inflation [103]. For example per capita real wealth in Britain between 1760-2000 increased roughly ten fold [103, 80].[1]

We shall focus on the distribution of wealth at the level of households which we refer to as agents throughout the thesis. By distribution we mean the abstract mathematical probability distribution which is an approximation for how wealth is distributed in reality. We view wealth as a continuous variable taking a continuous distribution. One of the interesting aspects of the probability distribution of wealth,

[1] In these references we find real wealth increased roughly one hundred fold whilst population increased only roughly ten fold.

as with income, is its heavy tailed nature and the presence of a power law for the very richest in society. The heavy tails reflect the skew in wealth at the top and the possibility of comparatively extreme wealth values [146, 33].

1.2 Discussion of Wealth Data Sources

Alvaredo et al [8] mentions four main sources of wealth data in the UK: tax on estates, tax on investment income, surveys and rich lists. Tax data is found very early in civilisation and is recorded in Mesopotamian texts as early as 2500BC [108]. Taxation was also a key motivation behind the Domesday book, one of the first surveys of wealth completed in the year 1086 by order of King William the conqueror [102]. One of the first to estimate wealth using tax records was Vilfredo Pareto. Using tax data from multiple countries and as early as the 16th century, Pareto discovered the presence of power law tails by plotting on logged scale engineering paper [117, 101].

In the modern age there have been two national sample savings surveys for Great Britain in 1953 and 1954 which allowed for estimates of wealth [97]. The potential of wealth surveys in the 1970s were discussed by The Royal Commission on the Distribution of Income and Wealth, also referred to as the Diamond Commission [129]. These surveys however were never carried through because the response rate was thought to be too low [8]. The UK government's view on household wealth surveys changed when the British Household Panel Survey (BHPS) started collecting household financial wealth data from 1995 [8]. Then in 2000 the Office for National Statistics (ONS) planned for a longitudinal wealth and assets survey (WAS) for Great Britain starting in 2006 [8, 40]. The WAS conducted by the ONS aimed to address gaps in the knowledge of household wealth [40].

Rich lists attempt to quantify wealth of the richest in society. Two well known rich lists are the Sunday Times rich list which has run annually from 1989 and the annual Forbes rich list starting in 1982 [53, 93]. The Sunday Times rich list is conducted by the Sunday Times newspaper and is specifically for the richest households in the UK. The Forbes rich list is conducted by the Forbes American business magazine and looks at the world's dollar billionaires. Rich lists are subject to bias as they are unlikely to capture everyone who should be on the list [53]. Also they are subject to human judgement as the lists are compiled by employees who have access only to incomplete data [53, 93].

There are issues with both taxation and survey data. For example, problems accurately measuring taxation data include avoidance and evasion and difficulties tracking and valuing assets [8]. Surveys present their own issues mainly in the form of potential differential non-response survey bias [145]. Differential non-response survey bias is when particular groups of the population do not respond to the survey. If these non-responder groups are more wealthy than is seen in the data then the wealth distribution will be biased and may seem more equal than it actually is. Plausible reasons for why non-responder groups could be more wealthy include hiding wealth and miscalculation of assets. The WAS does attempt to compensate for differential non-response bias via weighting, see Section 10.6 in [115].

We focus on three sources of wealth for our data analysis: the WAS biennial data from 2008-2016[2] for Great Britain and the Forbes and Times rich lists. Thus although throughout the thesis we discuss UK wealth we are mainly analysing Great British wealth plus potential Northern Irish rich households from the rich list. An advantage of these data sources is that they are measured at the household level which makes estimation of the probability distribution possible. In contrast, many early empirical studies of wealth were only in certain brackets of the population making it hard to have a good sense of the whole probability distribution [97]. Indeed, the Forbes and Times rich lists aim to capture the entirety of the richest households.

1.3 Heavy Tails and Power Laws

We now summarise some key definitions and theory on heavy tails to set the scene before we fit to the wealth distribution. Letting W be a random variable (RV) with probability measure $\mathbb{P}$, then the tail is the probability that W takes a value greater than some value w: $\mathbb{P}(W > w)$. In other words it is one minus the cumulative distribution function (CDF) of W. We can approximate the tail of W with the empirical tail:

Definition 1 (Empirical Tail)**.** *Let $\{w_1, w_2, \ldots, w_N\}$ be an i.i.d. sample from a random variable W. We define the empirical tail* [3] *of W as*

$$\mathbb{P}_N(W > w) := \frac{1}{N}\sum_{i=1}^{N} \mathbf{1}_{w_i > w} \tag{1.1}$$

where $\mathbf{1}$ *is the indicator function.*

Note that the empirical tail is one minus the empirical cumulative distribution

[2] We note that this survey is on-going but we only accessed fine-grained data during this period.
[3] Which more specifically is the empirical complementary cumulative distribution function.

function (ECDF) [143]:

$$\mathbb{P}_N(W \leq w) = 1 - \mathbb{P}_N(W > w) = \frac{1}{N}\sum_{i=1}^{N} \mathbf{1}_{w_i \leq w}\,.$$

Many results for the ECDF can by simple extension be applied to the empirical tail. For example we note that the ECDF tends towards the cumulative distribution function as the sample number increases to infinity, see specifically the Glivenko-Cantelli Theorem in Chapter 19 of [143] for details. Thus by simple extension the empirical tail tends towards the tail distribution as the sample number increases to infinity:

$$\mathbb{P}_N(W > w) \to \mathbb{P}(W > w) \text{ as } N \to \infty\,.$$

An important tail to consider is that of an exponential distribution with tail

$$\mathbb{P}(W > w) = e^{-\lambda w}, \;\; \lambda > 0, w > 0\,.$$

On a graph of w versus $\mathbb{P}(W > w)$ with a logged scale y-axis (a semi-log plot) the tail of the exponential distribution is then a straight line. A heavy-tailed distribution, such as a lognormal distribution, turns upwards on a semi-log, and a light tailed distribution such as the Weibull distribution, which we shall see in Section 1.4.5, turns downwards, see Figure 2.8 of [41]. A distribution that is heavy tailed has a tail distribution that approaches 0 less rapidly than an exponential tail and so has a greater probability of higher values than an exponential. Conversely a light tailed distribution has a tail distribution that approaches 0 more rapidly than an exponential tail and so predicts higher values less often than an exponential and consequently also a heavy tail [41]. Note we can compare two tails by saying one is lighter/heavier than another if it approaches 0 more/less rapidly.

There are several ways to mathematically define a heavy tailed distribution, see Chapter 2 of [61]. We choose the following definition of a heavy tailed distribution relating the tail convergence to the exponential tail convergence described in the previous paragraph:

Definition 2 (Heavy-tailed distribution)**.** *We call the probability distribution of W heavy-tailed if the following is satisfied:*

$$\lim_{w\to\infty} \frac{\mathbb{P}(W > w)}{e^{-\lambda w}} = \infty \quad \forall\ \lambda > 0\,. \tag{1.2}$$

This definition is the one used in [109] which goes into much more detail on what we overview in this section. We note that the faster the limit (1.2) tends to infinity the heavier the tail.

We now briefly present some classes of heavy tailed distribution. Firstly, we consider subexponential distributions which is proved to be heavy tailed in Lemma 3.2 of [61]. The name subexponential was originally chosen as the tail of a subexponential distribution decays more slowly than an exponential tail which is precisely the definition of a heavy tail given above in Definition 2. We present the following definition seen in [69]:

Definition 3 (Subexponential distribution)**.** *Suppose $W_1, W_2, \ldots W_n$ are i.i.d. positive RVs with some probability distribution. Then this distribution is subexponential if*

$$\lim_{w\to\infty} \frac{\mathbb{P}(\max(W_1, \ldots, W_n) > w)}{\mathbb{P}(W_1 + \cdots + W_n > w)} = 1\,. \tag{1.3}$$

We can see from Definition 3 that subexponential distributions are related to a monopoly effect that will be seen in models presented later in the thesis. By monopoly we mean that the wealthiest agent takes almost all the total wealth as the total wealth in the system grows. This can be seen more concretely by the following formulation of (1.3):

$$\lim_{w\to\infty} \mathbb{P}(\max(W_1, \ldots, W_n) > w) | W_1 + \cdots + W_n > w) = 1 \tag{1.4}$$

which is true as the $\max(W_1, \ldots, W_n) > w$ necessarily implies $W_1 + \cdots + W_n > w$.

A second class of heavy tailed distributions is called regularly varying. Suppose we have a function f such that

$$\lim_{w\to\infty} \frac{f(\lambda w)}{f(w)} = \frac{1}{\lambda^\beta} \tag{1.5}$$

for any real $\lambda > 0$ and $\beta \in \mathbb{R}$. If $\beta = 0$ then f is called a slowly varying function and if $\beta > 0$ then f is called a regularly varying function, see [133, 20] for extensive theory relating to these functions. It can be shown that any regularly varying function $f(x)$ can be written as $f(x) = l(x)x^\beta$ where $l(x)$ is a slowly varying function [20]. When the tail function is regularly varying we have the following definition [81]:

Definition 4 (Regularly Varying Distribution)**.** *The probability distribution of W is called regularly varying if its tail distribution is regularly varying i.e. for some*

$\beta \in \mathbb{R}_{>0}$

$$\lim_{w\to\infty} \frac{\mathbb{P}(W > \lambda w)}{\mathbb{P}(W > w)} = \frac{1}{\lambda^\beta} \quad \forall\ \lambda > 0\,.$$

Regular variation can also be written as

$$\mathbb{P}(W > \lambda w) \simeq \frac{1}{\lambda^\beta}\mathbb{P}(W > w)$$

using the notation [81][4]:

$$f(x) \simeq g(x) \Leftrightarrow \lim_{x\to\infty} \frac{f(x)}{g(x)} = 1\,. \tag{1.6}$$

We now summarise some properties of regular variation:

RV.1 Subclass of subexponential: Regularly varying distributions are also subexponential distributions but not vice versa, see Chapter 2 of [61]. For example a lognormal distribution is subexponential but not regularly varying [98, 109].

RV.2 Moments: If W is a regularly varying distribution with parameter β then the moments for $k > 0$, $\mathbb{E}[W^k]$, are finite if $k < \beta$ and infinite otherwise [20].

RV.3 Closure: The weighted sum of i.i.d. regularly varying RVs is again a regularly varying RV, under certain conditions the product of regularly random variables is again regularly varying and a regularly varying RV raised to some power is again regularly varying, see respectively Lemma 3.3, Section 4 and Section 5 of [81]

Taking $\mathcal{H}$, $\mathcal{S}$ and $\mathcal{R}$ as the class of heavy tailed, subexponential and regularly varying distributions respectively then as stated above we have the following containment

$$\mathcal{R} \subset \mathcal{S} \subset \mathcal{H}\,.$$

A regularly varying distribution of particular interest is the power law tail [64, 100]. This is defined as

$$\mathbb{P}(W > w) = \frac{\alpha}{w^\beta}\,, \quad \text{for } w > w_m > 0 \tag{1.7}$$

[4] Here $\simeq$ is often written as $\sim$.

for some real parameters

$$w_m,\ \alpha \text{ and } \beta \tag{1.8}$$

greater than zero. We refer to β as the **power law exponent**. We note that equality in (1.7) is often taken asymptotically (1.6). Indeed whether (1.7) has equality or is asymptotic is referred to as strong and weak Pareto law respectively in early work by Mandelbrot [99]. In this thesis we do not emphasise distinguishing between the two. The Pareto distribution is defined by (1.7) with

$$\alpha = w_m^\beta$$

and is the probability distribution for which the entire tail is a power law. It is named after Vilfredo Pareto as mentioned in Section 1.2.

Power laws present themselves as straight lines on the tail plot of w versus $\mathbb{P}(W > w)$ where both the axes are logged (also often called a 'log-log plot'). The parameter α shifts the tail up or down and the power law exponent β determines the gradient in log-log space. Power laws with a smaller β or shallower negative gradient are heavier. Tails which are less heavy than a power law, curve downwards on a log-log scale. This shall be seen in plots in Section 1.4.

There is a scale invariance property of the power law:

$$\mathbb{P}(cW > w) = c^\beta \mathbb{P}(W > w)\,, \quad \text{for some constant } c\,. \tag{1.9}$$

This implies that if wealth was inflated equally among the richest members of the population that the power law exponent will not change and that inequality amongst these members would remain the same.

By the fundamental theorem of calculus for continuous variables the probability density of W also has a power law with

$$f_W(w) = \frac{\alpha\beta}{w^{\beta+1}} \text{ for } w > w_m > 0\,. \tag{1.10}$$

Due to the high level of variability, unpredictability and complexity in the social sciences trying to find economic laws as compared to scientific laws may be futile. Despite this we could say that economics has particular tendencies, one of which could be [91]

The wealth distribution has a power law tail.

Pareto posed that this tendency holds across all time and in all countries [29]. Originally Pareto thought that the entire distribution was a power law (i.e. a Pareto distribution) however it was found that the power law only holds for the upper tail (for wealth greater than some w_m) [29]. Pareto also hypothesised that the power law exponent was roughly constant and held at around $\beta = 1.5$ [29] however again there is evidence that there is more fluctuation of the β across time and place. We shall see in the next section evidence of the power law only holding in the upper tail and relatively wide fluctuations in the β across time. Since Pareto there have been many studies suggesting this power law tendency including evidence of power laws in pre-industrial societies such as for Hungarian aristocrats in 1550 [130].

Wealth is not the only economic variable that has a power law distribution. Related quantities such as income, city size, firm size, land ownership and financial returns all exhibit power law distributions [64]. Indeed, power law distributions appear in variables outside of economics and not only for continuous variables but for discrete variables such as the degree distribution of internet topology, the number of species per genus, the number of times words appear in a text (commonly known as Zipf's law), citation counts and YouTube views [31, 55, 104].

1.4 Fitting the Wealth Distribution

We shall use WAS survey data and rich lists to fit the UK wealth distribution. Details of how we extract the data to form the empirical tail as seen in Figure 1.1 are in Appendix C.2. Our extraction of the survey data relied on fine grained cumulative household versus cumulative wealth data. This enabled us to find an approximation of N empirical tail points:

$$(w_i, \mathbb{P}_N(W > w_i))_{i=1}^{N} \tag{1.11}$$

Thus we did not produce the empirical tail directly from a sample[5] as in Definition 1.

1.4.1 Power Laws in the Positive Wealth Tail

We note Figure 1.1 covers only positive wealth for the UK. The vast majority of the survey empirical tail over the 2008-2016 time period is below £100 million (£10^8). The Forbes (dollar) billionaires converted to pounds over the same time period start from roughly £500 million whereas the Times rich lists start from roughly £100 million. The colour coded empirical tails are plotted for biennial years from

[5] It is noted that direct WAS sample data is available from the UK data service [5]. However the data downloaded there is presented in a complex form and it would be future work to analyse it properly.

2008 to 2016 along with the Forbes rich lists forming a gap from roughly £50 million (£5 $\cdot$ 10^7) to £500 million (£5 $\cdot$ 10^8). Additionally we plot Times rich list data for 2019, 2020 and 2021.

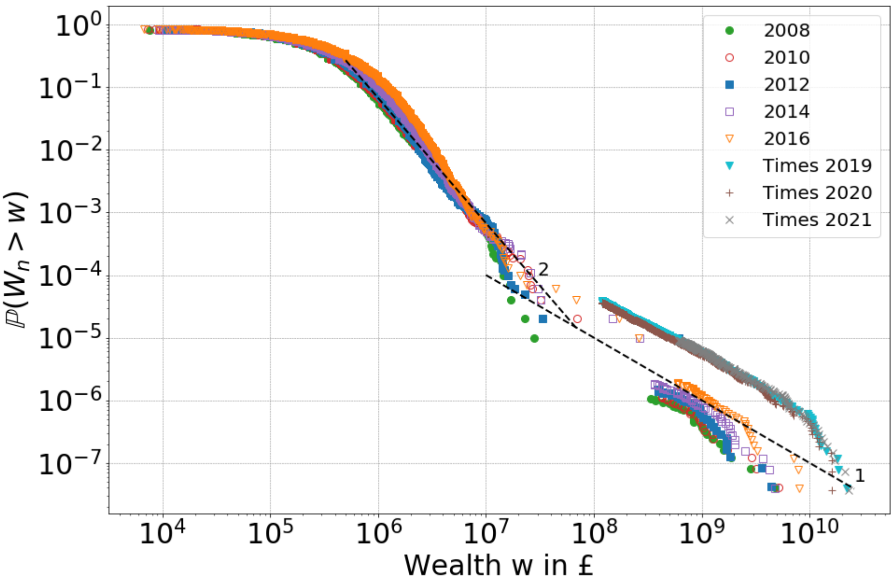

Figure 1.1: Empirical tail distribution of positive UK household wealth for five consecutive biennial time periods 2008, 2010, 2012, 2014, 2016 from WAS [116], together with Forbes rich list data on dollar billionaires [1], and UK Times rich list data from 2019, 2020 and 2021 (see appendix C.1). Dashed lines indicate power law tails with power law exponents 2 and 1 for comparison.

Figure 1.1 is on a double log or, as mentioned previously a log-log plot, where both the x and y axes are in a log base 10 scale. The standard base 10 is chosen for ease of readability of the plot values. As mentioned in Section 1.3 power laws are seen as straight lines on log-log plots. We can see immediately that the UK wealth distribution over the time period does not follow a Pareto distribution as the entire tail is not a straight line on the logged axes. In Figure 1.1 we plot power laws with exponents $\beta = 2$ and $\beta = 1$ for comparison between the WAS rich and Forbes dollar billionaires. We shall see later in the chapter that these exponents are not exact fits but illustrate the clear difference in tail gradient from the survey rich to billionaires.

The power law exponent for both Forbes and Times rich lists appear in a similar range. This rough agreement of exponent from two different data sources indicates some accuracy in the rich list data. The difference in exponents from WAS to rich lists appears to be significant and could suggest the distribution has two power law tails with different exponents. However as we have mentioned in Section 1.2, the seemingly two tailed power law structure could be due to differential survey bias, see Section 1.5, and so in reality there may be only one power law tail. There is evidence that this might be the case as when the WAS data is adjusted it is possible to match the power laws of the survey and rich lists, see [7].

1.4.2 Wealth Mixture Distribution

Now assume the wealth distribution with RV W is on the domain $(-\infty, \infty)$ [6]. Suppose W is a mixture distribution of two continuous distributions, one for negative wealth $(-\infty, 0)$ with RV W_- and the other for non-negative wealth $[0, \infty)$ with RV W_+. We note this mixture distribution is a simplification of what is found in [34] which was originally presented in [43].[7] The density of W is written in terms of the densities of W_- and W_+ as the following:

$$f_W(w) = \theta f_{W_-}(w) + (1-\theta) f_{W_+}(w)$$

with parameter $\theta \in (0,1)$. Then the tail of W is written in terms of the tails of W_- and W_+ as follows

$$\mathbb{P}(W > w) = \theta \mathbb{P}(W_- > w) + (1-\theta)\mathbb{P}(W_+ > w)\,. \tag{1.12}$$

We note $\mathbb{P}(W_- > w) = 0$ for $w \geq 0$ and $\mathbb{P}(W_+ > 0) = 1$ so that $1 - \theta = \mathbb{P}(W \geq 0)$[8] and $\theta = \mathbb{P}(W < 0)$. Therefore

$$\mathbb{P}(W > w) = \mathbb{P}(W \geq 0)\mathbb{P}(W_+ > w)\,, \quad \text{for } w > 0\,. \tag{1.13}$$

Likewise the CDF of the mixture distribution is

$$\mathbb{P}(W < w) = \theta \mathbb{P}(W_- < w) + (1-\theta)\mathbb{P}(W_+ < w)\,. \tag{1.14}$$

where in particular $\mathbb{P}(W_+ < w) = 0$ for $w \leq 0$.

Now let us consider the absolute value of negative wealth or debt $|W|$, $W < 0$.

[6] Although we note that in reality wealth is bounded.

[7] In particular we avoid a separate unit mass distribution for zero wealth.

[8] By assumed continuity of the distribution we can disregard the inequalities thus $\mathbb{P}(W < w) = \mathbb{P}(W \leq w)$ and similarly for the tail.

For debt $|w|$ with $w < 0$ we arrive at the following formulation of the tail

$$\begin{aligned}\mathbb{P}(|W| > |w|) &= \mathbb{P}(W < w) \\ &= \theta\mathbb{P}(W_- \leq w)\,, \quad \text{from (1.14)} \\ &= \mathbb{P}(W < 0)\mathbb{P}(|W_-| > |w|)) \quad \text{for } w < 0\,. \end{aligned} \tag{1.15}$$

1.4.3 Methods for Fitting Distributions to Wealth

We shall now describe how we fit parameters of chosen distributions to the wealth data. It will be seen that the exponential and lognormal distributions both fit well positive wealth before the power law tail. Therefore we first describe how we fit the exponential and lognormal distributions. First from the data we estimate $\mathbb{P}(W \geq 0)$. Then we estimate the mean and median numerically of W_+ using the alternative expectation formula [77] and the 50th percentile of W_+ as follows

$$\begin{aligned}\mu_+ := \mathbb{E}[W_+] &= \int_0^\infty \mathbb{P}(W_+ > w)\mathrm{d}w \\ m_+ := \mathrm{Med}[W_+] &= w \text{ s.t. } \mathbb{P}(W_+ > w) = \frac{1}{2} \Rightarrow \\ &= w \text{ s.t. } \mathbb{P}(W > w) = \frac{\mathbb{P}(W \geq 0)}{2} \text{ from (1.13)}\,.\end{aligned}$$

Then the exponential and lognormal fits to W for $w > 0$ give

$$\begin{aligned}\mathbb{P}(W > w) &= \mathbb{P}(W \geq 0)\exp\left(-\frac{1}{\mu_+}w\right) \text{ with } W_+ \sim \mathrm{Exp}\left(\frac{1}{\mu_+}\right) \\ \mathbb{P}(W > w) &= \mathbb{P}(W \geq 0)\mathbb{P}_{\mathrm{Lognorm}(k,s)}(W > w) \text{ with } W_+ \sim \mathrm{Lognorm}(k,s), \\ & k = \log(m_+) \text{ and } s \text{ chosen numerically with NLLS described below.}\end{aligned}$$

We now propose a method of non-linear least squares (NLLS), see Section 2 of [132], to fit wealth numerically the parameters Θ of a chosen probability distribution $D(\Theta)$. For N empirical tail values (1.11) we define the sum of squares of the difference of the empirical tail and distribution tail across all empirical tail points w_i :

$$S_{D(\Theta)} = \sum_{i=1}^{N}(\mathbb{P}_N(W > w_i) - \mathbb{P}_{D(\Theta)}(W > w_i))^2\,.$$

We then minimise numerically over $S_{D(\Theta)}$ to find approximate parameters for the distribution:

$$\hat{\Theta} = \arg\min_{\Theta} S_{D(\Theta)}\,. \tag{1.16}$$

We note we do not analyse the residuals, $\epsilon_i = \mathbb{P}_N(W > w_i) - \mathbb{P}_{D(\hat{\Theta})}(W > w_i)$, and assume they are roughly i.i.d. normal with zero mean. The cost function for the approximated parameters $\hat{\Theta}$ is then

$$C_D(\hat{\Theta}) = S_{D(\hat{\Theta})}\,. \tag{1.17}$$

Finally we note that maximum likelihood estimation (MLE) is commonly used to fit distributions. It is particularly convenient to use MLE when it gives analytic estimators as for the Pareto distribution which we shall discuss in Section 1.4.6. However for those distributions with no analytic MLE estimators numerical methods have to be used and are not a topic of this thesis.

1.4.4 Empirical Results: Positive Wealth and the κ-generalised Distribution

We now focus on fitting distributions to the empirical tail for positive wealth[9] from the WAS (excluding rich lists) for the particular years 2008 and 2016. We can see the exponential and lognormal distributions fit positive wealth for roughly the first 90% of the wealth distribution reasonably well in both 2008 and 2016, see Figures 1.2 and 1.3. In 2008 the lognormal distribution fits to a reasonable extent up to the first 99% of the positive wealth distribution, see Figure 1.2. However in 2016 the lognormal fits less well the top 10%, see Figure 1.3. In Appendix A.2 there are exponential and lognormal parameter fits, see Tables A.1 and A.2, and tail fits for all biannual years 2008-2016, see Figure A.1.

To fit the tail for positive wealth (1.13) we are interested in the tail value $\mathbb{P}(W \geq 0)$ which is the probability that a randomly selected member of the population has positive wealth. Equally the CDF at 0, $\mathbb{P}(W < 0) = 1 - \mathbb{P}(W \geq 0)$, is the probability that a randomly selected household has negative wealth. From our empirical tail data we estimate the EDF at 0: $\mathbb{P}_N(W < 0)$, as roughly 0.15 for each of the biennial years 2008 to 2016. This means roughly 15% of households have negative wealth which is likely an overestimate. It is stated elsewhere in the WAS data that roughly 15% of households have wealth less than £20000 over the time period [116].

As seen in Figures 1.2 and 1.3 we see visual evidence of power laws in the upper tail for roughly the richest top 10% illustrated by the rough straight lines. Thus we see a two part structure in the positive wealth distribution: the

[9] Strictly speaking two of the distributions we fit, the exponential and κ-generalised, will be for non-negative wealth $[0, \infty)$ whilst the lognormal fit is for strictly positive wealth $(0, \infty)$.

first 90% or so of the distribution (discounting negative wealth) follows roughly an exponential distribution and the remaining 10% or so follows a power law. This two part structure has been observed previously, see for example Figure 5 in [146]. A distribution with these characteristics is the κ-generalised (κ-gen) distribution [34]. For this distribution we first introduce the generalised exponential (gen-exp) function $\exp_\kappa : \mathbb{R} \to \mathbb{R}$:

$$\exp_\kappa(x) = (\sqrt{1+\kappa^2 x^2} + \kappa x)^{1/\kappa}. \tag{1.18}$$

for $x \in \mathbb{R}$ and deformation parameter $\kappa \in \mathbb{R}$. The gen-exp function (1.18) was first discovered in the context of non-linear kinetics in particle systems and has many mathematical properties [83]. For example the gen-exp function has many properties of the regular exponential e.g. $\exp_\kappa(x)\exp_\kappa(-x) = 1$, tends to a regular exponential function as $\kappa \to 0$ or $x \to 0$ and is also a regularly varying function (1.5).

Using (1.18) the κ-gen distribution is defined by the following density[10]

$$f_{\kappa\text{-gen}}(w) = \frac{\alpha}{\beta}\left(\frac{w}{\beta}\right)^{\alpha-1} \frac{\exp_\kappa\left(-\left(\frac{w}{\beta}\right)^\alpha\right)}{\sqrt{1+\kappa^2\left(\frac{w}{\beta}\right)^{2\alpha}}} \tag{1.19}$$

with tail function

$$\mathbb{P}_{\kappa\text{-gen}}(W > w) = \exp_\kappa\left(-\left(\frac{w}{\beta}\right)^\alpha\right) \tag{1.20}$$

for $w \geq 0$, $\alpha, \beta > 0$ and $\kappa \in (0,1)$. To fit the positive wealth distribution with κ-gen we have from (1.13)

$$\mathbb{P}(W > w) = \mathbb{P}(W \geq 0)\mathbb{P}_{\kappa\text{-gen}}(W > w), \text{ for } w > 0\,.$$

It can be shown that the κ-gen distribution is a regularly varying distribution (1.6), in particular it has a (asymptotic) power law tail (1.7), with

$$\mathbb{P}_{\kappa\text{-gen}}(W > w) \simeq \left(\frac{\beta^\alpha}{2\kappa}\right)^{1/\kappa} \frac{1}{w^{\alpha/\kappa}}\,.$$

Thus, in particular, we can approximate the κ-gen power law exponent as

$$\beta_{\kappa\text{-gen}} = \frac{\alpha}{\kappa}\,. \tag{1.21}$$

Fits with the κ-gen distribution are shown in Figures 1.2 and 1.3. The

[10] We shall make it clear when we use β differently from the power law exponent.

parameter fits for α, β and κ were found using NLLS (1.16).[11] Several other distributions also have a two part structure with a power law tail including Mittag-Leffler [66], Stable [42], Dagum, Generalised Beta, Inverse Generalised Gamma [44] and double Pareto Lognormal [124].[12]

In Appendix A.2 there are parameter fits for all WAS biennial years 2008-2016 positive wealth data with the κ-gen distribution, see Table A.3 and tail fit plots for the remaining years, see Figure A.1.

If we include rich list data we have a two power law structure as seen in Figure 1.1. We find the κ-generalised distribution gives a similar fit with and without the Forbes rich lists and is thus unable to detect the small number of UK Forbes dollar billionaires. We are unaware of a non-mixture distribution for the entirety of positive wealth that has two different power laws in the tail.

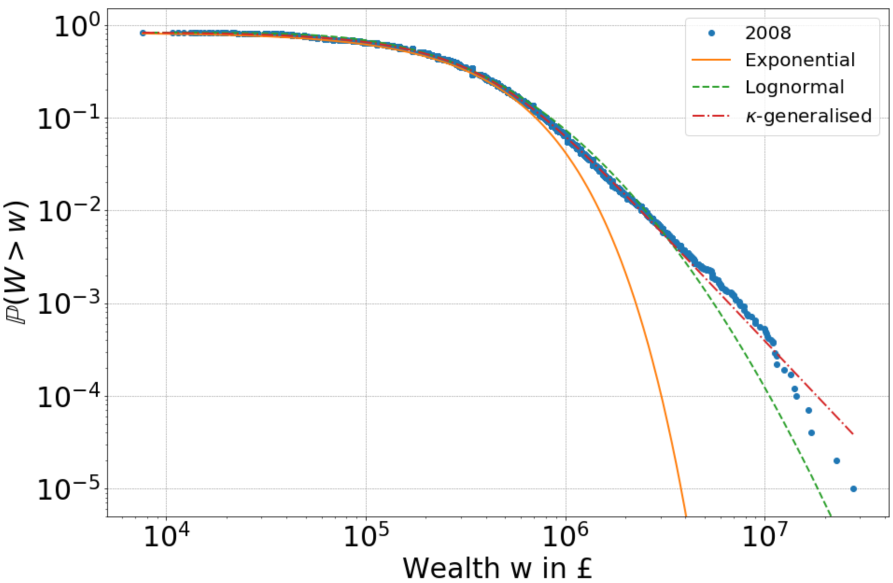

Figure 1.2: Empirical tail distribution of positive UK household wealth in 2008 with exponential, lognormal and κ-gen fits to WAS [116]. Parameter fits (to 2 decimal places or nearest whole number): $\hat{\mathbb{P}}(W \geq 0) = 0.83$, exponential $\hat{\mu}_+ = 336276$, lognormal $\hat{m}_+ = 257057$, $\hat{s} = 1.01$ and κ-generalised $\hat{\alpha} = 1.13$, $\hat{\beta} = 341405$, $\hat{\kappa} = 0.50$.

[11] MLE estimation is also possible for κ-gen, however as there are no analytic solutions, numerical methods would need to be used. See Section 3.1.6 of [32] for an overview of MLE applied to the κ-gen distribution.

[12] We attempted to fit a number of these distributions with NLLS but could not find fits as good as the κ-generated distribution. We do not show these fits and further analysis with for example MLE would be needed to confidently rule out that these distributions do not fit the data well.

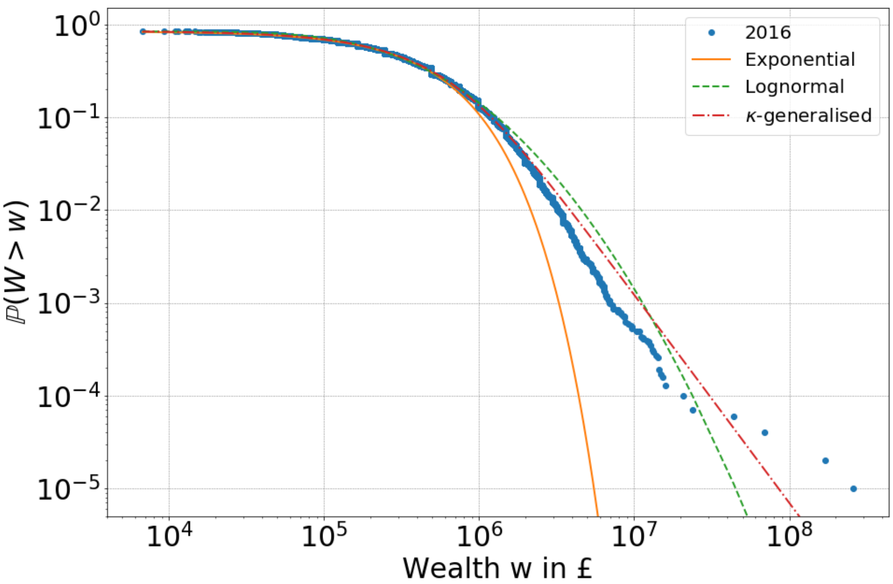

Figure 1.3: Empirical tail distribution of positive UK household wealth in 2016 with exponential, lognormal and κ-gen fits to WAS [116]. Parameter fits (to 2 decimal places or nearest whole number): $\hat{\mathbb{P}}(W \geq 0) = 0.84$, exponential $\hat{\mu}_+ = 485223$, lognormal $\hat{m}_+ = 328702$, $\hat{s} = 1.16$ and κ-generalised $\hat{\alpha} = 1.00$, $\hat{\beta} = 480765$, $\hat{\kappa} = 0.44$.

1.4.5 Empirical Results: Debt

Instead of using negative wealth let us consider the absolute value of negative wealth $|W_-|$. We fit this to a Weibull distribution as was done in Clementi et al. [34]. The tail of the Weibull is found to be

$$\mathbb{P}_{\text{Weib}}(|W_-| > |w|) = \exp\left(-\left(\frac{|w|}{\lambda}\right)^s\right).$$

for wealth $w < 0$ and $s, \lambda > 0$. We note that the Weibull distribution is a generalisation of the exponential distribution which occurs when $s = 1$. The exponential distribution was fitted to debt in the original mixture model [43]. We note for $s > 1$ the Weibull distribution is light tailed which implies that high values of debt are much less probable than high values of wealth which as we have seen is heavy tailed.

Then from (1.15) we find the tail of debt $|W|$ for $w < 0$ is

$$\mathbb{P}(|W| > |w|) = \mathbb{P}(W < 0) \exp\left(-\left(\frac{|w|}{\lambda}\right)^{s}\right). \tag{1.22}$$

See Figures 1.4 and 1.5 for the fit to debt (1.22) for 2008 and 2016 respectively where we estimate $\mathbb{P}(W < 0)$ from the data and then using the NLLS method, as described in Section 1.4.3, we fit parameters λ and s.

In Appendix A.3 we have the parameter fits in Table A.4 and tail plot fits in Figure A.2 for all WAS debt data for biennial years 2008-2016 [116].

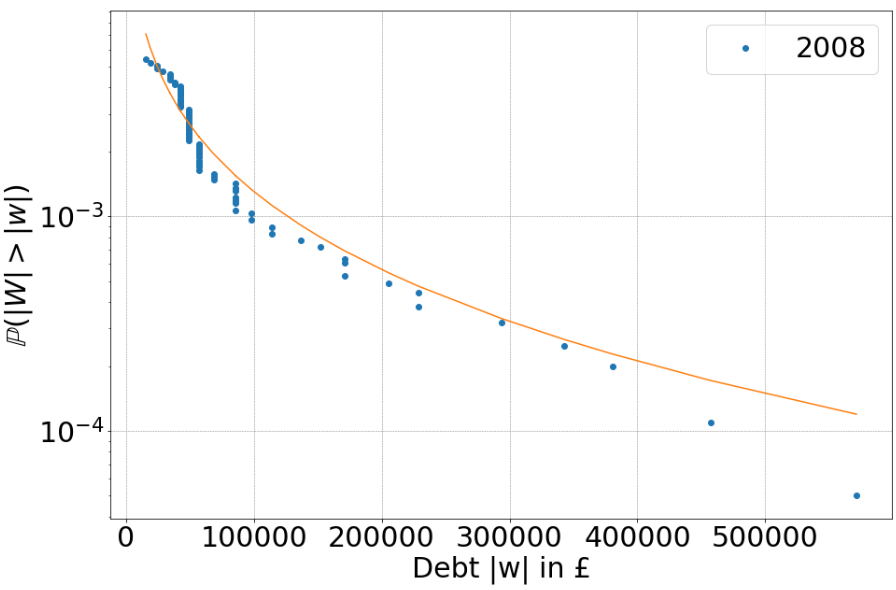

Figure 1.4: Empirical tail distribution of UK household debt from WAS [116] in 2008 with Weibull fit (1.22). Parameter fits (to 2 decimal places or nearest whole number) $\hat{\mathbb{P}}(W < 0) = 0.17$, $\hat{s} = 0.23$, $\hat{\lambda} = 100$.

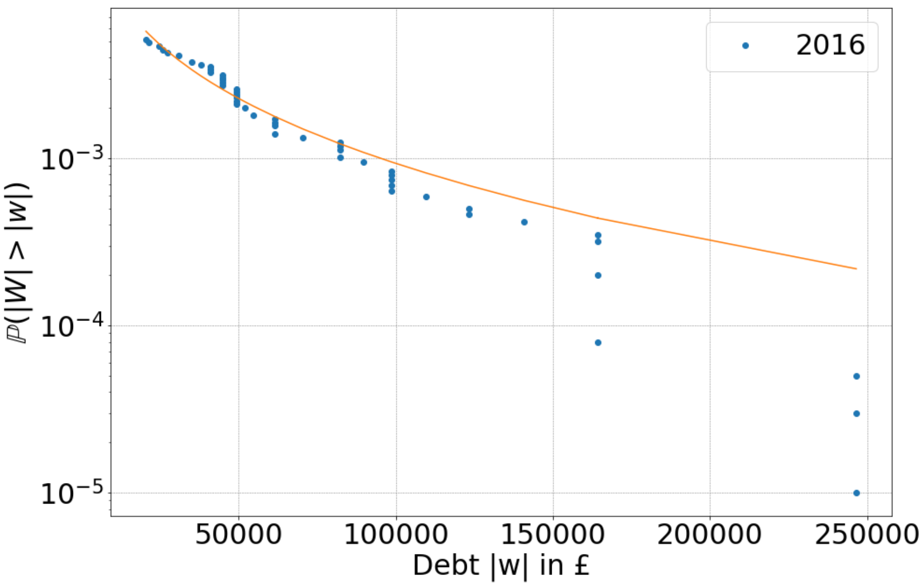

Figure 1.5: Empirical tail distribution of UK household debt from WAS [116] in 2016 with Weibull fit (1.22). Parameter fits (to 2 decimal places or nearest whole number) $\hat{\mathbb{P}}(W < 0) = 0.16$, $\hat{s} = 0.28$, $\hat{\lambda} = 261$.

1.4.6 Fitting Power Law Tails

We now focus on fitting a power law (1.7) to the upper end of the tail. Three questions with regards to power law fitting of the wealth distribution tail are:

PL.1 Does the (upper) tail follow a power law?

PL.2 Assuming the distribution has a power law tail how do you fit the power law: in particular how do you estimate the power law parameters? These are w_m, the value where the power law starts, α the log shift factor and β the power law exponent (1.8).

PL.3 Are there potentially two (or more) power laws in the tail?

We shall address point **PL.1** with discussion of a goodness of fit test in the next Section 1.4.7. A basic visual test of **PL.1** is, as mentioned previously in Section 1.3, the straight line seen in the empirical tail in log-log space. As mentioned in Section 1.4.1 we plot in log base 10. Now this test can be misleading as non-power law distributions can appear to exhibit a straight line log-log empirical tail such as the lognormal [98]. However in our case we have reason to rule out in particular a lognormal as its fit does not capture the straight line shape of the upper tail, see Figures 1.2 and 1.3.

Making the assumption that the tail is power law we shall now consider question **PL.2**. To fit w_m we again use a visual approach and set

$$w_m : \mathbb{P}(W > w_m) = 0.1 \,. \tag{1.23}$$

This seems to roughly agree where the exponential distribution stops fitting the tail which is roughly for the top 10% richest households, see Figures 1.2 and 1.3. We note that Section 3.3 of [31] contains a detailed description of more rigorous methods to fit w_m.[13]

Once our w_m is chosen via (1.23) we then estimate the remaining power law parameters α and β. We note that fitting the remaining parameters α and β with a greater w_m should theoretically give the same fits as for w_m. This is because we are just fitting the slope and shift further down the line in log-log space. We now show three methods for fitting α and β given w_m.

The first method is using ordinary linear regression (OLS). We present two models with different error formulations on which we apply OLS. The first model described in [110, 52] has multiplicative errors ε_1 :

$$\mathbb{P}(W > w) = \alpha w^{-\beta} \varepsilon_1 \,. \tag{1.24}$$

Then taking logs we have

$$\begin{aligned} \log_{10} \mathbb{P}(W > w) &= \log_{10} \alpha - \beta \log_{10} w + \log_{10} \varepsilon_1 \Rightarrow \\ y &= a + bx + \varepsilon \end{aligned} \tag{1.25}$$

where $y = \log_{10} \mathbb{P}(W > w)$, $a = \log_{10} \alpha$, $b = -\beta$, $x = \log_{10} w$ and $\varepsilon = \log_{10} \varepsilon_1$ is the error term. By the Gauss-Markov Theorem (G-M) the OLS estimates $\hat{a}$ and $\hat{b}$ for a and b respectively are unbiased and have minimum variance if the errors ε have zero mean, finite variance and are uncorrelated [106]. Taking for simplicity

$$\log_{10} \varepsilon_1 \sim \mathcal{N}(0, 1) \tag{1.26}$$

i.i.d. then the G-M conditions are satisfied. We can then estimate the OLS power

[13] This was attempted but did not appear to give good results for our data and so is not presented in the thesis and is left for future work.

law parameters α and β as

$$\hat{\alpha}_{\text{OLS}} = 10^{\hat{a}}, \quad \hat{\beta}_{\text{OLS}} = -\hat{b}. \tag{1.27}$$

However transforming back (1.25) with estimators (1.27) gives

$$\mathbb{P}(W > w) = \hat{\alpha}_{\text{OLS}} w^{-\hat{\beta}_{\text{OLS}}} \varepsilon_1 \,.$$

Therefore if we predict with

$$\hat{\mathbb{P}}(W > w) = \hat{\alpha}_{\text{OLS}} w^{-\hat{\beta}_{\text{OLS}}} \tag{1.28}$$

we will be ignoring the factor ε_1. Now assuming (1.26) we have that ε_1 is a lognormal distribution and so $\mathbb{E}[\varepsilon_1] > 0$. Therefore we will be biasing the prediction by on average $\mathbb{E}[\varepsilon_1]$. Estimating ε_1 with the residuals to account for this prediction bias is discussed in [110].

The second model is the following with additive errors

$$\mathbb{P}(W > w) = \alpha w^{-\beta} + \varepsilon_2 \,, \tag{1.29}$$

with the errors, ε_2, following the conditions of G-M. Assume for simplicity that $\varepsilon_2 \sim \mathcal{N}(0,1)$. Then

$$\log_{10} \mathbb{P}(W > w) = \log_{10} \alpha - \beta \log_{10} w + \log_{10}(1 + \tilde{\varepsilon}_2)$$

with $\tilde{\varepsilon}_2 = \varepsilon_2/(\alpha w^{-\beta})$. By Jensen's inequality, as log is a strictly concave function and all errors are assumed not to be equal, we have

$$\mathbb{E}[\log_{10}(1 + \tilde{\varepsilon}_2)] < \log_{10} \mathbb{E}[1 + \tilde{\varepsilon}_2] = 0 \,.$$

Therefore the errors $\log_{10}(1+\tilde{\varepsilon}_2)$ break the G-M condition and so the OLS estimates are not guaranteed to be unbiased. Methods for correcting the prediction bias in this case are seen in [52, 148].

In summary OLS provides a simple way to fit power laws however depending on the errors the fits may be biased. We do not carry out an analysis of the errors in this thesis and so use no corrections on the OLS prediction (1.28).

A second method to fit the power law parameters α and β is using NLLS, see Section 1.4.3. Assuming the additive error model (1.29) then the NLLS estimators

(1.16) are the following

$$(\hat{\alpha}_{\text{NLLS}}, \hat{\beta}_{\text{NLLS}}) = \arg\min_{(\alpha,\beta)} \sum_{i=1}^{N} \left(\mathbb{P}_N(W > w_i) - \frac{\alpha}{w_i^{\beta}} \right)^2 . \tag{1.30}$$

We note that (1.30) has no analytic solutions, see Example 2.1, p22 [132] and so numerical methods must be used. Assuming the model (1.29), in particular that the errors are i.i.d. with mean zero, and that (α, β) is in a bounded domain then as the power law is continuous for α and β greater than zero the estimate (1.30) exists, see Chapter 12 of [132]. Under a few more conditions (1.30) is consistent and asymptotically unbiased, see again Chapter 12 of [132]. We note that (1.30) is identical to maximum likelihood estimation on the model (1.29) assuming the errors ϵ_2 are i.i.d. normal, see Section 2.2 of [132].

The final method we use is maximum likelihood estimation (MLE) on the sample from the power law tail [31]. MLE assumes a distribution and maximises over the parameter space to find which parameter value is most likely to generate the sample. Now as we have mentioned, our data is of the form of N empirical points (1.11). Note first we have the minimum wealth value for the power law $\hat{w}_m$ already chosen from (1.23). To generate an i.i.d. sample $\mathbf{w} = \{w_1, w_2, \ldots, w_n\}$ from the power law tail, $w_i \geq \hat{w}_m$, for all $i = 1, 2, \ldots, n$, we select $u_i \sim$Uniform$[0, 0.1]$ [14] and find the closest w_i such that $\mathbb{P}(W > w_i) = u_i$. As the entire sample is from a power law we assume this sample is from a Pareto distribution, see section 1.3. Thus the density of W given the sample $\mathbf{w}$ is assumed of the form

$$f_{W|\mathbf{w}}(w) = \frac{\beta \hat{w}_m^{\beta}}{w^{\beta+1}} .$$

The likelihood function, $L(\mathbf{w}; \beta)$, is the product distribution of the sample:

$$\begin{aligned} L(\mathbf{w}; \beta) &= f_{W|\mathbf{w}}(\mathbf{w}) \\ &= \prod_{i=1}^{n} \frac{\beta \hat{w}_m^{\beta}}{w_i^{\beta+1}} \qquad \text{as } w_i \text{ i.i.d.} \end{aligned}$$

To maximise the likelihood function over β it is convenient to maximise instead over the log-likelihood function $l(\mathbf{w}; \beta) = \log L(\mathbf{w}; \beta)$. As the log function is monotonically increasing, maximising over the log-likelihood is the same as maximising over the

[14] Assuming the power law part of the tail is for the top 10% as in (1.23).

likelihood. Thus

$$\begin{aligned} l(\mathbf{w};\beta) &= \sum_{i=1}^{n}(\log\beta + \beta\log\hat{w}_m - (\beta+1)\log w_i) \\ &= n\log\beta - \beta\sum_{i=1}^{n}\log\frac{w_i}{\hat{w}_m} - \sum_{i=1}^{n}\log w_i \end{aligned}$$

Then maximising over the log-likelihood by taking the partial derivative and setting equal to 0, $\frac{\partial}{\partial\beta}l(\mathbf{w};\beta) = 0$, gives the maximum likelihood estimator for β as [15]

$$\hat{\beta}_{\text{MLE}} = \frac{n}{\sum_{i=1}^{n}\log\frac{w_i}{\hat{w}_m}}. \tag{1.31}$$

For a summary of several properties of the estimator (1.31), in particular that it is consistent and efficient, see [31]. We note however that the estimator (1.31) is not unbiased. It can be seen, see Example 7.3.1, p428 of [76] that

$$\mathbb{E}[\hat{\beta}_{\text{MLE}}] = \frac{n\beta}{n-1}$$

so that (1.31) is biased however it is an asymptotically unbiased estimator. It is proven in Theorem 6.2.2, p369 [76] that under certain mild regularity conditions MLE estimators are asymptotically unbiased.

Then the prediction for the tail fitted to the sample is

$$\hat{\mathbb{P}}(W > w|\mathbf{w}) = \left(\frac{\hat{w}_m}{w}\right)^{\hat{\beta}_{\text{MLE}}}.$$

Now given $w > w_m$ where in our case $\hat{w}_m$ is chosen from (1.23)

$$\begin{aligned} \mathbb{P}(W > w|\mathbf{w}) &= \mathbb{P}(W > w|W > \hat{w}_m) \\ &= \frac{\mathbb{P}(W > w)}{\mathbb{P}(W > \hat{w}_m)} \Rightarrow \\ \mathbb{P}(W > w) &= \mathbb{P}(W > \hat{w}_m)\mathbb{P}(W > w|\mathbf{w}) \end{aligned} \tag{1.32}$$

Therefore our power law prediction with MLE for the distribution W by (1.32) for

[15] We know this is a maximum by the second derivative test and assuming $\hat{\beta} > 0$.

$w > \hat{w}_m$ is

$$\hat{\mathbb{P}}(W > w) = \mathbb{P}(W > \hat{w}_m)\left(\frac{\hat{w}_m}{w}\right)^{\hat{\beta}_{\text{MLE}}}.$$
$$= 0.1\left(\frac{\hat{w}_m}{w}\right)^{\hat{\beta}_{\text{MLE}}}, \quad \text{by (1.23)}$$

and thus

$$\hat{\alpha}_{\text{MLE}} = 0.1\hat{w}_m^{\hat{\beta}_{\text{MLE}}}. \tag{1.33}$$

We note that the MLE estimators for α and β, respectively 1.33 and 1.31, are the same as the Hill estimators [73].

See Figures 1.6 and 1.7 for power law parameter fits with OLS (1.27) (with no bias correction), MLE (1.31), (1.33) and NLLS (1.30) for 2008 and 2016 respectively. We note in these figures we only show the tail from w_m (1.23) onwards. The power law exponents β roughly agree for MLE and NLLS but give significantly different results compared to OLS especially in 2016. In 2016, visually the OLS method seems to work better than the MLE and NLLS method however the log-log plot can be deceiving due to most points occurring for lower wealth. We note the erratic nature of the data points in this year.

We also plot the κ-gen fits in Figures 1.6 and 1.7 which are the same fits from Figures 1.2 and 1.3 respectively. We can see that the κ-gen fits are in the same approximate range with power law exponents (1.21) agreeing with the power law fits in some years better than others.

In Appendix A.4 Tables A.5 and A.6 show the power law parameter fits and Figure A.3 shows the tail plot power law fits for WAS biennial years 2008-2016 [116].

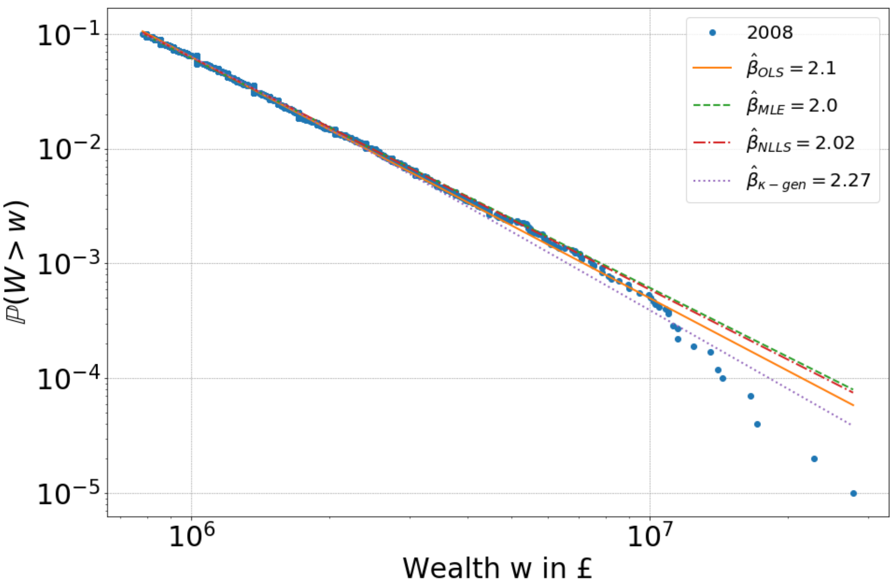

Figure 1.6: Empirical tail distribution of positive UK household wealth greater than w_m (1.23) in 2008 with κ-gen fit and power law fits using OLS (1.27), MLE (1.31), (1.33) and NLLS (1.30) from WAS [116].

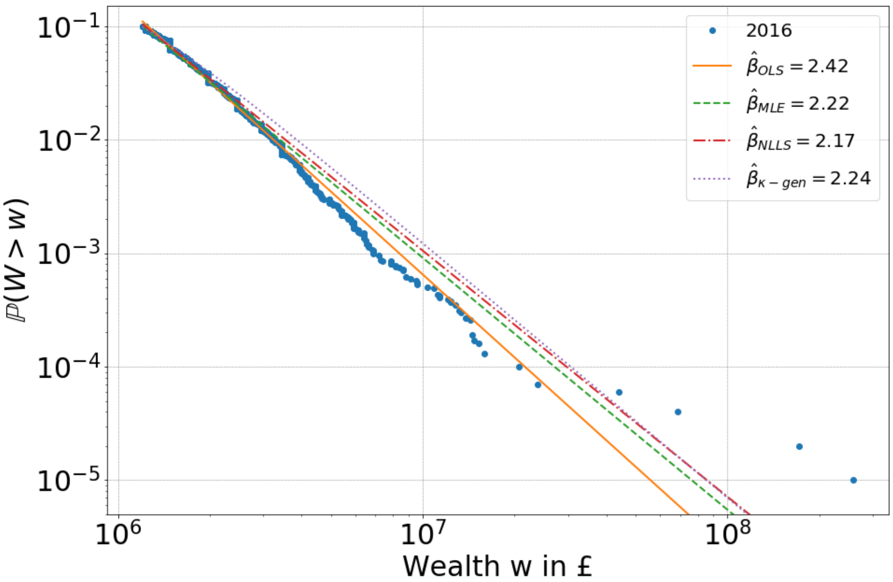

Figure 1.7: Empirical tail distribution of positive UK household wealth greater than w_m (1.23) in 2016 with κ-gen fit and with power law fits using OLS (1.27), MLE (1.31), (1.33) and NLLS (1.30) from WAS [116].

Finally, motivated from Figure 1.1 we consider the possibility that the tail is approximated by two power laws **PL.3**:

$$\mathbb{P}(W > w) = \begin{cases} \dfrac{\alpha_1}{w^{\beta_1}} & \text{for } 0 < w_{1,m} < w \leq w_{2,m} \\ \dfrac{\alpha_2}{w^{\beta_2}} & \text{for } w > w_{2,m} \end{cases} \tag{1.34}$$

with all parameters in $\mathbb{R}_{>0}$ and $\beta_1 \neq \beta_2$. We know of no literature for this case and instead it has been assumed that the appearance of two power laws is due to survey bias as discussed in Sections 1.2 and 1.5. Thus it has been thought that there is in reality only one power law in the tail [145]. We shall show reasoning for this bias distortion in Section 1.5. However the extent of the difference in the power law exponent between survey and rich lists gives plausibility to the two power law hypothesis. Beyond the UK we see that the US does not exhibit two power laws whereas many European countries such as France and the Netherlands do, see appendix of [145]. One could hypothesise that a higher level of redistribution avoided by the those in the rich lists could be responsible for the difference in power law exponents in European countries.

Figures 1.8 and 1.9 show fits for OLS (1.27) and MLE (1.33) and (1.31) to the Forbes rich lists [1] as well as previous fits to the WAS richest 10% [116] as seen in Figures 1.6 and 1.7. We note that in this case (possibly due to the small size of the rich lists) we find no numerical solutions for NLLS parameter fits to the rich lists.

In Appendix A.5 Tables A.7 and A.8 contain the power law parameter fits and Figure A.4 shows the tail plot fits for Forbes rich lists biennial years 2008-2016 [1].

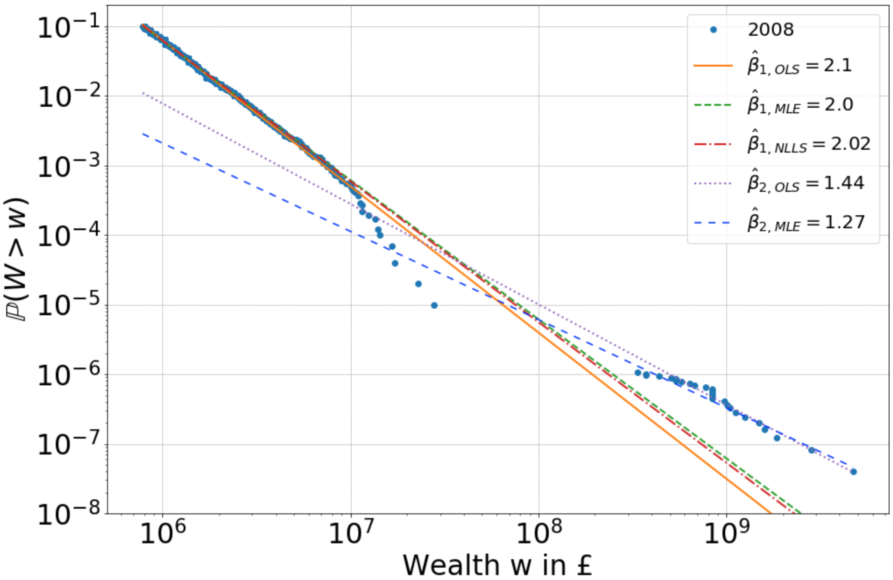

Figure 1.8: Empirical tail distribution of positive UK household wealth greater than w_m (1.23) in 2008 with power law fits using MLE (1.31), (1.33) and OLS (1.27) from WAS [116] and Forbes rich list [1].

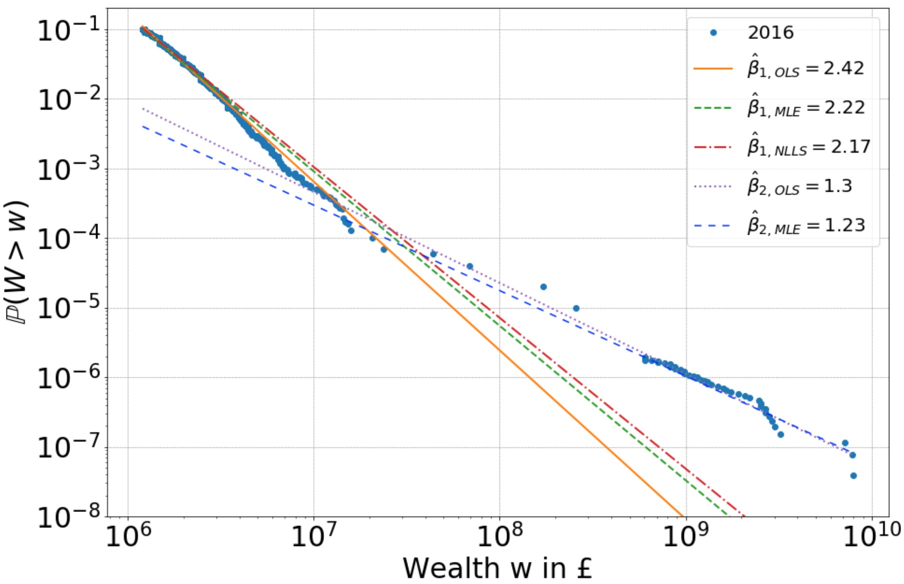

Figure 1.9: Empirical tail distribution of positive UK household wealth greater than w_m (1.23) in 2016 with power law fits using MLE (1.31), (1.33) and OLS (1.27) from WAS [116] and Forbes rich list [1].

1.4.7 Goodness of Fit

Throughout this section we have fitted various distributions to wealth data. We can compare fits of different distributions by using a measure such as the cost function (1.17) which is proportional to the mean squared error. However suppose we want to compare the fitted distribution directly with the data itself. In particular we want to know how likely the data comes from a specific distribution. For this purpose we can use a goodness of fit test. One popular goodness of fit test is the Kolmorogov-Smirnoff test (KS) [136, 75]. We can formulate the null hypothesis for the KS test as

$$\mathrm{H}_0 : \text{the sample data is generated by the particular distribution in question.}$$

The KS test statistic is then defined as

$$\hat{T} := \max_{w > w_m} |\mathbb{P}_{D(\hat{\Theta})}(W > w) - \mathbb{P}_N(W > w)|$$

where $\mathbb{P}_{D(\hat{\Theta})}(W > w)$ is the tail fit of a particular distribution $D(\Theta)$ to the empirical tail $\mathbb{P}_N(W > w)$ of N data points. However if we estimate the parameters for the distribution using the data the KS test is no longer valid as the data and the fit are not independent. To avoid this independence issue a Monte-Carlo approach can be used that is outlined in [111] and described below:

MC.1 First fit the data with the chosen distribution. Then calculate the KS test statistic $\hat{T}$.

MC.2 From the fitted distribution generate n samples and for each sample i fit again with the distribution and calculate the KS test statistic $\hat{T}_i$ to the fit.

MC.3 The p value for H_0 is then the proportion of sample test statistics greater than the data test statistic

$$p = \frac{1}{n}\sum_{i=1}^{n} \mathbf{1}_{\hat{T}_i > \hat{T}} \,.$$

We attempted this Monte-Carlo approach for the power law fits on the power law proportion of the WAS data (not including the rich list data) and found p-values of 0 thus rejecting the power law hypothesis. This agrees with wealth data studied by Clauset et al [31]. For this Monte-Carlo method to give p-values that do not reject H_0 the original data would have to fit the distribution with no more error than artificially sampling from the distribution. This test may be too harsh and a more lenient method may be studied for future work. The existence of the power law assumption in the wealth distribution in this thesis therefore rests primarily on the inexact log-log visual straight line test.

1.5 Bias in Survey Data

As previously discussed in Sections 1.2 and 1.4.6 the presence of bias in the survey may account for the difference in the exponents of the power laws between the survey and rich lists as seen in Figures 1.1, 1.8 and 1.9. To show why this is possible we will set up a simulated sampling situation which is a simplified version of what is done in [145]. Assume there are 27 million households, a rough estimate for the number of households in the UK [116], and that the richest top 10% follow a Pareto distribution with $w_m = 10^6$ and $\beta = 1.3$. From the top 10% of the population we sample without replacement $n = 3000$ households, roughly the same number of the top 10% in the wealth and asset surveys [116], in two ways. Firstly we sample uniformly, secondly according to the rule

$$\mathbb{P}(\text{household } i \text{ sampled}) = \frac{w_i^\gamma}{\sum_{j=1}^{n} w_j^\gamma} \tag{1.35}$$

where w_i is the wealth of household i and $\gamma < 0$ so that we bias against choosing wealthier households. As mentioned in Section 1.2 this is an example of differential non-response bias. Figure 1.10 shows the results of the simulations with the two types of sample. As can be seen the uniform sample has very similar exponent to the population but the biased sample with $\gamma = -1$ has an exponent that increases significantly. Thus with high bias it is possible that the survey data has an exponent that is significantly higher than the rich list exponent. However we question whether this level of bias is present in the survey data. As mentioned in the previous section the presence of both one and two power law tails are apparent in other countries with for example two tails in France whereas just one in the USA, see online appendix of [145]. It would be future research to completely determine whether one or two power laws are present in wealth data and how this differs between countries.

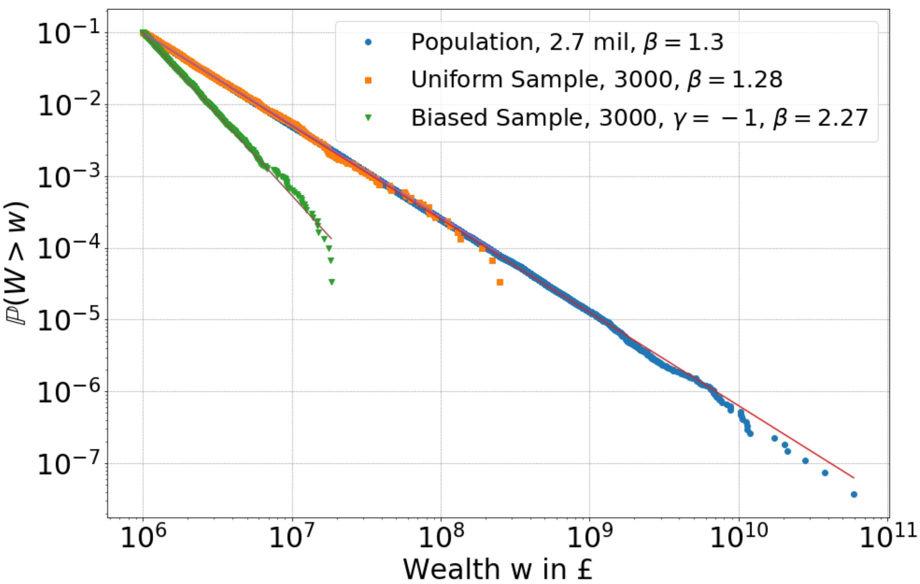

Figure 1.10: Sampling experiment: start with population top wealthiest 10% 2.7 million from Pareto$(10^6, 2)$ (blue) then sample 3000 from this top 10% uniformly (orange) or with power law rule (1.35) with $\gamma = -1$ biasing against choosing wealthy households (green). Power law fits are with MLE (1.31), (1.33)

1.6 Wealth Inequality

Though not the main focus in this thesis, wealth inequality has been a major topic of investigation in economics [121, 44, 71] and is measured for simulations in Chapter 4. How wealth is distributed among society matters to a great extent as high levels of inequality has likely been one of the reasons for social unrest throughout history [39]. On the other hand it has yet to be seen if a relatively equal modern society is possible with equalising attempts often leading to more harm than good [87]. The presence of heavy tails and power laws in the wealth distribution indicates there is vast inequality [17]. Large inequality seems to have occurred at a similar time to when complex civilisation originated [128, 46]. It has been noted that since the 1980s inequality has been rising amongst many nations globally [121, 9]. Potential contributing factors include globalisation, financialisation, decreased taxes, increased tax evasion and avoidance, increased inheritance and domination of the technological sector [121, 9, 12]. There is on going debate amongst economists about what policies, if any, should be implemented to reduce inequality.

There are several measures of inequality each with their various merits [26, 37, 121]. In this thesis we use the standard Gini coefficient $g \in [0, 1]$ and top 1% wealth share $s_{0.01} \in (0, 1]$. The Gini coefficient, often used for measuring income inequality, can be thought of as a measure of the difference between any two randomly selected agents wealth. The top 1% wealth share is defined as the proportion of wealth held by the richest 1% of the population. The standard formulation of the Gini is calculated only for positive wealth as including negative wealth values can give a $g > 1$ [28].

For a non-decreasing ordered sample of N agents' wealth $w_1 \leq w_2 \leq \cdots \leq w_N$ with total wealth $\mu = \sum_{i=1}^{N} w_i$ we define[16]

$$s_{0.01} := \sum_{i>0.99N}^{N} w_i \Big/ \mu \quad \text{and} \quad g := \frac{2}{N} \sum_{i=1}^{N} i w_i \Big/ \mu - \frac{N+1}{N}. \tag{1.36}$$

We note that higher g and $s_{0.01}$ indicates higher inequality bounded by the two extreme cases:

1. Perfect or total equality: $w_1 = w_2 = \cdots = w_N \Rightarrow g = 0$ and $s_{0.01} = 0.01$;
2. Perfect or total inequality: $w_i = 0$ for $i = 1, 2, \ldots, N-1$ and $w_N > 0 \Rightarrow g = 1$ and $s_{0.01} = 1$.

[16] We note there are various formulations for the Gini but we use the one given in [78].

For a continuous probability distribution for positive wealth W_+ defined by probability $\mathbb{P}$ then g can be formulated as

$$g = 1 - \frac{1}{\mathbb{E}[W_+]} \int_0^\infty \mathbb{P}(W_+ > w)^2 \mathrm{d}w \tag{1.37}$$

The formula for g (1.37) was first found by Dorfman [47]. We note that g is only well defined for power law tails with exponent $\beta > 1$ corresponding to a finite mean. One of the primary issues with the Gini is that it does not cater well for extremes, for example the Gini for a Pareto distribution with $\beta > 1$ is $1/(2\beta - 1)$ which for all $\beta > 3/2$ is less than the Gini for an exponential distribution which is always 0.5 [78].

Estimates for the UK top one percent wealth share (see Figure 1.11, copied here from [9]) decreased significantly from 1895 until around 1985, and has since been increasing.[17] We note that in the US estimates for top wealth shares have increased in a more dramatic fashion as compared to the UK, see Figure 1 in [127].

From the WAS we find estimated Gini for UK total wealth has remained relatively stable over the years 2008-2016 at around 0.6 [116]. We note that the Gini can be calculated for different components of wealth as was done in WAS [116]. From 2008-2016 the Gini was highest for UK financial wealth estimated above 0.8 [116]. See Figure 1.12 for the Gini and top 1% wealth share for total wealth and wealth components estimated from the WAS data over 2008-2016 using the Lorenz curves [116]. We see a roughly flat measure of total inequality during this time, slightly rising property wealth inequality and notable jumps in financial and pension wealth inequality The WAS survey data, as well as likely exhibiting bias, may be undervaluing certain assets and thus underestimating inequality [7]. As noted above the Gini may not be adequate to pick up significant changes in the very wealthy. It also requires a fair amount of granulated data to estimate perhaps explaining the lack of historical Gini measures of wealth in the literature.

Another way of estimating inequality is looking at the value of the power law β (1.7). Lower β indicates higher inequality as it means there is a higher probability of richer agents. Looking specifically at the MLE estimates of β (1.31), we see in the WAS that there were lower values in 2008 and 2010 and higher values in 2012, 2014 and 2016, see Table A.6 and Figure A.3. Looking at the Forbes rich lists we see no real relationship in the βs, see Table A.8 and Figure A.4. This may be because in the UK over these years we do not have many data points. Thus over this relatively

[17] Or more specifically non-decreasing may be a more accurate term as we can see from Figure [9] there may be times of roughly flat wealth inequality in the 2000s.

short time period of 2008-2016 we do not find any exact trends for inequality in terms of the βs.

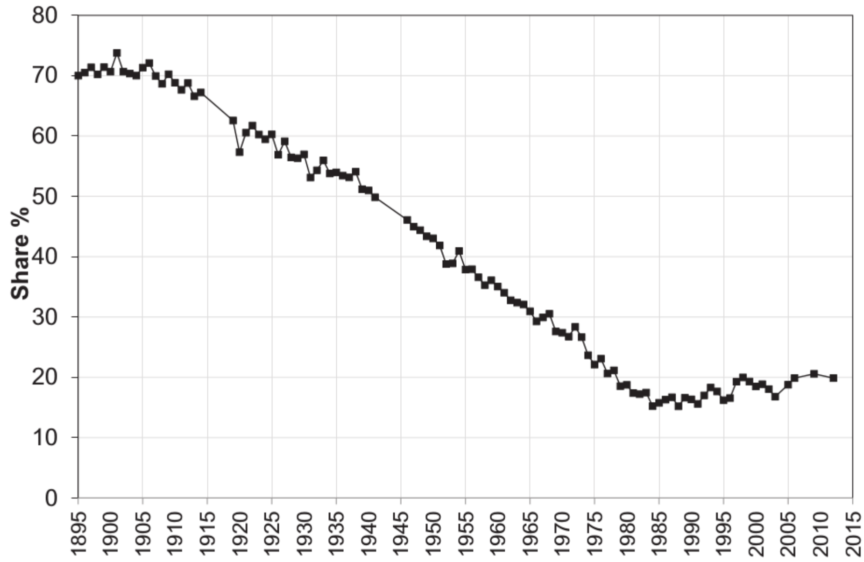

Figure 1.11: Top 1% UK wealth share, $s_{0.01}$, from 1895-2013: Figure 1 in [9]

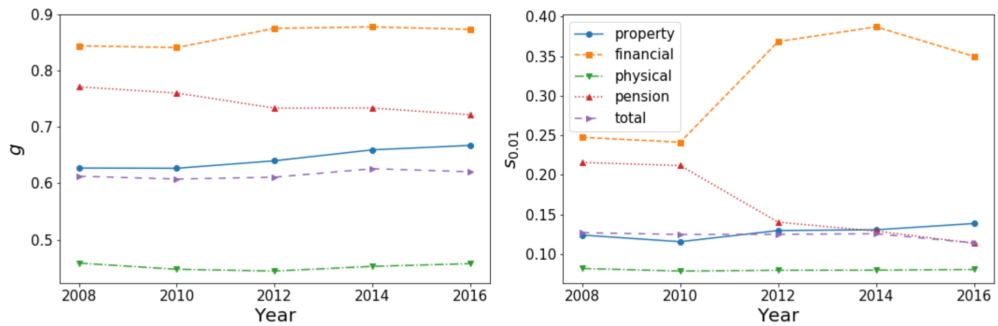

Figure 1.12: Gini, g, and top 1% UK wealth share, $s_{0.01}$, estimated using the Lorenz curve from WAS biennially 2008-2016 for total wealth and wealth components [116]. Note the likely underestimate of $s_{0.01}$ for total wealth in comparison to Figure 1.11

1.7 Discussion

This chapter's main focus was fitting the wealth distribution using the WAS and rich list data. We discussed how it can be useful to use a mixture distribution to separate negative and positive wealth. Using the NLLS method we find an adapted Weibull and κ-generalised distribution fit well the debt and positive wealth distribution respectively. An analysis of residuals, different methods of fitting such as MLE and a comparison of different distribution fits for wealth is left to future work.

We also gave three methods for fitting power laws to the upper tail. Linear regression in log-log space, has been criticised because of issues in the errors which we outlined. However we see that the linear regression method does give a rough approximation for the β parameter. The NLLS method avoids the error transform issue but as it is a numerical method the numerical solution may not be the optimal one. Finally the MLE fitting method is both convenient as there is an analytical solution and again suffers less of a bias issue compared to OLS as it is asymptotically unbiased, see [30] for simulations demonstrating the bias. Thus when accuracy and efficiency is required MLE is likely the best choice. We shall present various models generating power laws in the upper tail in the following chapters.

We discussed that the KS Monte-Carlo test outlined in Section 1.4.7 was insufficient to conclude the wealth data had a power law. But we felt it was inconclusive due to the slightly erratic nature of the data. It is left to future work to test more rigorously distributions fitted to wealth.

Finally we noted the change in exponent in the power law from WAS to rich lists and showed how bias can cause a change in exponent. However we think that this level of bias may be unlikely and that it could instead be caused by different means of measuring wealth between the two data sources [7]. More extensive study would be necessary to totally rule out that two power laws is not an artefact of wealth data.

Chapter 2

An Overview of Wealth Models

2.1 Introduction

In this chapter we review models that produce heavy tails or power laws which as seen in Chapter 1, are a key feature of the wealth distribution. The stochastic feedback model referred to as a balls in bins process with feedback or non-linear generalised Pólya urn in the combinatorics literature, shall be analysed further for particular cases in Chapter 3 and has been fitted to US wealth data [142]. In Chapter 4 we shall run our non-linear extension of the Kesten process with realistic UK wealth 2008 initial conditions using data presented in Chapter 1.

There have been several reviews of such models producing heavy tails, see for instance [94, 123, 64, 111, 134]. We shall present a selection of models from the literature as well as some variations. These models are by no means an exhaustive list however they show that many types of mechanisms can produce heavy tailed distributions and power laws. Throughout this chapter we view W_n or W_t as wealth at discrete time n or continuous time t. However in these models we are quite loose with what we mean by wealth and present some models that previously were applied to related quantities like money or stocks.

As mentioned in the introduction the models fall into four categories of hierarchy, exchange, feedback and multiplicative process. We further subdivide the multiplicative processes into discrete and continuous time leaving the following five categories:

1. Hierarchy (Section 2.2). This is a static model where we translate heavy tailed distributions into a hierarchy. Hierarchies are seen in many social structures including ones relating to wealth and we extend on Simon's early 1957 model [135] to see the link with the κ-generalised distribution.

2. Exchange (Section 2.3). We review models in econophysics that relate the exchange of money between agents to the transfer of energy between particles. The basic model leads to the Boltzmann-Gibbs distribution and under certain extra conditions we can produce heavy tails and power laws. A review of some of these models can be seen in Yakovenko [146].

3. Feedback (Section 2.4). We focus on the model that has been referred to as both a balls in bins process with feedback and a non-linear generalised Pólya urn in the literature [49, 118] and, as mentioned above, has been fitted explicitly to US wealth. We review key theoretical results from the combinatorics literature. Further results as well as simulations for particular power law feedback functions will be presented in Chapter 3.

4. Discrete time stochastic multiplicative process (Sections 2.5, 2.6 and 2.7). We discuss proportionate growth which was first studied by Robert Gibrat [68] and used in the seminal paper modelling income by D. G. Champernowne [25]. Then we present key theory on the Kesten process, an extension of proportionate growth, first analysed in detail by Harry Kesten [88]. A new non-linear generalisation of the Kesten process is analysed in Chapter 4.

5. Stochastic differential equations or SDEs (Sections 2.8, 2.9, 2.10, and 2.11). SDEs are the the continuous time versions of the discrete time multiplicative processes. Geometric Brownian motion has famously been applied to stock prices and produces a lognormal distribution. We discuss non-linear extensions of GBM as well as the stationary Fokker-Planck equation to find stationary solutions of SDEs often used in the econophysics literature. The well known paper by Jean-Phillipe Bouchaud was one of the first to apply this method to a model of wealth [23] and we also find an SDE that produces a stationary κ-generalised distribution with the same methodology.

The models discussed in this chapter can also be classified into two classes: stationary and non-stationary. In the stationary case we have

$$W_t \to \mathcal{D}(\Theta) \quad \text{as} \quad t \to \infty$$

where the quantity W_t tends to $\mathcal{D}(\Theta)$, some distribution with time independent parameters Θ after some large time. In the non-stationary case, although W_t may asymptotically be described by a particular distribution, the parameters of the distribution will be time dependent. As noted in Section 1.1 wealth has been growing rapidly since the industrial revolution suggesting that a non-stationary model is best.

However a normalisation of wealth may be stationary and so in this case a stationary heavy tailed distribution may be suitable.

Another important factor to consider is the initial distribution. For some models it does not matter what the initial condition W_0 is, the asymptotic distribution W_t as $t \to \infty$ will be the same. However for other cases the asymptotic dynamics will depend non-linearly on the initial condition. We shall focus more explicitly on initial conditions for the models in Chapters 3 and 4.

Finally related to the initial condition one can consider the meaning of time. How long the model captures certain dynamics and at which time the model starts are important questions to ask. We shall run models with these points in mind in Chapter 4.

The following quote sums up well the limits of modelling social phenomena (where we can replace income with wealth):

> The forces determining the distribution of incomes in any community are so varied and complex, and interact and fluctuate so continuously, that any theoretical model must either be un-realistically simplified or hopelessly complicated.

The quote is by D. G. Champernowne, see Section 2 of [25], who was one of the first to model income with stochastic methods. Like Champernowne we focus on models that are un-realistically simplified and that inevitably do not explicitly take into account many potentially important factors underlying the wealth distribution.

2.2 Hierarchy

Many human and animal societies are arranged in a hierarchy. The formation of hierarchies and their stability can be analysed with models [72]. We shall present a static hierarchical model for wealth which under certain circumstances produces a power law. Our presentation here will generalise slightly the model first introduced in [135] to produce a κ-generalised distribution. We note also that a static hierarchy model has recently been extended to a dynamic model with multiple hierarchies [58].

Suppose we have a hierarchical structure where at each level $l = 1, 2, \ldots, n$ there are a proportionally decreasing number of agents. In particular, suppose at level l there are r^{n-l} agents, $r \in \mathbb{N}_{>1}$, such that each agent at level l has wealth value $w_l > 1$. We note that the wealth is increasing with the levels thus

$$w_1 \leq w_2 \leq \cdots \leq w_n \,.$$

See Figure 2.1 and Table 2.1 for an illustration of this set up. For example with $r = 2$ as seen in Figure 2.1 the number of agents halves as the level goes up by one.

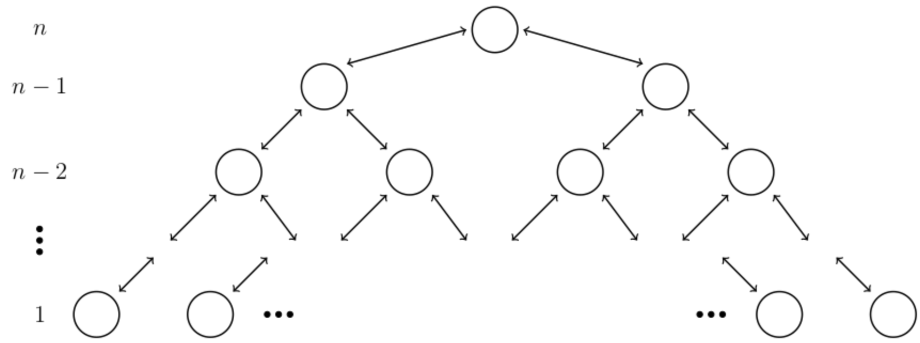

Figure 2.1: Hierarchical structure with r^{n-k} (here $r = 2$) agents at each level $k = 1, 2, \ldots, n$. Every agent at each level has the same wealth with higher wealth at a higher level. Arrows are merely for display however one could think of a dynamic model of agents going up and down the hierarchy as considered in [58].

Level	number of agents	wealth
1	r^{n-1}	w_1
2	r^{n-2}	w_2
$\vdots$	$\vdots$	$\vdots$
n	$r^0 = 1$	w_n

Table 2.1: Hierarchical structure levels with $w_1 \leq w_2 \leq \cdots \leq w_n$.

In total then, there are

$$N = \sum_{l=1}^{n} r^{n-l} = \sum_{i=0}^{n-1} r^i = \frac{r^n - 1}{r - 1}$$

agents. With W_N as the discrete RV for wealth an agent chosen uniformly at random from the N agents then we have for $\omega \in [\omega_l, \omega_{l+1})$

$$\begin{aligned}
\mathbb{P}(W_N > w) &= \frac{1}{N} \sum_{i=l+1}^{n} r^{n-i} = \frac{1}{N} \sum_{i=0}^{n-l-1} r^i \\
&= \frac{(r^{n-l} - 1)(r - 1)}{(r - 1)(r^n - 1)} = \frac{r^{-l} - r^{-n}}{1 - r^{-n}} \approx r^{-l} \qquad \text{for large } n
\end{aligned}$$

Note in particular we have $\mathbb{P}(W_N > \omega) = 1\,, \quad \omega < \omega_1$ and $\mathbb{P}(W_N > \omega) = 0\,, \quad \omega > \omega_n$. Note also we shall assume we can approximate W_N as a continuous distribution for large N. Now we shall consider wealth growing as a function of the levels l and some constant value $w_0 > 1$ thus

$$w_l = f(l, w_0)\,.$$

First suppose wealth grows linearly with the levels:

$$w_l = l w_0\,. \tag{2.1}$$

Then we have for $\omega \in [\omega_l, \omega_{l+1})$

$$\begin{aligned}
\log \mathbb{P}(W_N > w) &\approx -l \log r \\
&= -\frac{\log r}{w_0} w_l \;\Rightarrow \\
\mathbb{P}(W_N > w_l) &\approx \exp\left(-\frac{\log r}{w_0} w_l\right)
\end{aligned}$$

implying that in this case W_N can be approximated by an exponential distribution with parameter $\dfrac{\log r}{w_0}$.

Now suppose instead that wealth grows multiplicatively with the levels:

$$w_l = w_0^l\,. \tag{2.2}$$

Then we have for $\omega \in [\omega_l, \omega_{l+1})$

$$\begin{aligned}
\log \mathbb{P}(W_N > w) &\approx -l \log r \\
&= -\frac{\log r}{\log w_0} \log w_l \;\Rightarrow \\
\mathbb{P}(W_N > w_l) &\approx w_l^{-\log r/\log w_0}
\end{aligned}$$

which implies W_N is approximated by a Pareto distribution with exponent $\beta = \dfrac{\log r}{\log w_0}$. We see (counter intuitively) that higher w_0 wealth at the lowest level and lower r meaning smaller numbers of agents at each level corresponds to lower β or higher inequality.

We know from our empirical investigation from Section 1.4 that positive wealth is fitted by an exponential distribution in the lower tail and a power law in the upper tail. Our best fit for this distribution, see Section 1.4.4, was the

κ-generalised distribution. Thus we want (1.20) for $\omega \in [\omega_l, \omega_{l+1})$

$$\begin{aligned}\mathbb{P}(W_N > w) &\approx r^{-l} \\ &= \exp_\kappa\left(-\left(\frac{w_l}{\beta}\right)^\alpha\right)\end{aligned}$$

with $\alpha, \beta > 0$ and $\kappa \in (0,1)$. Using the fact that the inverse of $\exp_\kappa(x)$ (1.18) is[33]

$$\log_\kappa(x) := \frac{x^\kappa - x^{-\kappa}}{2\kappa}$$

and using this to solve for $r^{-l} = \exp_\kappa\left(-\left(\frac{w_l}{\beta}\right)^\alpha\right)$ we find the wealth at level l to be

$$\begin{aligned}w_l &= \beta\left(\frac{r^{\kappa l} - r^{-\kappa l}}{2\kappa}\right)^{1/\alpha} \\ &= \beta\left(\frac{(aw_0)^{\kappa l} - (aw_0)^{-\kappa l}}{2\kappa}\right)^{1/\alpha}. \qquad (2.3)\end{aligned}$$

where (2.3) is by setting $r = aw_0$, $a > 0$.

With $a = \alpha = 1$, $\beta = 2$ and $\kappa \approx 1$ we have by substituting in (2.3) that $w_l \approx w_0^l - w_0^{-l} \approx w_0^l$ for high levels l. Also it can be seen using L'Hôpital's rule that $\lim_{\kappa \to 0} \log_\kappa(x) = \log x$ and thus with $\beta = \alpha = 1$ and $\kappa \approx 0$ and again substituting in (2.3) then $w_l \approx l \log(aw_0)$ and thus taking $a = e^{w_0}/w_0$ we have $w_l \approx lw_0$. Therefore we can see that (2.3) is a generalisation of both (2.1) and (2.2).

2.3 Econophysics Exchange Models

Econophysics applies mathematical and modelling techniques from physics to economics and was first coined by the physicist Eugene Stanley in 1995 [146]. These physics techniques have been applied to many economic and financial phenomena [6, 138]. In particular econophysics methods have been applied to model money, income and wealth of which many studies have been compiled by Chatterjee et. al. and Yakovenko [24, 146]. In this section we model money, a subset of wealth, such that we view total money as fixed in the system and as a medium of economic exchange [146]. We relate Boltzmann's kinetic theory of gases to the exchange of money, where the transfer of energy between molecules or atoms is equivalent to the transfer of money between agents. Agents can be thought of as individuals or households in the economy. This analogy was first made by Yakovenko in 2001 [48] although Boltzmann himself remarked how his theory could potentially extend from mechanical objects to sociology [146]. The Boltzmann-Gibbs theory states that the probability of a system (we can think of this as a molecule or atom) with energy w is the Boltzmann-Gibbs distribution with density

$$\mathbb{P}(w) = Ce^{-w/T} \tag{2.4}$$

where C is a normalising constant and T is the temperature of the system. Note in the continuous case with $w \in \mathbb{R}_{\geq 0}$ that (2.4) is identical to the probability density of an exponential distribution with $C = 1/T$.

When making the analogy to money we set $W_n(i)$ to be the random variable for the amount of money an agent i has in a population of $N \in \mathbb{N}$ agents at time n. We take time $n \in \mathbb{N}_0 = \{0, 1, 2, \dots\}$ to be discrete.

Assume two different agents i and j chosen at random with money $W_n(i)$ and $W_n(j)$ respectively at discrete time n. Consider an exchange of an amount $\Delta_n(i) \in \mathbb{R}$ from i to j at time n:

$$W_{n+1}(i) = W_n(i) - \Delta_n(i) \tag{2.5}$$

$$W_{n+1}(j) = W_n(j) + \Delta_n(i) \tag{2.6}$$

and thus

$$W_n(i) + W_n(j) = W_{n+1}(i) + W_{n+1}(j) \, .$$

Therefore the total money of the two agents before and after the exchange is the same and money is conserved in the transaction. As this is true for all agents exchanging with each other, money, like energy, is conserved in the system.

Now set the exchange amount to be fixed across time and agents: $\Delta_n(i) = \Delta > 0$. Also impose a boundary condition of $W_n(i) \geq 0$ for all n and i. Then repeatedly choosing uniformly at random two unique agents i and j, $i \neq j$, from the set of N agents and performing the exchange from agent i to j, see (2.5) and (2.6), the distribution of W_n tends towards a stationary Boltzmann distribution (2.4) as $n \to \infty$ [146]. We note however that this approximation is only true if the total money in the system is much greater than the number of agents in the system [131]. The exact stationary probability mass function solution is detailed also in [131].

Without imposing any lower bound on the value of $W_n(i)$ there is no convergence to a stationary distribution. In this case agents form a roughly normal distribution with increasing variance as agents money moves increasingly in both the positive and negative direction, see [79] for an exact solution. Imposing a lower 'debt' bound of $-\infty < -w_m < 0$ such that $W_n(i) \geq -w_m$ also gives a stationary distribution W which is proportional to a Boltzmann-Gibbs distribution (2.4) for $w \geq 0$ with $T = w_m + \mathbb{E}[W]$ [146] [1]

Suppose we are again running repeated exchanges (2.5) and (2.6). However we let the exchange amount be a proportion $\lambda_n(i) \in (0,1)$, potentially dependent on agent i and time n, of the payers money:

$$\Delta_n(i) = \lambda_n(i) W_n(i)\,. \tag{2.7}$$

Assuming each agent has a positive amount of money then no agent can go into debt but may go very close to 0. This form of exchange (2.7) is referred to as multiplicative exchange [79]. Let us first focus on fixed $\lambda_n(i) = \lambda \in (0,1)$ independent of agent i and time n. In this case there is a stationary distribution W [79]. If $\lambda = 0.5$ for all agents, W is again an exponential distribution, for general $\lambda \neq 0.5$ we have that W is well fitted by a Gamma distribution. Exact solutions for the distribution of W are found by using the master equation approach which we omit here[2] [79]. It can be seen that the higher the λ the less equal the stationary distribution with the critical point at $\lambda = 0.5$ above which 'unprofitable interactions are sufficiently devastating that a large and persistent underclass is formed' [79].

Now suppose the exchange amount is again a proportion of the payers money (2.7) but the proportions for each agent are independent of the time step n but dependent on the agent: $\lambda_n(i) = \lambda(i)$. In this case choosing $\lambda(i)$ from a distribution with support $[0,1]$ such as a uniform or beta distribution we find evidence that the distribution W_n as n gets large leads to power laws, see left of Figure 2.2. We note that this power law formation is found in [27] for a more sophisticated exchange but still based on fixed agent proportions $\lambda(i)$ independent of time.

Another way to obtain a power law from an exchange model is related to the feedback function in the balls in bins process in Section 2.4 and Chapter 3. Assume that every agent's money is positive with $W_n(i) \in \mathbb{R}_{>0}$ at each time n. Instead of choosing the two agents for exchange in (2.5) and (2.6) uniformly at random, we now consider choosing the agents for exchange with the following probability:

$$\mathbb{P}(\text{Agent } i\,, j \text{ chosen for exchange at iteration } n|\mathbf{W}_n) = \frac{W_n(i)^\gamma \cdot W_n(j)^\gamma}{\left(\sum\limits_{k=1}^{N} W_n(k)^\gamma\right)^2} \tag{2.8}$$

[1] Though we note that after results in [131] that this is again only true for $\sum_i W_n(i) >> N$.

[2] We do however use this method in the balls in bins process in Chapter 3. It is future work to analyse these exchange models in more detail using the master equation.

with $\gamma \in \mathbb{R}_{\geq 0}$ and $\mathbf{W}_n = \{W_n(1), W_n(2), \ldots W_n(N)\}$. Thus agents with more money are more likely to exchange for $\gamma > 0$ and we have the previous uniform case for $\gamma = 0$. Now choose some fixed exchange amount $\Delta_n = \Delta > 0$ with a boundary condition $W_n(i) > 0$ for all n and i. With $\gamma > 1$ we obtain a power law for the distribution of W_n after a large n number of exchanges ((2.5) and (2.6)) with agents chosen according to (2.8), for a particular case with $\gamma = 3$, see right of Figure 2.2.

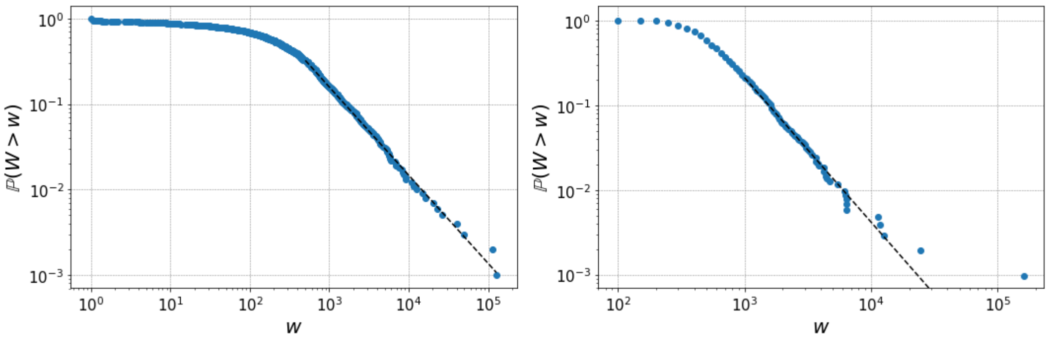

Figure 2.2: Power law tails fitted with MLE (1.31), (1.33), for two repeated exchange scenarios. Left: $N = 1000$ agents, all agents starting with $w_0 = 1000$, 10^6 repeated exchanges, fixed lower bound at 1, with exchange amount proportional to the payers money (2.7) with time independent, agent dependent proportion $\lambda_n(i) = \lambda(i)$ chosen from a uniform distribution between $(0, 1)$. Right: $N = 1000$ agents, all agents starting with $w_0 = 1000$, 10^8 repeated exchanges, with agents chosen for exchange chosen by rule (2.8) with $\gamma = 3$ and fixed exchange amount $\Delta = 50$.

2.4 Balls in Bins Process with Feedback

Let us consider $N \in \mathbb{N}$ agents with agent j having positive wealth $W_n(j) \in \mathbb{N}$ at discrete time $n \in \{0, 1, 2, \ldots\}$. We introduce the probability that an agent gains a certain amount of wealth $\omega \in \mathbb{N}$ at time $n + 1$ as follows, letting $\mathbf{W}_n = \{W_n(1), W_n(2), \ldots W_n(N)\}$, then

$$\mathbb{P}(W_{n+1}(j) = W_n(j) + \omega | \mathbf{W}_n) = \frac{f(W_n(j))}{\sum\limits_{i=1}^{N} f(W_n(i))} \tag{2.9}$$

where $f : \mathbb{N} \to \mathbb{R}_{>0}$ is some continuous function referred to as the **feedback function**. Thus the probability that the agent j gains ω at time n is proportional to some function of the agent's current wealth.

We have made the simplifications that both wealth and ω are strictly positive and discrete and that f goes to a positive co-domain to guarantee a non-zero denominator in (2.9). Therefore each agents wealth will either stay the same or increase at each iteration. We also assume for simplicity ω is fixed though it could be extended to a random value. We assume wealth is discrete rather than continuous to make analogy to the following so-called balls in bins process with feedback [114].

Definition 5 (Balls in bins process with feedback)**.** *Set $W_n(j)$ as the number of balls in bin j out of N distinct bins at time n and* (2.9) *as the probability that ω new balls go into bin j at time $n+1$. Then repeated applications of* (2.9) *is called a **balls in bins process with feedback** f.*

We shall refer to the process in Definition 5 simply as the balls in bins process. We note that as (2.9) only depends on the previous time then the discrete time RVs $\mathbf{W}_n$ forms a Markov chain. Another way to think of the balls in bins process is with a Pólya urn. For this analogy set $W_n(j)$ as the number of balls of colour i out of N distinct colours in an urn at time n and (2.9) as the probability that ω new balls of colour i go into the urn at time $n+1$. Then repeated applications of (2.9) in this case is known in the literature as an example of a **generalised Pólya urn** [118]. This model was considered in the context of the economy by W. B. Arthur a well-known figure in non-mainstream studies of economics [10].

We shall now introduce a definition of monopoly in the balls in bins process:

Definition 6 (Balls in Bins Monopoly)**.** *We define monopoly by bin j_m in the balls in bins process if there exists an $n_m \in \mathbb{N}$ such that for all iterations $n \geq n_m$ bin j_m takes every new ball added to the system almost surely.*

Thus Definition 6 means that monopoly by an agent j_m is achieved when for all $n \geq n_m$

$$\mathbb{P}(W_{n+1}(j_m) = W_n(j_m) + \omega | \mathbf{W}_n) \stackrel{a.s.}{=} 1\,.$$

We note that Definition 6 is stronger than one bin taking all but a negligible fraction of the balls in the limit. Definition 6 states that one bin gets all but **finitely many** balls. The following theorem gives a condition on the feedback function f for monopoly to occur.

Theorem 1 ([90],[113])**.** *Monopoly for some agent j_m (as defined in Definition 6) occurs with probability* 1 *in the balls in bins process with feedback $f : \mathbb{N} \to \mathbb{R}_{>0}$ if the following is satisfied*

$$\sum_{i=1}^{\infty} \frac{1}{f(i)} < \infty\,. \tag{2.10}$$

Theorem 1 was first proved for the power function

$$f(x) = x^{\gamma} \tag{2.11}$$

with $\gamma > 1$ in Proposition 2 of [90]. It was extended to the more general form seen in Theorem 1 which we see in [114] and was first developed in [113]. We present essentially the same proof in [113] for feedback (2.11) in Chapter 3 where we relate the discrete time model to continuous time. Although not explicitly stated as such the balls in bins process with feedback (2.11) was used to model US wealth [142].

We shall now prove the following proposition (see Proposition 1 in [114]) which gives a class of feedback functions f such that (2.10) holds and so by Theorem 1 monopoly in the balls in bins process occurs.

Proposition 1 ([114]). *Suppose we extend the domain of the feedback function f from $\mathbb{N}$ to $\mathbb{R}_{\geq 1}$, i.e. $f : \mathbb{R}_{\geq 1} \to \mathbb{R}_{>0}$. Suppose also f is differentiable on $\mathbb{R}_{\geq 1}$, increasing with $f(1) > 0$ and satisfies the following limit:*

$$\lim_{x\to\infty} \frac{xf'(x)}{f(x)} > 1. \tag{2.12}$$

Then for this f, (2.10) *holds.*

Proof. If f satisfies (2.12) then by definition there exists $n \in \mathbb{N}$ and $c > 1$ such that

$$\begin{aligned} \frac{xf'(x)}{f(x)} &\geq c \quad \forall\, x \geq n \Leftrightarrow \\ x\frac{\mathrm{d}}{\mathrm{d}x}\log f(x) &\geq c \quad \forall\, x \geq n \Leftrightarrow \\ \log f(x) &\geq c(\log x + d) \quad \forall\, x \geq n, \text{ some } d \in \mathbb{R} \Leftrightarrow \\ f(x) &\geq Dx^{c} \quad \forall\, x \geq n, \text{ and } D = e^{cd} > 0 \Rightarrow \\ \sum_{i=n}^{\infty} \frac{1}{f(i)} &\leq \frac{1}{D}\sum_{i=n}^{\infty}\frac{1}{i^{c}} < \infty\,. \end{aligned}$$

As $f(1) > 0$ and f is increasing we know $f(x) > 0 \quad \forall\, x$. Thus $\sum_{i=1}^{n-1} \frac{1}{f(i)} < \infty$ and so f satisfies (2.10). ∎

Thus if a feedback function f satisfies the conditions of Proposition 1 then monopoly occurs in the balls in process. Examples of functions that satisfy (2.12) are included in Table 2.2.

$f(x)$
$\alpha x^\gamma (\log x)^\beta$ with fixed $\alpha > 0$, $\beta \geq 0$ and $\gamma > 1$
$x^\gamma(\alpha + \sin x)$ with $\alpha > 1$ and $\gamma > 1$
$\alpha e^{\beta x}$ with fixed $\alpha, \beta > 0$
$f(x)$ differentiable on $\mathbb{R}_{>1}$ s.t. $\exists\ c > 1$ s.t. $f(x) \geq \alpha x^\gamma\ \forall\ x \geq c$ with fixed $\alpha > 0$ and $\gamma > 1$

Table 2.2: Examples of functions satisfying (2.12).

Monopoly is an interesting result of the balls in bins process as we often see a less extreme version of it in the economy. For instance market domination of companies in multiple sectors such as steel and oil has happened numerous times leading to very wealthy individuals such as Andrew Carnegie [125]. More recently the advent of new technology has seen companies such as Google and Amazon dominate their respective markets leading to their owners being at the top of the rich lists [107]. Thus the monopoly of the balls in bins process occurring for feedback function f satisfying Theorem 1 could lead one to suspect that a similar iterative mechanism as in (2.9) does happen in the economy. However we recognise the limitations of the balls in bins process type of monopoly modelling reality: in the real world monopolies are not as strong and do not last ad infinitum.

If f satisfies (2.10) as well as two extra conditions which we do not include here, see Definition 1 in [114], then f is called a **valid feedback function** and we have the following two properties for the model with **two bins**:

- **B.1 Initial conditions matter**: if one bin contains more than half of the initial number of balls it is likely with high probability to never have less than half the proportion of balls throughout the process and thus be the monopolising bin, see Theorem 3 in [114].

- **B.2 The number of balls in the losing bin has a heavy tail**: let L be the number of balls in the losing (non-monopolising) bin then for large $w \in \mathbb{N}$ the tail of the distribution of L is heavy, in particular if f is the power function (2.11) with $\gamma > 1$ then the tail of L is a power law: $\mathbb{P}(L > w) \simeq c/w^{\gamma-1}$ for some $c \in \mathbb{R}_{>0}$, see Corollary 2 in [114].

We could hypothesise that something similar is true for more general $N \in \mathbb{N}$ bins. We shall see evidence for this in Chapter 3. Assuming **B.1** and **B.2** do generalise to N agents then we can relate this phenomenon to the wealth distribution. The fact that initial conditions dominate the outcome of the balls in bins process as in **B.1** could be compared to a lack of mobility within the wealth distribution. We also know from our empirical analysis in Section 1.4 the wealth distribution exhibits power law tails thus we also see the potential link to **B.2**.

2.5 Proportionate Growth

For this section and the next two Sections 2.6 and 2.7 we have N agents with the wealth of agent $i \in \{1, 2, \dots, N\}$ at discrete time $n \in \{0, 1, 2, \dots\}$ denoted as $W_n(i)$. However for convenience we often do not refer to a particular agent and thus write W_n instead of $W_n(i)$.

Let us assume the following update rule for each of the N agents

$$W_{n+1} = A_{n+1} W_n \tag{2.13}$$

with A_n drawn i.i.d. from a RV at each time n and for each agent. Thus we have N independent agents running in parallel. The solution to (2.13) for general n is

$$W_n = W_0 \prod_{k=1}^{n} A_k. \tag{2.14}$$

We could also write (2.13) in vector-matrix form:

$$\mathbf{W}_{n+1} = \mathbf{A}_{n+1} \mathbf{W}_n \tag{2.15}$$

where $\mathbf{W}_n$ is a vector of size N of the agents wealth and $\mathbf{A}_n$ is an N by N diagonal random matrix with the A_n of each agent on the diagonal. We note that we could have dependence between the wealth of the agents if non-diagonal elements $A_{i,j}$ of $\mathbf{A}_n$ is drawn from a random variable that is not degenerate at 0.[3] The solution to (2.15) for general n is

$$\mathbf{W}_n = \mathbf{A}_n \mathbf{A}_{n-1} \dots \mathbf{A}_1 \mathbf{W}_0\,. \tag{2.16}$$

Various convergence conditions related to the norm of the matrix product $\mathbf{A}_n \mathbf{A}_{n-1} \dots \mathbf{A}_1$ have been studied [62, 89]. We note here for the case (2.13) that if A_n takes strictly positive values then W_n becomes unbounded. This result is mentioned in [89] and is the result of Kolmogorov's three series Theorem, see Chapter IX Section 9 [57].

Equation (2.13) is known as Gibrat's proportionate growth, named after Robert Gibrat, who first studied the equation to consider a multitude of size distributions

[3] A RV X is not degenerate at 0 if $\mathbb{P}(X \neq 0) > 0$.

including firm and city size [68]. Gibrat studied the case under which W_n approximates a lognormal which can be seen as an application of the central limit theorem to the log of the product in (2.14), see **PG.2** below. We note that a particular form of this model (2.13) was considered by Yule for the size distribution for the genera of species [147]. A similar case was considered in the seminal paper for income generation by Champernowne (although he was looking specifically at agents in different income classes) [25]. Levy and Solomon considered this model specifically for generating wealth through investment returns [95].

As we previously mentioned if $A_n > 0$ then (2.14) is unbounded, thus to keep wealth bounded it can be convenient to introduce normalised wealth:[4]

$$w_n := \frac{W_n}{\langle W_n \rangle} \in [0, 1] \tag{2.17}$$

where $\langle W_n \rangle := \frac{1}{N} \sum_{i=1}^{N} W_n(i)$ is the ensemble average over the N agents at time n.

We now consider some specific cases of (2.13) producing heavy tailed distributions labelled **PG.1** and **PG.2**.

PG.1 $A_n = a > 1$ such that n is from some time independent distribution : $n \sim \mathcal{D}(\Theta)$.

We can think of this case as every agent compounding some fixed amount a at different speeds or over varying lengths of time. In this case (2.14) becomes:

$$W_n = W_0 a^n.$$

Now suppose we approximate n with an exponential distribution, commonly used for waiting times, then $n \sim \text{Exp}(\lambda)$. Then $\mathbb{P}(n > s) = e^{-\lambda s}$ and so

$$\begin{aligned}
\mathbb{P}(W_n > w) &= \mathbb{P}(W_0 a^n > w) \\
&= \mathbb{P}\left(n > \frac{1}{\log a} \log\left(\frac{w}{W_0}\right)\right) \\
&= \exp\left(\frac{-\lambda}{\log a} \log\left(\frac{w}{W_0}\right)\right) \\
&= \left(\frac{W_0}{w}\right)^{\lambda/\log a}.
\end{aligned}$$

Therefore $W_n \sim \text{Pareto}(W_0, \lambda/\log a)$. Taking $a = e$ we arrive at the well known relation that if $X \sim \text{Exp}(\lambda)$ then $W_0 e^X \sim \text{Pareto}(W_0, \lambda)$.

[4] Note w_n is N times the wealth fraction.

PG.2 Suppose A_n are i.i.d. RVs with $A_n > 0$ and n large.

Taking logs of (2.14) gives

$$\log W_n = \log W_0 + \sum_{i=1}^{n} \log A_i \Rightarrow$$
$$\log W_n - \log W_0 = \sum_{i=1}^{n} \log A_i$$

Now as the $\log A_i$ are also i.i.d. RVs with $\mu := \mathbb{E}[\log A_i] \in \mathbb{R}$ and $\sigma^2 := \mathrm{Var}(\log A_i) \in (0, \infty)$ then by the central limit theorem

$$\frac{\sum_{i=1}^{n} \log A_i - n\mu}{\sqrt{n\sigma^2}} \xrightarrow{\mathrm{d}} Z \quad \text{as} \quad n \to \infty$$

where $Z \sim \mathcal{N}(0, 1)$. Thus we have the approximation for large n:

$$\begin{aligned}\frac{\log W_n - \log W_0 - n\mu}{\sqrt{n\sigma^2}} &\approx Z \Leftrightarrow \\ W_n &\approx W_0 \exp\left(\sqrt{n\sigma^2} Z + n\mu\right) \Leftrightarrow \\ W_n &\sim W_0 \cdot \mathrm{Lognormal}(n\mu, n\sigma^2)\,.\end{aligned}$$

We also have the approximations for mean and variance of W_n by using the formulae for mean and variance of the lognormal:

$$\begin{aligned}\mathbb{E}[W_n] &= W_0 \exp\left(n\mu + n\sigma^2/2\right) \\ \mathrm{Var}(W_n) &= W_0^2 (\exp(n\sigma^2) - 1) \exp(2n\mu + n\sigma^2) \\ &= (\exp(n\sigma^2) - 1)\mathbb{E}[W_n]^2\,.\end{aligned}$$

However it is shown that these are poor approximations until a sufficiently large n is reached [122].

We could also consider the bounded situation with normalised wealth w_n (2.17). Supposing A_n takes some value a with density $f(a)$ then

$$\begin{aligned}\mathbb{P}(w_{n+1} > w) &= \mathbb{P}(a w_n > w) \\ &= \mathbb{P}(w_n > w/a) \\ &= \int_0^\infty \mathbb{P}(w_n > w/a) f(a) \mathrm{d}a \quad \text{by the law of total probability.}\end{aligned}$$

Now suppose there exists a stationary distribution $w_n \to w_\infty$ as $n \to \infty$ which has a power law with tail $\mathbb{P}(w_\infty > w) = \alpha/w^\beta$ with $\alpha > 0$ and $\beta > 0$. Then substituting into the above gives

$$\begin{aligned} \alpha/w^\beta &= \alpha \int_0^\infty (a/w)^\beta f(a) \mathrm{d}a \Leftrightarrow \\ \mathbb{E}[A_n^\beta] &= 1 \,. \end{aligned} \tag{2.18}$$

Thus assuming there is a stationary distribution and A_n i.i.d. satisfies (2.18) then the distribution has a power law tail. This heuristic solution is discussed in the review in Gabaix [63]. It can be shown that if in addition to (2.18) a positive lower barrier to wealth exists thus $w_n \geq w_l > 0$ for all n and $\mathbb{E}[\log A_n] < 0$ then the stationary distribution w_∞ with a power law tail exists [137]. These conditions are similar to when the Kesten process exhibits a stationary distribution which we see in the next section.

2.6 Kesten Process

Let us now consider a similar multiplicative process to the proportionate growth model (2.13) differing only by the additional RV B_n. Let us assume the following update rule for each of the N agents:

$$W_{n+1} = A_{n+1} W_n + B_{n+1} \tag{2.19}$$

where A_n and B_n are random variables and A_n is i.i.d. Then it can be shown, see [88], that the solution to (2.19) for general n is

$$W_n = \prod_{k=1}^{n} A_k \left(W_0 + \sum_{k=1}^{n} B_k \prod_{i=1}^{k} A_i^{-1} \right) . \tag{2.20}$$

Assuming both A_n and B_n are both i.i.d. random variables and B_n is not degenerate at 0 then (2.19) defines a **Kesten process** first rigorously studied by Kesten [88]. We also assume that the variance of both A_n and B_n are finite. The Kesten process has been used to model wealth where A_n is related to the wealth rate of return and B_n is earnings after consumption or in other words savings [18]. For an overview of Kesten models applied to wealth see Section 3 of [17].

We can write (2.19) in matrix form:

$$\mathbf{W}_{n+1} = \mathbf{A}_{n+1}\mathbf{W}_n + \mathbf{B}_{n+1}$$

where $\mathbf{W}_n$ is a vector of size N of each agent's wealth, $\mathbf{A}_n$ is an N by N diagonal matrix with the A_n of each agent on the diagonal and $\mathbf{B}_n$ is a vector of size N of the additive term B_n for each agent. Kesten considered the more general case where $\mathbf{A}_n$ is non-diagonal and so the wealth of an agent at the next iteration is dependent on other agents current wealth. We do not discuss here the dependent scenario.

Defining

$$\mu := \mathbb{E}\big[\log|A_k|\big] \in \mathbb{R} \quad \text{and} \quad \nu^2 := \mathrm{Var}\big[\log|A_k|\big] \in (0,\infty) \tag{2.21}$$

then we consider the following two cases on μ ; one leading to a stationary distribution and the other non-stationary:

Suppose $\boldsymbol{\mu} < \mathbf{0}$. This is the stationary case where Kesten proved the following result with exact details found in Theorem 5 in [88]: suppose there exists a $\beta > 0$ such that $\mathbb{E}[|A_n|^\beta] = 1$ and provided several other mild regularity conditions on the distributions of A_n and B_n are satisfied,

$$W_n \to W_\infty := \sum_{k=1}^{\infty} B_k \prod_{i=1}^{k-1} A_i \quad \text{in distribution as } n \to \infty\,, \tag{2.22}$$

for all initial conditions W_0. The stationary distribution exhibits a power law in one or both tails with parameter β, i.e. the following limits

$$\lim_{w\to\infty} w^\beta \mathbb{P}(W_\infty > w) \quad \text{and} \quad \lim_{w\to\infty} w^\beta \mathbb{P}(W_\infty < -w)\,, \tag{2.23}$$

exist and are finite, with at least one of them strictly greater than zero.

Suppose $\boldsymbol{\mu} > \mathbf{0}$. Following relatively recent results, see Theorem 2 ii) in [74], this non-stationary case can be analysed as follows. Taking absolute values and logarithms in (2.20) we have

$$\frac{\log|W_n| - \mu n}{\sqrt{n}\nu} = \frac{\sum_{k=1}^{n} \log|A_k| - \mu n}{\sqrt{n}\nu} + \frac{\log\left(\left|W_0 + \sum_{k=1}^{n} B_k \prod_{i=1}^{k} A_i^{-1}\right|\right)}{\sqrt{n}\nu}\,. \tag{2.24}$$

By the CLT for i.i.d. random variables A_k we have

$$\frac{\sum_{k=1}^{n} \log |A_k| - \mu n}{\sqrt{n}\nu} \to \mathcal{N}(0,1) \quad \text{in distribution as } n \to \infty \,.$$

Since $\mathbb{E}[\log |A_i|^{-1}] = -\mu < 0$ we have $\left| W_0 + \sum_{k=1}^{n} B_k \prod_{i=1}^{k} A_i^{-1} \right| \xrightarrow{n\to\infty} \widetilde{W}_\infty$ corresponding to the limit in the stationary case. This implies $\frac{1}{\sqrt{n}\nu} \log \left(\left| W_0 + \sum_{k=1}^{n} B_k \prod_{i=1}^{k} A_i^{-1} \right| \right) \to 0$ and

$$\frac{\log |W_n| - \mu n}{\sqrt{n}\nu} \to Z \sim \mathcal{N}(0,1) \quad \text{in distribution as } n \to \infty \,. \tag{2.25}$$

Therefore

$$\frac{|W_n|^{1/(\sqrt{n}\nu)}}{e^{\mu\sqrt{n}/\nu}} \xrightarrow{d} e^Z \sim \text{Lognorm}(0,1) \quad \text{in distribution as } n \to \infty \,.$$

To leading order of the second summand of (2.24) we get a linear dependence on the initial condition as $n \to \infty$ [5]

$$|W_n| \asymp |W_0| \exp(\mu n + \sqrt{n}\nu Z) \,.$$

We note that [74] also includes the case with $\mu = 0$ which we do not discuss here.

2.7 Generalising the Kesten Process

We could generalise the random iterative models in the preceding sections to the following form

$$W_{n+1} = f_{\theta_{n+1}}(W_n) \quad n = 0, 1, 2, \ldots \tag{2.26}$$

where $f_{\theta_{n+1}}$ is some (random) function with argument (θ_{n+1}, W_n) where $\theta_{n+1} \in \mathbb{R}^k$, $k \in \mathbb{N}$ are i.i.d. RVs. The models analysed in Sections 2.5 and 2.6 as well as the model we will see in Chapter 4 are summarised in Table 2.3.

[5] Here the symbol $\asymp$ means that $W_n = W_0 \exp\left(\mu n + \sqrt{n\nu^2} Z + o(\sqrt{n})\right)$ as $n \to \infty$, with Bachmann-Landau (or little o) notation such that $o(a_n)/a_n \to 0$ for all positive sequences $(a_n : n \in \mathbb{N})$.

Label	θ_{n+1}	$f_{\theta_{n+1}}(W_n)$
Proportionate growth (2.13)	A_{n+1}	$A_{n+1}W_n$
(Linear) Kesten process (2.19)	(A_{n+1}, B_{n+1})	$A_{n+1}W_n + B_{n+1}$
Non-linear Kesten process, see Chapter 4, (4.3)	(α_{n+1}, S_{n+1})	$W_n + \alpha_{n+1}W_n^{\gamma} + S_{n+1}$

Table 2.3: Summary of discrete random iterative models.

It is known that under certain conditions the generalised process (2.26) produces a unique stationary distribution independent of initial conditions W_0, see Section 2.4 of Collamore et al. [35]. We also know, see Section 2.6, that the linear Kesten stationary case produces power law tails. An interesting question is whether there are models of the form (2.26) that produce non-stationary distributions with power laws. We shall see evidence from simulations in Chapter 4 that the non-linear Kesten process does produce such power laws.

2.8 Stochastic Differential Equations

In the remainder of the chapter we would like to find continuous time analogues of the random discrete time iterative model (2.26) in the previous Section 2.7 that could potentially model the wealth distribution. One way to do this is through a stochastic differential equation (SDE). We define an SDE in the Itô convention noting that there are other ways of defining an SDE such as Langevin and Stratonovich [65].

We shall now view W_t as the wealth of an agent at time $t \in \mathbb{R}_{\geq 0}$, although in the subsequent sections we may view W_t as modelling related quantities such as stocks. Suppose we have an SDE in Itô form:

$$\mathrm{d}W_t = a(W_t, t)\mathrm{d}t + b(W_t, t)\mathrm{d}B_t \tag{2.27}$$

where $a(W_t, t)$ is the drift coefficient, $b(W_t, t)$ is the diffusion coefficient and B_t is Brownian motion. We note that it is possible to use the power law tailed Levy process L_t instead of Brownian motion B_t. SDEs defined with Levy process noise works well to fit power law tailed online video views [104]. We refer the reader to Chapter 5 of [112] for conditions on a and b for uniqueness and existence results of the (strong) solution to (2.27). We assume in general that strong solutions exist to the SDEs presented in this chapter.

Now let $p(w, t)$ be the probability density of a random variable W_t generated by the SDE (2.27). Then the Fokker-Planck equation is defined as follows, see Section of 5.4 [65]:

$$\frac{\partial p(w,t)}{\partial t} = -\frac{\partial}{\partial w}(a(w,t)p(w,t)) + \frac{\partial^2}{\partial w^2}\left(\frac{b(w,t)^2 p(w,t)}{2}\right). \tag{2.28}$$

Now suppose W_t tends to a stationary distribution W with density $p(w)$ independent of t so that $\frac{\partial p(w)}{\partial t} = 0$.[6] Then $a(w,t)$ and $b(w,t)$ also become independent of t which we now denote $a(w)$ and $b(w)$ respectively. Therefore (2.28) becomes

$$\frac{\mathrm{d}}{\mathrm{d}w}a(w)p(w) = \frac{\mathrm{d}^2}{\mathrm{d}w^2}\left(\frac{b(w)^2 p(w)}{2}\right)$$

which has the following solution

$$p(w) = \frac{p(w_0)b(w_0)^2}{b(w)^2}\exp\left(\int_{w_0}^{w}\frac{2a(s)}{b(s)^2}\mathrm{d}s\right) \tag{2.29}$$

where w_0 is an arbitrary constant in the domain of W [94, 65].

2.9 SDE Agent Interaction: Bouchaud's Model

We now present an SDE model of wealth with dependence between agents. We denote the wealth of agent i at time t as $W_t(i)$ and assume $N \in \mathbb{N}$ agents. We note in this model that wealth is assumed to be non-negative: $W_t(i) \geq 0$. Consider the following SDE for agent i:

$$\mathrm{d}W_t(i) = \mu W_t(i)\mathrm{d}t + \sigma W_t(i)\mathrm{d}B_t(i) + \sum_{j\neq i}\left(a_{ij}W_t(j) - a_{ji}W_t(i)\right)\mathrm{d}t \tag{2.30}$$

where $B_t(i)$ is Brownian motion for agent i and a_{ij} is the proportion of wealth that agent i gains by exchange from agent j (and vice versa). Thus the summation term in (2.30) can be viewed as economic exchange across all agents in the system. This model defined by (2.30) was first introduced by Bouchaud and Mézard [23]. We note that (2.30) is presented with the Itô convention as written in [94] rather that the Langevin equation form seen in [23].

To make matters simpler we assume all agents exchange with the same fixed rate $a_{ij} = \tau/N$ for some constant $\tau \in \mathbb{R}$. We define the ensemble mean $\overline{W_t} := (1/N)\sum_{i=1}^{N} W_t(i)$ and the proportional wealth of agent i as $w_t(i) := W_t(i)/\overline{W_t}$.

[6] The stationary solution may only exist under constraints such as a reflecting boundary condition, see Section 5.2.1 [65] for an overview.

Substituting the fixed rates into (2.30) we find

$$\mathrm{d}W_t(i) = \mu W_t(i)\mathrm{d}t + \sigma W_t(i)\mathrm{d}B_t(i) - \tau(W_t(i) - \overline{W_t})\mathrm{d}t\,. \tag{2.31}$$

We note that (2.31) is also referred to as reallocating geometric Brownian motion (RGBM) [19] and can be seen to be geometric Brownian motion (GBM), see next Section 2.10, with a mean reverting term determined by the value and particularly the sign of the reallocation rate parameter τ. For $\tau = 0$ (2.31) is GBM, for $\tau > 0$ we have that growth in wealth reverts to the mean and that W_t becomes stationary with an inverse Gamma distribution, see (2.34), and for $\tau < 0$ we have mean repulsion so that agents wealth is pushed either more positive or negative in an exponential manner and no stationary solution exits [19]. Fitting (2.31) to US top wealth share data from 1913 to 2014 gives a fluctuating τ that is on average positive until around 1980 after which it is on average negative [19] coinciding with the decrease then increase in inequality as seen in Figure 1.11.

We now find the stationary distribution of (2.31) with $\tau > 0$. Summing these equations (2.31) over the N agents and dividing by N we obtain

$$\mathrm{d}\overline{W_t} \simeq \mu\overline{W_t}\mathrm{d}t \tag{2.32}$$

with solution $\overline{W_t} \simeq W_0\exp(\mu t)$. Now

$$\begin{aligned}
\mathrm{d}w_t(i) &= \frac{1}{\overline{W_t}}(\mathrm{d}W_t(i) - w_t(i)\mathrm{d}\overline{W_t}) \qquad \text{by quotient rule and cancellation of } \mathrm{d}t \\
&\simeq \frac{1}{\overline{W_t}}\left(\mu W_t(i)\mathrm{d}t + \sigma W_t(i)\mathrm{d}B_t(i) - \tau(W_t(i) - \overline{W_t})\mathrm{d}t - w_t(i)\mu\overline{W_t}\mathrm{d}t\right) \\
&= -\tau(w_t - 1)\mathrm{d}t + \sigma w_t\mathrm{d}B_t
\end{aligned}$$

where the second line above is by substitution of (2.31) and (2.32). For ease of notation we drop the i and so we have the following SDE equation to solve for:

$$\mathrm{d}w_t \simeq -\tau(w_t - 1)\mathrm{d}t + \sigma w_t\mathrm{d}B_t\,. \tag{2.33}$$

Note that (2.33) is no longer μ dependent. Using (2.29) to find a stationary solution of (2.33) we have $a(s) = -\tau(s - 1)$ and $b(s) = \sigma s$ and assuming $W_t \geq w_0$ for all t

we have

$$\begin{aligned}\int_{w_0}^{w} \frac{2a(s)}{b(s)^2}\mathrm{d}s &= \int_{w_0}^{w} \frac{2\tau(1-s)}{\sigma^2 s^2}\mathrm{d}s \\ &= -\frac{2\tau}{\sigma^2}\left[\frac{1}{s} + \log s\right]_{w_0}^{w} \\ &= \frac{2\tau}{\sigma^2}\left(\frac{1}{w_0} - \frac{1}{w}\right) + \log\left(\frac{w_0}{w}\right)^{2\tau/\sigma^2}\end{aligned}$$

and so by (2.28) the stationary density to (2.33) is

$$p(w) = \frac{C}{w^{2(1+\tau/\sigma^2)}} \exp\left(-\frac{2\tau}{\sigma^2 w}\right), \quad w > 0 \tag{2.34}$$

which is an inverse Gamma distribution with $C = \dfrac{(\alpha-1)^\alpha}{\Gamma(\alpha)}$ where $\alpha = 1 + 2\tau/\sigma^2$ and $\Gamma(\alpha)$ is the Gamma function. See [94] for this result which was first found for the Langevin equivalent in [23]. We see that for large w this density (2.34) tends to a power law and thus the stationary distribution of (2.31) has a power law tail with exponent $\beta = \alpha$. We see from (2.34) higher inequality either with smaller τ meaning a smaller rate of agent interaction or higher σ^2 corresponding to a higher rate of diffusion.

An extension of the Bouchaud model that produces a generalisation of (2.34) with another free parameter and incorporates negative wealth is seen in [67].

2.10 Analysis of Non-linear GBM

In this section we will analyse the following SDE

$$\mathrm{d}W_t = \mu W_t^{\gamma_1}\mathrm{d}t + \sigma W_t^{\gamma_2}\mathrm{d}B_t \tag{2.35}$$

with parameters $\gamma_1, \gamma_2, \mu, \sigma \in \mathbb{R}$. We shall call (2.35) non-linear geometric Brownian motion (non-linear GBM) if γ_1 or γ_2 are not both equal to 0 or 1. If $\gamma_1 = \gamma_2 = 1$ we have the well-known GBM with solution

$$W_t = W_0 \exp\left(\left(\mu - \frac{\sigma^2}{2}\right)t + \sigma B_t\right) \tag{2.36}$$

and thus W_t has a lognormal distribution. GBM was assumed to model stock prices in the derivation of the famous Black-Scholes equation for call options [21]. However using GBM to model stock prices implies normal returns which has been found to be problematic due to evidence that stock returns can be heavy tailed [105]. Thus to create heavier returns one may consider (2.35) but with different γ_1 or γ_2 so that the growth and or random noise changes non-linearly.

With $\gamma_1 = 1$ and $\gamma_2 = \gamma \in \mathbb{R}$ (2.35) is known as the constant elasticity in returns variance (CEV) used as an alternative to GBM to model stock prices [38]. It assumes that the diffusion term increases non-linearly with the stock price W_t. Exact transition density solutions to the CEV model have been found involving modified Bessel functions [96]. They were first studied for $0 < \gamma < 1$ [38] but have since been extended to $\gamma > 1$ [51] (where $\gamma = 1$ is exactly GBM). When fitting to options data it has been observed that γ may oscillate above one but in the main is below one [51]. For $\gamma > 1$ it has been found that the stationary solution is a power law with exponent $\beta = 2\gamma - 1$ [70].

Relatively recently (2.35) has been analysed more generally in [70] and they found that if $\gamma_2 > 1$ and $\gamma_1 < 2\gamma_2 - 1$ then there is a stationary power law with exponent $\beta = 2\gamma_2 - 1$ (see (2.43) below). Note that the exponent only depends on the exponent in the diffusive term.

It is interesting to observe when (2.35) reaches explosion ($W_t = \infty$ for $t < \infty$). As noted in [70] the pure growth process reaches explosion for $\gamma_1 = \gamma > 1$. Thus with $\mu \neq 0$, $\sigma = 0$ then the pure growth ODE

$$\mathrm{d}W_t = \mu W_t^{\gamma} \mathrm{d}t \tag{2.37}$$

has solution

$$W_t = \left(W_0^{-(\gamma-1)} - \mu(\gamma-1)t\right)^{-\frac{1}{\gamma-1}}$$

which explodes in finite time for $\gamma > 1$ at

$$t = \frac{1}{W_0^{\gamma-1}(\gamma-1)\mu} \, .$$

Now considering $\mu = 0$, $\sigma \neq 0$ we have the purely diffusive process

$$\mathrm{d}W_t = \sigma W_t^{\gamma} \mathrm{d}B_t \, . \tag{2.38}$$

The solution to this equation does not explode as in the pure growth process but does exhibit ‘bursting’ behaviour where the process can reach arbitrarily large values [70]. It is indicated in [70] that in general (2.35) is explosive when $\gamma_1 > 2\gamma_2 - 1 > 1$ however it would be interesting to confirm this using Feller’s test of explosions [84]

A particular case of the model (2.35) has been studied as an SDE approximation to a GARCH process of price volatility and produces a stationary distribution with power laws under reflective boundary conditions [92]. As is done in [94, 92, 70] and illustrated in the previous Section 2.9 we shall find the stationary solution of (2.35) using the stationary density formula from the Fokker-Planck equation (2.29). Following the notation from (2.29) we have $a(s) = \mu s^{\gamma_1}$ and $b(s) = \sigma s^{\gamma_2}$ and so

$$\int_{w_0}^{w} \frac{2a(s)}{b(s)^2} \mathrm{d}s = \frac{2\mu}{\sigma^2} \int_{w_0}^{w} s^{\gamma_1 - 2\gamma_2} \mathrm{d}s$$

$$= \begin{cases} \frac{2\mu}{\sigma^2} [\log s]_{w_0}^{w} & \text{if } \gamma_1 - 2\gamma_2 = -1 \quad (2.39) \\ \frac{2\mu}{\sigma^2} \left[\frac{s^{\gamma_1 - 2\gamma_2 + 1}}{\gamma_1 - 2\gamma_2 + 1} \right]_{w_0}^{w} & \text{if } \gamma_1 - 2\gamma_2 \neq -1 \,. \quad (2.40) \end{cases}$$

Let us first analyse the case $\gamma_1 - 2\gamma_2 = -1$ (2.39) which was studied in [92]. Note that GBM $\gamma_1 = \gamma_2 = 1$ (but with a reflecting barrier) satisfies this case. If we choose $w_0 > 0$ as a reflecting barrier [94] then

$$\int_{w_0}^{w} \frac{2a(s)}{b(s)^2} \mathrm{d}s = \frac{2\mu}{\sigma^2} (\log w - \log w_0) \quad \text{if } \gamma_1 - 2\gamma_2 = -1$$

and thus by substituting into (2.29) we have that the stationary distribution is a Pareto distribution with density:

$$p(w) = \frac{C}{w^{\beta+1}} \quad \text{if } \gamma_1 - 2\gamma_2 = -1 \tag{2.41}$$

with $\beta = 2(\gamma_2 - \frac{\mu}{\sigma^2}) - 1$ and $C = \beta w_0^{\beta}$, see [94] for this result. Note for the Pareto distribution to be well defined we assume $\beta > 0$. For $\gamma_1 = \gamma_2 = 1$ we see that $\beta > 0$ translates to $\mu < \sigma^2/2$ and thus GBM (2.36) tends to 0 as $t \to \infty$ which is prevented by the boundary condition. From the power law exponent β we see that fixing γ_2 there is greater inequality either with higher μ or smaller σ^2. We also see counter-intuitively greater inequality with smaller γ_2. However from preliminary simulations the presence of the 'bursting' phenomena discussed above is more likely for higher γ_2.

Now consider the other case where $\gamma_1 - 2\gamma_2 \neq 1$ (2.40) and in particular when

$$\gamma_1 < 2\gamma_2 - 1 \quad \text{and} \quad \mu, \sigma > 0 \,. \tag{2.42}$$

Then, again assuming the reflecting lower barrier $w_0 > 0$:

$$\int_{w_0}^{w} \frac{2a(s)}{b(s)^2} \mathrm{d}s = \frac{2\mu}{\sigma^2(\gamma_1 - 2\gamma_2 + 1)}(w^{\gamma_1 - 2\gamma_2 + 1} - w_0^{\gamma_1 - 2\gamma_2 + 1}) \quad \text{if } \gamma_1 < 2\gamma_2 - 1$$

and thus by substituting into (2.29) we have that the stationary distribution is of the following form

$$p(w) = C w^{\tau} \exp(-a w^{\alpha}) \quad \text{if } \gamma_1 < 2\gamma_2 - 1 \tag{2.43}$$

where $\tau = -2\gamma_2 < 0$, $a = \dfrac{-2\mu}{\sigma^2(\gamma_1 - 2\gamma_2 + 1)} > 0$ and $\alpha = \gamma_1 - 2\gamma_2 + 1 < 0$. We see that if $\gamma_1 > 2\gamma_2 - 1$ then $\alpha > 0$ and $a < 0$ leading to no stationary solution. With $\gamma_1 > 1$ the process is always finite and greater than 0 [70]. Then integrating over $(0, \infty)$ (2.43) gives result 4 in [70] with the constant C calculated below for $w_0 \to 0$. For $p(w)$ to be a density, which we see is a generalisation of the inverse Gamma distribution seen in the last section, we must have

$$\begin{aligned}
1 &= C \int_{w_0}^{\infty} w^{\tau} \exp(-a w^{\alpha}) \mathrm{d}w \\
&= C \alpha^{-1} a^{-\frac{\tau+1}{\alpha}} \int_{a w_0^{\alpha}}^{0} z^{\frac{\tau+1}{\alpha} - 1} \exp(-z) \mathrm{d}z, \quad \text{where } z = a w^{\alpha} \quad \text{with } a > 0,\, \alpha < 0 \\
&= -C \alpha^{-1} a^{-\frac{\tau+1}{\alpha}} \int_{0}^{a w_0^{\alpha}} z^{\frac{\tau+1}{\alpha} - 1} \exp(-z) \mathrm{d}z \quad \Rightarrow \\
C &= \frac{-\alpha}{a^{-\frac{\tau+1}{\alpha}} \gamma\left(\frac{\tau+1}{\alpha}, a w_0^{\alpha}\right)} \xrightarrow{w_0 \to 0} \frac{-\alpha}{a^{-\frac{\tau+1}{\alpha}} \Gamma\left(\frac{\tau+1}{\alpha}\right)}
\end{aligned}$$

where $\gamma(x, y) := \int_0^y s^{x-1} e^{-s} \mathrm{d}s$ is the lower incomplete Gamma function and $\Gamma(x)$ is the regular Gamma function. We note that for large w the density (2.43) approaches a power law density (1.10). This indicates the stationary solution of (2.35) with condition (2.42) has a power law tail. Again we see higher inequality for lower γ_2 but note again higher γ_2 indicates a greater level of 'bursting' [70].

2.11 SDE with κ-generalised Solution

We now show that the κ-generalised distribution which is a good model for positive wealth, see Section 1.4.4, is the stationary distribution of the following SDE:

$$\mathrm{d}W_t = -\frac{1}{2} \mathrm{d}t + \left(\frac{\sqrt{1 + \kappa^2 \left(W_t / \beta\right)^{2\alpha}}}{\alpha / \beta \left(W_t / \beta\right)^{\alpha - 1}} \right)^{1/2} \mathrm{d}B_t \tag{2.44}$$

with $\alpha, \beta > 0$ (noting here β is not the power law exponent, instead we shall use $\beta_{\kappa\text{-gen}}$ as in Section 1.4.4) and $\kappa \in (0,1)$ with a reflecting boundary condition such that $W_t \geq 0$ for all t.

We see that the diffusion coefficient scales like $\simeq \sqrt{\kappa/\alpha}\sqrt{W_t}$ for large W_t so let us first analyse the simpler SDE

$$\mathrm{d}W_t \simeq -\frac{1}{2}\mathrm{d}t + \sqrt{\kappa/\alpha}\sqrt{W_t}\mathrm{d}B_t\,.$$

We see this SDE is a special asymptotic case of the non-linear GBM (2.35) with $\mu = -1/2$, $\gamma_1 = 0$, $\sigma = \sqrt{\kappa/\alpha}$ and $\gamma_2 = 1/2$. Now as $\gamma_1 - 2\gamma_2 = -1$ we have the density solution scales as (2.41)

$$p(w) \simeq \frac{C}{w^{\beta_{\kappa\text{-gen}}+1}}$$

with $\beta_{\kappa\text{-gen}} = 2(\gamma_2 - \mu/\sigma^2) - 1 = \alpha/\kappa$ which agrees with the κ-gen exponent (1.21). This means the distribution is more unequal for smaller α/κ or equivalently greater κ/α.

Going back to the original SDE (2.44)we again use (2.29) to find the stationary solution. We have

$$a(s) = -1/2 \quad \text{and} \quad b(s) = \left(\frac{\sqrt{1+\kappa^2\,(s/\beta)^{2\alpha}}}{\alpha/\beta\,(s/\beta)^{\alpha-1}}\right)^{1/2}.$$

Thus

$$\int_{w_0}^{w} \frac{2a(s)}{b(s)^2}\mathrm{d}s = \int_{w_0}^{w} -\frac{\alpha/\beta\,(s/\beta)^{\alpha-1}}{\sqrt{1+\kappa^2\,(s/\beta)^{2\alpha}}}\mathrm{d}s\,. \tag{2.45}$$

As seen in (1.18) in Section 1.4.4 we have the generalised exponential function

$$\exp_\kappa(s) := (\sqrt{1+\kappa^2 s^2} + \kappa s)^{1/\kappa}\,.$$

Setting $f(s) = \exp_\kappa(g(s))$ where $g(s)$ is another differentiable function, it can be shown using the chain rule and the property $\frac{\mathrm{d}}{\mathrm{d}s}\exp_\kappa(s) = \frac{\exp_\kappa(s)}{\sqrt{1+\kappa^2 s^2}}$ [33] that $f'(s) = \frac{g'(s)}{\sqrt{1+\kappa^2 g(s)^2}}f(s)$. Thus we have

$$\int_{w_0}^{w} \frac{g'(s)}{\sqrt{1+\kappa^2 g(s)^2}}\mathrm{d}s = \int_{w_0}^{w} \frac{f'(s)}{f(s)}\mathrm{d}s = [\log f(s)]_{w_0}^{w} = \log\exp_\kappa(g(w)) - \log\exp_\kappa(g(w_0))$$

and setting $g(s) = -(s/\beta)^\alpha$ we have by substitution

$$\int_{w_0}^{w} -\frac{\alpha/\beta\,(s/\beta)^{\alpha-1}}{\sqrt{1+\kappa^2\,(s/\beta)^{2\alpha}}}\mathrm{d}s = \log(\exp_\kappa(-(w/\beta)^\alpha)) - \log(\exp_\kappa(-(w_0/\beta)^\alpha))\,.$$

Therefore by (2.29) the stationary density solution to (2.44) is

$$p(w) = \frac{\alpha}{\beta}\left(\frac{w}{\beta}\right)^{\alpha-1}\frac{\exp_\kappa\left(-\left(w/\beta\right)^\alpha\right)}{\sqrt{1+\kappa^2\left(w/\beta\right)^{2\alpha}}}\,, \qquad w \geq 0 \tag{2.46}$$

which is the probability density of the κ-generalised distribution. We note that (2.44) is only one of many possibilities that gives the κ-generalised density as a solution. Any drift and diffusion term satisfying (2.45) would be suitable and there was no particular reason for the choice in (2.44).

2.12 Discussion

In this chapter we analysed a variety of agent based models from the literature along with slight extensions that produce heavy tails and power laws under the general themes of hierarchy, exchange, preferential attachment and multiplicative processes. We presented models that are both stationary and non-stationary and emphasise that non-stationary models may be more appropriate as wealth is generally a growing quantity.

The hierarchy models show how wealth can be arranged in space whereas the remainder of the models are concerned with how wealth changes amongst agents through time. In the exchange models we model only a fixed quantity 'money', a subset of wealth, that is repeatedly exchanged amongst agents. For the multiplicative models we mostly forget exchange and let wealth grow over time. In the balls in bins process with feedback, units of wealth enter the system and are received by agents with probability proportional to their wealth. In the multiplicative process models the wealth of agents grow through time by repeated multiplication of random factors. The balls in bins process depends on every agent at each iteration whereas the majority of the multiplicative process models are at the opposite extreme with agents iterating independently.

Although we fit the non-linear Kesten process to the wealth data in Chapter 4, future work would be to fit more of the models to wealth. Comparison of the models, for instance by analysing the returns, could show if there is a link between them.

Chapter 3

Balls in Bins Process with Feedback

3.1 Introduction

In this chapter we shall analyse the balls in bins process with feedback, which we shall refer to as the **feedback model** and was summarised in Section 2.4. In particular we will be interested in the resulting probability distribution using the master equation approach. We are particularly interested in the formation of power laws and discuss in more detail the monopoly result as well as an approximation of the expectation of wealth given the initial condition.

We can think of the feedback model as follows: at each time iteration a unit of wealth enters the economy and goes to a particular agent with some probability based on the agent's existing wealth, w, proportional to w^γ. The strength of the feedback is dominated by a parameter γ where above the critical value $\gamma_c = 1$ the model produces power laws with exponent $\beta = \gamma - 1$ and monopoly in the wealth distribution.

For overviews of the feedback model see [113, 114, 49, 149]. We note there are extensive theoretical results in the literature some of which were summarised in Section 2.4. Many results relate to the case of $N = 2$ agents. Our main aim in this chapter is to focus on the general $N \in \mathbb{N}$ case via the master equation which we believe has not been done before. We shall first focus on the feedback function of $f(w) = w^\gamma$ and then consider the more general fitness feedback function $f(w) = \alpha w^\gamma$ where $\alpha > 0$ is the fitness. We note that we do not fit the model to wealth data but this has been done for US wealth in [142].

3.2 Model

We shall now set up the model. We denote the wealth of $N \in \mathbb{N}$ agents at iteration $n = 0, 1, 2, \ldots$ by the set

$$\mathbf{I}_n = \{I_n(1), I_n(2), \ldots, I_n(N)\}$$

where $I_n(j)$ is the wealth of agent j at discrete time n. We use I for wealth at discrete time to differentiate from the continuous time process in the following sections. For this model we consider only cases of positive wealth of a non-zero integer value: $I_n(i) \in \mathbb{N}$ for all $i \in \{1, 2, \ldots, N\}$ and $n \in \{0, 1, 2, \ldots\}$.

At each iteration n, a wealth packet, $\omega_p \in \mathbb{N}$, is given to an agent j using the following update rule[1]:

$$\mathbb{P}(\mathbf{I}_{n+1} = \mathbf{I}_n + \omega_p \mathbf{E}_j | \mathbf{I}_n) = \frac{I_n(j)^\gamma}{\sum\limits_{i=1}^{N} I_n(i)^\gamma} \tag{3.1}$$

for real $\gamma \geq 0$ where $\mathbf{E}_j$ is the basis vector of N elements with zeros everywhere but the j^{th} place which is a one. Thus (3.1) is the probability that an agent j gains wealth ω_p and we see from the memoryless property of (3.1) that $\mathbf{I}_n$ is a Markov chain on the state space $S = \mathbb{N}^N$. For $\gamma > 0$ the update rule (3.1) gives a wealth advantage as the higher an agent's wealth the more likely it is to grow further in wealth. The case $\gamma = 1$ is the classical Pólya urn model of which a special case we analyse in Section 3.5. We focus primarily in this chapter on the most applicable case to wealth of $\gamma > 1$ which produces monopoly and power laws however theory is also known for the $\gamma < 1$ case where each agent in the limit gains the same fraction of wealth. Both the cases of $\gamma = 1$ and $\gamma < 1$ are discussed in [90, 49]. As mentioned in the introduction, we shall refer to the model for $\mathbf{I}_n$ defined by the update rule (3.1) as the **feedback model**. We note in this model agents can only gain wealth and the overall wealth in the system is growing at a fixed rate of ω_p and is an example of a pure birth process [85].

Note the update rule is invariant under non-zero multiplication:

$$\frac{(\alpha I_n(j))^\gamma}{\sum\limits_{i=1}^{N} (\alpha I_n(i))^\gamma} = \frac{I_n(j)^\gamma}{\sum\limits_{i=1}^{N} I_n(j)^\gamma}$$

where $\alpha \in \mathbb{N}$ and so the probability of a given agent gaining wealth is the same

[1]Feedback function $f(x) = x^\gamma$ in (2.8).

under non-zero scalar multiplication of the system $\mathbf{I}_n$. For the remainder of the chapter we set the wealth packet as one: $\omega_p = 1$.

3.3 Markov Process and Master Equation

We shall now relate the discrete time feedback model to a continuous time jump chain with independent agents, see Chapter 5 of [144], Chapter 7 [65] and Section 6.9 of [139] for background theory in this section.

Let

$$\mathbf{W}_t = \{W_t(1), W_t(2), \ldots, W_t(N)\}$$

be N agents at continuous time $t \in \mathbb{R}_{\geq 0}$ where $W_t(j) \in \mathbb{N}$ is the wealth of agent j at time t. For an agent $W_t(j) := W_t$ we define the transition probabilities as

$$p_t(\omega, \nu) := \mathbb{P}(W_t = \nu | W_0 = \omega),$$

the transition rates

$$g(\omega, \nu) := \frac{\mathrm{d}p_t(\omega, \nu)}{\mathrm{d}t}\Big|_{t=0}$$

and let $p_t(\omega) := \mathbb{P}(W_t = \omega)$ where ν, $\omega \in \mathbb{N}$. We define the rates to be

$$g(\omega, \nu) = \begin{cases} \omega^\gamma & \text{if } \nu = \omega + 1, \\ -\omega^\gamma & \text{if } \nu = \omega, \\ 0 & \text{otherwise.} \end{cases}$$

The **master equation** for the feedback model outlined in Section 3.2 for a single agent is

$$\begin{aligned} \frac{d}{dt} p_t(\omega) &= \sum_{\nu \neq \omega} (p_t(\nu) g(\nu, \omega) - p_t(\omega) g(\omega, \nu)) \\ &= p_t(\omega - 1)(\omega - 1)^\gamma - p_t(\omega)\omega^\gamma. \end{aligned} \tag{3.2}$$

The rates indicate the agents jump to a new wealth value of $\omega_p = 1$ above their current wealth. The intervals of time that the jumps happen are called the holding times. The **holding time** (or waiting time) $H_t(\omega)$, is defined as the time until an agent jumps to another value (when currently at value ω at time τ), and is exponentially distributed with mean $1/|g(\omega, \omega)| = \dfrac{1}{\omega^\gamma}$ i.e.

$$H_t(\omega) := \inf\{t > 0 : W_{\tau+t} = \omega + 1 \ : W_\tau = \omega\} \sim \mathrm{Exp}(\omega^\gamma). \tag{3.3}$$

Jump times are then defined as the cumulative sums of the holding times and are the continuous times when an agent jumps in wealth. The process W_t is known as a continuous time jump chain. As the holding times are independently distributed exponential RVs the agents run independent of each other and never jump at the same time. It can be shown that the distribution of W_t at the jump times is the same as the distribution of the balls in bins process with (3.1) by **exponential embedding** [113].[2] In mathematical terms suppose t_n , $n \in \mathbb{N}$, are the jump times then

$$\mathbf{I}_n = \mathbf{W}_{t_n}$$

and

$$(\mathbf{W}_t)_{t \in \mathbb{R}_{\geq 0}}$$

is a Markov process on the state space $S = \mathbb{N}^N$.

3.4 Explosion Time and Monopoly

For a particular agent j at time t with wealth $W_t := W_t(j)$, we find using the master equation (3.2) a relationship for the time differential of the expectation. Thus

$$\begin{aligned} \frac{d}{dt}\sum_{\omega=1}^{\infty} \omega p_t(\omega) &= \sum_{\omega=1}^{\infty} \omega \frac{d}{dt} p_t(\omega) \\ &= \sum_{\omega=1}^{\infty} \omega p_t(\omega-1)(\omega-1)^{\gamma} - \sum_{\omega=1}^{\infty} p_t(\omega)\omega^{\gamma} \\ &= \sum_{k=0}^{\infty} (k+1) p_t(k) k^{\gamma} - \sum_{\omega=1}^{\infty} p_t(\omega)\omega^{\gamma}, \qquad\qquad (k = \omega - 1) \\ &= \sum_{\omega=1}^{\infty} p_t(\omega)\omega^{\gamma}\,. \end{aligned}$$

Where we note that the interchange of the differentiation in the sum in the first equality above holds if $\sum_{\omega=1}^{\infty} \omega \frac{d}{dt} p_t(\omega)$ converges uniformly for $t \in \mathbb{R}$, see Theorem 7.17 of [126]. Therefore

$$\frac{d}{dt}\mathbb{E}[W_t] = \mathbb{E}[W_t^{\gamma}]\,. \tag{3.4}$$

When $\gamma = 1$ we see equation (3.4) holds exactly with solution

$$\mathbb{E}[W_t | W_0 = \omega_0] = \omega_0 e^t\,. \tag{3.5}$$

[2] We note that one of the earliest references to this technique is [11].

For a random variable X and a sufficiently differentiable function f on X we can find an approximation of the first moment (expectation) $\mathbb{E}[f(X)]$ using Taylor expansion. The approximation is (see Chapter 4 of [16])

$$\mathbb{E}[f(X)] \approx f(\mathbb{E}[X]) + \frac{f''(\mathbb{E}[X])}{2}\mathrm{Var}(X)\,. \tag{3.6}$$

Taking our random variable as W_t and feedback function $f(W_t) = W_t^\gamma$, with $\gamma > 0$ we find the approximation to be

$$\mathbb{E}[W_t^\gamma] \approx \mathbb{E}[W_t]^\gamma + \frac{\gamma(\gamma-1)(\mathbb{E}[W_t])^{\gamma-2}}{2}\mathrm{Var}(W_t)\,. \tag{3.7}$$

Thus assuming the second term in (3.7) is small (which for high $\mathbb{E}[W_t]$ and $\mathrm{Var}(W_t)$ holds for $\gamma < 2$)

$$\frac{d}{dt}\mathbb{E}[W_t] \approx \mathbb{E}[W_t]^\gamma \tag{3.8}$$

is a good approximation to (3.4).

Solving the first order ODE (3.8) for $\gamma > 1$ we find[3]

$$\mathbb{E}[W_t|W_0 = \omega_0] \approx \left(\frac{1}{\omega_0^{\gamma-1}} - (\gamma-1)t\right)^{-\frac{1}{\gamma-1}}\,. \tag{3.9}$$

We note if every agent starts with ω_0 then (3.5) and (3.9) will give the same expected value for every agent and we will have no idea of the order of poorest to richest agents. However if every agent starts with a different ω_0 then the order from poorest to richest shall be predicted to be the same as the initial order across time. We see also from (3.9) that for $\gamma > 1$ we expect a non-linear dependence on initial conditions.

For $\gamma > 1$, (3.9) 'explodes' ($\mathbb{E}[W_t] = \infty$ for $t < \infty$) at the **explosion time**

$$t = \frac{1}{\omega_0^{\gamma-1}(\gamma-1)}\,. \tag{3.10}$$

Another way to define the explosion time of an agent is the sum of the holding times that an agent spends at each state

$$\omega = \omega_0, \omega_0 + 1, \ldots\,.$$

[3] We note we solve the equivalent ODE in Section 2.10.

Thus a RV of the explosion time T can be written

$$T := \sum_{\omega=\omega_0}^{\infty} H_t(\omega)\,. \tag{3.11}$$

Then

Proposition 2. *The explosion time T (3.11) of an agent for the feedback model has finite expectation (expected explosion time) and variance for $\gamma > 1$.*

Proof. Noting the holding times for an agent are exponentially distributed $\mathrm{Exp}(\omega^{\gamma})$ with mean $\dfrac{1}{\omega^{\gamma}}$ (3.3)

$$\begin{aligned}\mathbb{E}[T] &= \mathbb{E}\left[\sum_{\omega=w_0}^{\infty} H_t(\omega)\right]\\ &= \sum_{\omega=\omega_0}^{\infty} \mathbb{E}[H_t(\omega)]\\ &= \sum_{\omega=\omega_0}^{\infty} \frac{1}{\omega^{\gamma}} < \infty \qquad \text{for } \gamma > 1\,.\end{aligned}$$

Where linearity of expectation holds in the second equality above as $H_t(\omega) \geq 0$ for all ω and $\mathbb{E}[H_t(\omega)]$ are finite, see Section 5.6 of [139]. Using the fact that the holding times have variance $\dfrac{1}{\omega^{2\gamma}}$ we have

$$\begin{aligned}\mathrm{Var}(T) &= \mathrm{Var}\left(\sum_{\omega=\omega_0}^{\infty} H_t(\omega)\right)\\ &= \sum_{\omega=\omega_0}^{\infty} \mathrm{Var}(H_t(\omega)) \qquad \text{by independence of } H_{\omega}\\ &= \sum_{\omega=\omega_0}^{\infty} \frac{1}{\omega^{2\gamma}} < \infty \qquad \text{for } \gamma > \frac{1}{2}\,.\end{aligned}$$

Again the linearity of variance holds in the second equality above as $H_t(\omega) \geq 0$ for all ω and $\mathrm{Var}(H_t(\omega))$ are finite by extension of the result in Section 5.6 of [139].

■

Therefore for each agent j the Markov process $(W_t(j))_{t\in\mathbb{R}_{\geq 0}}$ for the feedback model with $\gamma > 1$ can be described as a pure birth process that is explosive.

We now come to the following monopoly result. This is a particular case of the more general result given in Theorem 1 in Chapter 2.4. We note the existing proofs of the following result, see Proposition 2 [90] and Theorem 1 [114].

Proposition 3 ([114])**.** *For $\gamma > 1$, an agent j in the discrete time feedback model will achieve* ***monopoly*** *(see Section 2.4 Definition 6).*

Proof. We have N independent agents with independent explosion times T_i, $i = 1, 2, \ldots, N$ where T is defined in (3.11). As each T_i is a continuous random variable with finite mean and non-zero variance (Proposition 2) we have that $T_i \neq T_j$ for $i \neq j$ almost surely.

Thus we must have an ordering of the explosion times T_i :

$$T_{(1)} < T_{(2)} < \cdots < T_{(N)} < \infty\,.$$

Hence, letting $T_j = T_{(1)}$ we have $W_{T_j}(j) = \infty$ and

$$\sum_{i \neq j} W_{T_j}(i)$$

is a finite random variable. Thus relating the continuous process back to the discrete time feedback model this must mean that after a certain number of iterations every new wealth packet goes to the monopoly agent j . ■

We note in the proof above that the jump times of the joint process $\mathbf{W}_t$ do not go beyond the smallest explosion time $T_{(1)}$ i.e if t_n is the n^{th} jump time then $\lim_{n\to\infty} t_n = T_{(1)}$.

We also note that monopoly in the discrete feedback model means the rank ordering of agents from poorest to richest will after a certain amount of time become fixed. The rate at which this occurs is open to future research.

3.5 Solutions to the Master Equation and the Power Law Relationship

Solution for $\gamma = 0$

For $\gamma = 0$ the master equation (3.2) is

$$\frac{\mathrm{d}}{\mathrm{d}t} p_t(\omega) = p_t(\omega - 1) - p_t(\omega)\,. \tag{3.12}$$

It can be shown by substituting back into (3.12) that the mass function solution is of a Poisson distribution

$$p_t(\omega) = \frac{t^{\omega-\omega_0}}{(\omega-\omega_0)!}e^{-t}\,, \qquad \omega = \omega_0,\, \omega_0+1, \dots\,.$$

Solution for $\gamma = 1$ and $\omega_0 = 1$

For $\gamma = 1$, this is a case of the classical Pólya urn and the master equation (3.2) is

$$\frac{\mathrm{d}}{\mathrm{d}t}p_t(\omega) = p_t(\omega-1)(\omega-1) - p_t(\omega)\omega\,. \tag{3.13}$$

With $\omega_0(i) = 1$ for all $i = 1, 2, \dots, N$ it can be shown by substituting back into (3.13) that the solution to (3.13) is

$$p_t(\omega) = (1-e^{-t})^{\omega-1}e^{-t}\,, \qquad \omega = 1, 2, \dots.$$

This is the probability mass function for a geometric distribution, the discrete analogue of the exponential distribution, see Chapter 1 of Feller [56].

Solution for $\gamma = 2$ and $\omega_0 = 1$

For $\gamma = 2$ the master equation (3.2) is

$$\frac{d}{dt}p_t(n) = p_t(\omega-1)(\omega-1)^2 - p_t(\omega)\omega^2. \tag{3.14}$$

Assume $\omega_0(i) = 1$ for all $i = 1, 2, \dots, N$ and so

$$p_0(\omega) = \begin{cases} 1 & \text{if } \omega = 1 \\ 0 & \text{otherwise.} \end{cases} \tag{3.15}$$

For $\omega = 1$ the master equation (3.14) is

$$\frac{d}{dt}p_t(1) = -p_t(1)\,.$$

which has solution with initial conditions (3.15)

$$p_t(1) = e^{-t}\,. \tag{3.16}$$

For $\omega = 2$ the master equation (3.14) after substituting (3.16) becomes

$$\frac{d}{dt}p_t(2) + 4p_t(2) = e^{-t}$$

which has solution with initial conditions (3.15)

$$p_t(2) = \frac{1}{3}e^{-t} - \frac{1}{3}e^{-4t}.$$

Doing the same again for $\omega = 3$ we solve

$$\frac{d}{dt}p_t(3) + 9p_t(3) = \frac{4}{3}(e^{-t} - e^{-4t})$$

to find

$$p_t(3) = \frac{1}{6}e^{-t} - \frac{4}{15}e^{-4t} + \frac{1}{10}e^{-9t}\,.$$

From these solutions we make an ansatz for the general master equation (3.14) with initial conditions (3.15) as

$$p_t(\omega) = \sum_{j=1}^{\omega} a_{\omega,j}e^{-j^2 t} \tag{3.17}$$

with coefficients $a_{\omega,j} \in \mathbb{R}$. Substituting (3.17) into (3.14) and using the initial conditions (3.15) we can find the relationship between the coefficients for $n > 1$ given as

$$\begin{aligned} a_{\omega,j} &= \frac{(\omega-1)^2}{\omega^2 - j^2}a_{(\omega-1),j} \quad \text{ for } j = 1,2,\ldots,\omega-1 \quad \text{and} \\ a_{\omega,\omega} &= -\sum_{j=1}^{\omega-1} a_{\omega,j} \\ a_{1,1} &= 1\,. \end{aligned}$$

This shall be proved for general $\gamma \neq 0$ in Proposition 4 below.

General solution for $\gamma \neq 0$ with $\omega_0 = 1$

The method to reach the solution (3.17) for $\gamma = 2$ with initial condition $\omega_0(i) = 1$ for all $i = 1, 2, \ldots, N$ can be extended for general $\gamma \neq 0$:

Proposition 4. *The master equation* (3.2) *for the feedback model with* $\gamma \neq 0$ *and* *initial condition* $\omega_0 = 1$ *for all agents has the following solution for* $\omega > 1$

$$p_t(\omega) = \sum_{j=1}^{\omega} a_{\gamma,\omega,j} e^{-j^\gamma t} \tag{3.18}$$

with

$$\begin{aligned}
a_{\gamma,\omega,j} &= \frac{(\omega-1)^\gamma}{\omega^\gamma - j^\gamma} a_{\gamma,(\omega-1),j} \quad \textit{for } j = 1, 2, \ldots, \omega - 1\,, \\
a_{\gamma,\omega,\omega} &= -\sum_{j=1}^{\omega-1} a_{\gamma,\omega,j} \quad \textit{and} \\
a_{\gamma,1,1} &= 1\,.
\end{aligned}$$

Proof. Taking the derivative of (3.18) we find the left hand side of the master equation is

$$\frac{d}{dt} p_t(\omega) = \sum_{j=1}^{\omega} -j^\gamma a_{\gamma,\omega,j} e^{-j^\gamma t}\,. \tag{3.19}$$

The right hand side of the master equation is found by substitution of (3.18). We have for all $\omega = 2, 3, \ldots$

$$\begin{aligned}
p_t(\omega-1)(\omega-1)^\gamma - p_t(\omega)\omega^\gamma &= \sum_{j=1}^{\omega-1} a_{\gamma,\omega-1,j} e^{-j^\gamma t} (\omega-1)^\gamma - \sum_{j=1}^{\omega} a_{\gamma,\omega,j} e^{-j^\gamma t} \omega^\gamma \\
&= \sum_{j=1}^{\omega} ((\omega-1)^\gamma a_{\gamma,\omega-1,j} - \omega^\gamma a_{\gamma,\omega,j}) e^{-j^\gamma t} \qquad (3.20)
\end{aligned}$$

defining $a_{\gamma,\omega-1,\omega} := 0$. Setting the summands of (3.19) equal to (3.20) we have

$$\begin{aligned}
-j^\gamma a_{\gamma,\omega,j} &= (\omega-1)^\gamma a_{\gamma,\omega-1,j} - \omega^\gamma a_{\gamma,\omega,j} \Rightarrow \\
a_{\gamma,\omega,j} &= \frac{(\omega-1)^\gamma}{\omega^\gamma - j^\gamma} a_{\gamma,(\omega-1),j} \quad \text{for } j = 1, 2, \ldots, \omega - 1\,.
\end{aligned}$$

Now from the initial conditions (3.15) and substitution into (3.18) we have

$$\begin{aligned}
p_0(\omega) &= \sum_{j=1}^{\omega} a_{\gamma,\omega,j} = 0 \quad \forall\, \omega > 1 \Rightarrow \\
a_{\gamma,\omega,\omega} &= -\sum_{j=1}^{\omega-1} a_{\gamma,\omega,j}\,.
\end{aligned}$$

Finally again from the initial conditions (3.15) and substitution into (3.18) we have

$$p_0(1) = a_{\gamma,1,1} = 1\,.$$

■

We see for fixed ω that $p_t(\omega) \to 0$ as $t \to \infty$ which makes sense for $\gamma \leq 1$ as every agent will go to infinite wealth values. For $\gamma > 1$ this also makes sense as we have explosion at some finite time when viewing each agent individually. However as we noted above the joint process will not go beyond the smallest explosion time.

We plot $p_t(\omega)$ for an arbitrary value of $t = 3$ after numerically calculating (3.18) for $\gamma = 1,\ 1.3,\ 1.5,\ 2$ in Section 3.7 Figure 3.3. We note that the sum only converges numerically for appropriate t and ω that would be future research to analyse more exactly.

Power Law at $\gamma > 1$

We shall see from simulations, see Figure 3.2 Section 3.7, that in a certain region $[\omega_m, \omega_M] \subseteq \mathbb{N}$ a power law appears after a long enough time in the feedback model for $\gamma > 1$.

Thus for $\omega \in [\omega_m, \omega_M]$ assume

$$p_t(\omega) \to \frac{\alpha}{\omega^{\beta+1}} \qquad \text{as } t \to \infty \tag{3.21}$$

for some $\alpha, \beta > 0$. Then we have

$$\frac{\mathrm{d}p_t(\omega)}{\mathrm{d}t} \to 0 \qquad \text{as } t \to \infty\,. \tag{3.22}$$

Substituting (3.21) into the master equation (3.2) we have for
$\omega - 1,\, \omega \in [\omega_m, \omega_M]$

$$\begin{aligned}\frac{\mathrm{d}p_t(\omega)}{\mathrm{d}t} = p_t(\omega-1)(\omega-1)^\gamma - p_t(\omega)\omega^\gamma \to \frac{\alpha}{(\omega-1)^{\beta+1}}(\omega-1)^\gamma - \frac{\alpha}{\omega^{\beta+1}}\omega^\gamma \quad \text{as } t \to \infty \\ = \alpha\left((\omega-1)^{\gamma-(\beta+1)} - \omega^{\gamma-(\beta+1)}\right)\,.\end{aligned}$$

For (3.22) to hold with $\omega_M < \infty$ we require that

$$\beta \to \gamma - 1 \qquad \text{as } t \to \infty\,. \tag{3.23}$$

For $N = 2$ existing theory proves that (3.23) holds, see Corollary 2 [114]. We also see evidence that (3.23) holds for $N > 2$ agents - see Figure 3.2 for $N = 10$ agents. In Figure 3.2 we run the feedback model 2000 times and find the empirical tails of the aggregate of the runs excluding the monopoly agents.

3.6 Fitness

Now suppose to each agent j we attach a time and wealth independent 'fitness' $\eta_j > 0$ and adapt the update rule (3.1) so that an agent j gains a wealth packet ω_p at time k as follows:[4]

$$\mathbb{P}(\mathbf{I}_{n+1} = \mathbf{I}_n + \omega \mathbf{E}_j | \mathbf{I}_n) = \frac{\eta_j I_n(j)^\gamma}{\sum_{i=1}^{N} \eta_i I_n(i)^\gamma} \tag{3.24}$$

where we maintain the same notation and set up as in Section 3.1. We note again that this update rule is invariant under non-zero multiplication. We shall call the model defined by (3.24) the **fitness feedback model**. The fitness can be viewed as a fixed factor attached to each agent that leads to a greater chance of higher wealth. For instance this could be something like intelligence or productivity. This is a slightly more general model of the feedback model studied in the last section which occurs for equal fitness $\eta_i = 1$ for all agents i.

Applying exponential embedding so that the jump chain continuous time process $\mathbf{W}_t = \mathbf{I}_n$ at the jump times t_n and using exactly the same theory applied to the fitness feedback model as presented earlier in this chapter for the feedback model we find the following analogous results. The master equation for a single agent with fitness $\eta > 0$ is

$$\begin{aligned}\frac{d}{dt} p_t(\omega) &= p_t(\omega - 1)\eta(\omega - 1)^\gamma - p_t(\omega)\eta\omega^\gamma \\ &= \eta(p_t(\omega - 1)(\omega - 1)^\gamma - p_t(\omega)\omega^\gamma)\,. \end{aligned} \tag{3.25}$$

For $\gamma = 1$ we can derive the equation

$$\frac{d}{dt}\mathbb{E}[W_t] = \eta \mathbb{E}[W_t]$$

with solution

$$\mathbb{E}[W_t | W_0 = \omega_0] = \omega_0 e^{\eta t}\,.$$

[4]With feedback function $f(x) = \eta x^\gamma$ in (2.8).

An approximation for the time derivative of the expectation for $\gamma > 1$ is

$$\frac{d}{dt}\mathbb{E}[W_t] \approx \eta \mathbb{E}[W_t]^{\gamma} .$$

From this we find the approximate expected value of wealth at time t scaling with η compared to (3.9) for $\gamma > 1$:

$$\mathbb{E}[W_t | W_0 = \omega_0] \approx \left(\frac{1}{w_0^{\gamma-1}} - (\gamma - 1)\eta t \right)^{-\frac{1}{\gamma-1}} \tag{3.26}$$

which we see predicts higher expected wealth with both higher initial wealth and higher fitness. See Figure 3.4 for simulated W_t versus the predicted expectation (3.26).

It follows from (3.26) that monopoly occurs for $\gamma > 1$ with explosion time becoming proportionally smaller by $1/\eta$ compared to the feedback model (3.10):

$$t = \frac{1}{\omega_0^{\gamma-1}\eta(\gamma - 1)} .$$

As in Proposition 4 we find the following general solution for the master equation such that all agents have the same initial wealth $\omega_0 = 1$.

Proposition 5. *The master equation* (3.25) *for the fitness feedback model with* $\gamma \neq 0$ *and initial condition* $w_0 = 1$ *for all agents has the following solution for* $\omega > 1$

$$p_t(\omega) = \sum_{j=1}^{\omega} a_{\gamma,\omega,j} e^{-j^{\gamma}\eta t} \tag{3.27}$$

with

$$a_{\gamma,\omega,j} = \frac{(\omega - 1)^{\gamma}}{\omega^{\gamma} - j^{\gamma}} a_{\gamma,(\omega-1),j} \quad \textit{for } j = 1, 2, \ldots, \omega - 1 \quad \textit{and}$$

$$a_{\gamma,n,n} = -\sum_{j=1}^{\omega-1} a_{\gamma,n,j} \quad \textit{and}$$

$$a_{\gamma,1,1} = 1 .$$

We note that the only difference in the solution (3.27) compared to (3.18) is scaling the time t by the fitness η. This leads to the following particular solutions comparable to those seen in Section 3.5:

For $\gamma = 0$ the master equation is

$$\frac{d}{dt}p_t(\omega) = \eta(p_t(\omega - 1) - p_t(\omega))$$

with the following Poisson solution

$$p_t(\omega) = \frac{(\eta t)^{\omega - \omega_0}}{(\omega - \omega_0)!}e^{-\eta t}\,, \quad \omega \geq w_0$$

with mean ηt.

For $\gamma = 1$ the master equation (3.2) is

$$\frac{d}{dt}p_t(\omega) = \eta(p_t(\omega - 1)(\omega - 1) - p_t(\omega)\omega)\,.$$

With $\omega_0 = 1$ for all agents it can be shown the solution for $\omega \geq \omega_0 = 1$ is the following Geometric distribution

$$p_t(\omega) = (1 - e^{-\eta t})^{\omega - 1}e^{-\eta t}$$

with mean $e^{\eta t}$.

For $\gamma > 1$ we again hypothesise that the limiting distribution follows a power law for $\omega \in [\omega_m, \omega_M]$ and using the same substitution approach in the last section but with master equation (3.25) we see $p_t(\omega) = \alpha/\omega^{\beta+1}$ with $\beta \to \gamma - 1$ as $t \to \infty$.

3.7 Simulations

We note that using a binary tree search detailed in Appendix B can significantly speed up the run time of simulating the feedback models. The binary tree search is an alternative to the bisection and in parallel approach seen in [142].

We summarise the following points relating to the simulations:

S1. The model is dependent on initial conditions and the rank order of agents in terms of the magnitude of their wealth becomes fixed over time. We see evidence of this in Figure 3.4.

S2. We see the presence of monopoly for $\gamma > 1$ and power laws excluding the monopolistic richest agent. The higher the γ the more apparent this is as seen in Figures 3.1. We also see evidence of the relationship between the power law exponent, $\beta = \gamma - 1$, from Figure 3.2 and 3.3.

S3. We present simulations not fitted to real world data. This has been done in [142] however as they acknowledge it is hard to relate certain aspects of the model to the real world. Monopoly where almost all wealth belongs to a single household is of course an asymptotic concept, and for real world data only the transient behaviour of feedback models can be relevant. Therefore the feedback models can only be fitted over a finite time period.

Figure 3.1 show the tails of the feedback model for one run at iteration $n = 10^6, 10^7, 10^8, 10^9$ for $N = 1024$ agents[5], initial wealth of all agents the same at $\omega_0 = 1$, and $\gamma = 1,\ 1.1,\ 1.2,\ 1.3$. We see that for $\gamma = 1$ the simulation is well fitted by an exponential distribution in line with the theory in Section 3.5 and for $\gamma > 1$ the appearance of power law tails and monopoly also in line with theory in Sections 3.4 and 3.5. We can see the monopoly becomes more apparent for higher γ. It is interesting to note that for $\gamma > 1$ some agents never receive a wealth packet so stay at initial wealth $\omega_0 = 1$ whilst the monopoly agent receives close to the total number of wealth packets added.

The relationship of the power law exponent β and γ, $\beta = \gamma - 1$, seen in Section 3.5, is not apparent for one run of the feedback model as in Figure 3.1. However if we aggregate across runs (combine all values from each run) excluding the monopoly agents we obtain a more accurate distribution and do see evidence that $\beta \to \gamma - 1$. The example of $N = 10$ agents and aggregating over 2000 runs is seen in Figure 3.2. Checking the power law relationship for higher numbers of agents and different initial conditions and the speed of convergence through simulations is subject to future investigation. We note that higher γ seems to lead to faster convergence across runs when comparing $\gamma = 1.5$ and $\gamma = 2$ in Figure 3.2.

We numerically calculate the mass function $p_t(\omega)$ for the feedback model using the sum form (3.18) in Figure 3.3. We again see evidence for the exponential mass function for $\gamma = 1$ and the power law mass function $p_t(\omega) = \alpha/\omega^\gamma$ (meaning a power law exponent of $\beta = \gamma - 1$) for $\gamma > 1$. We note that we obtain good results from the sum (3.18) only for specific t and can only calculate the sum up to a certain ω before numerical divergence.

[5] We choose $N = 2^{10} = 1024$ agents as it is convenient for the binary search detailed in Appendix B

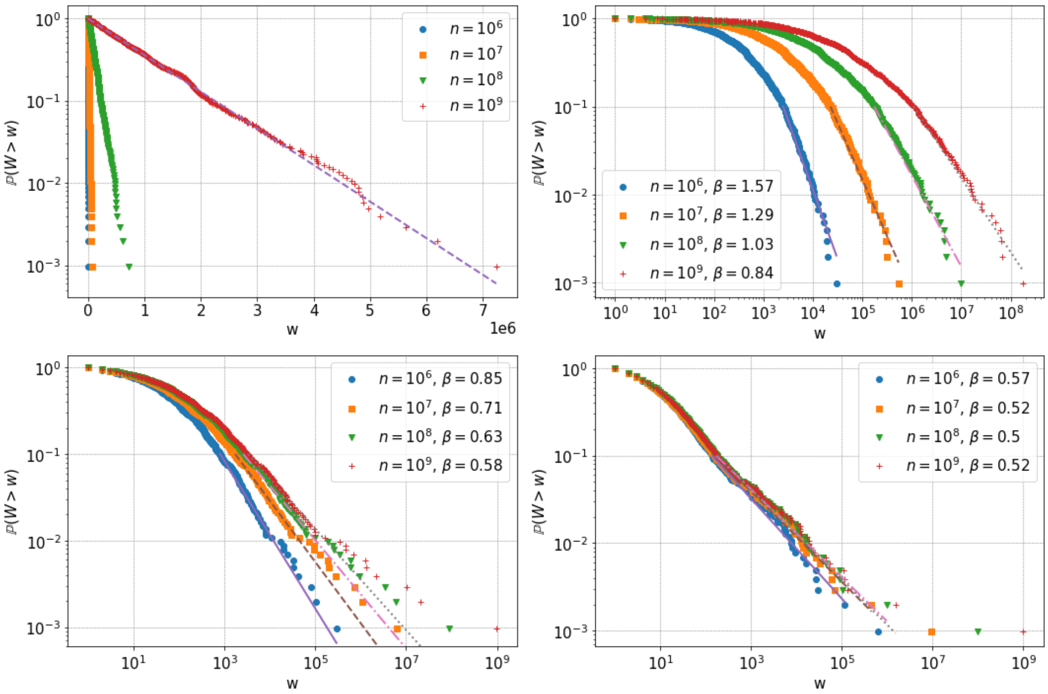

Figure 3.1: Tails of feedback model with initial wealth $\omega_0 = 1$ for each agent at iteration $n = 10^6, 10^7, 10^8, 10^9$ for $N = 2^{10} = 1024$ agents. Top left $\gamma = 1$, top right $\gamma = 1.1$, bottom left $\gamma = 1.2$ and bottom right $\gamma = 1.3$. The β values are MLE power law fits for the top 10% richest agents.

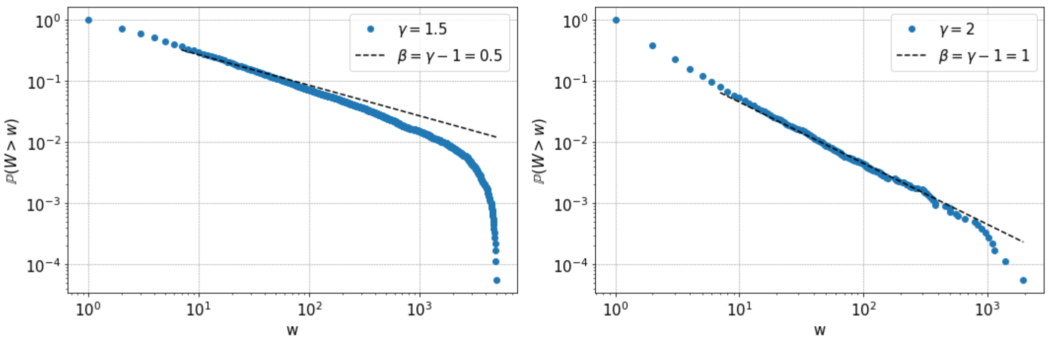

Figure 3.2: Tails of feedback model of aggregation of 2000 runs up to $n = 10^4$ iterations excluding monopoly agent with $N = 10$ agents, initial wealth $\omega_0 = 1$ with $\gamma = 1.5$, left and $\gamma = 2$, right. Power law tails plotted with exponents $\beta = \gamma - 1$.

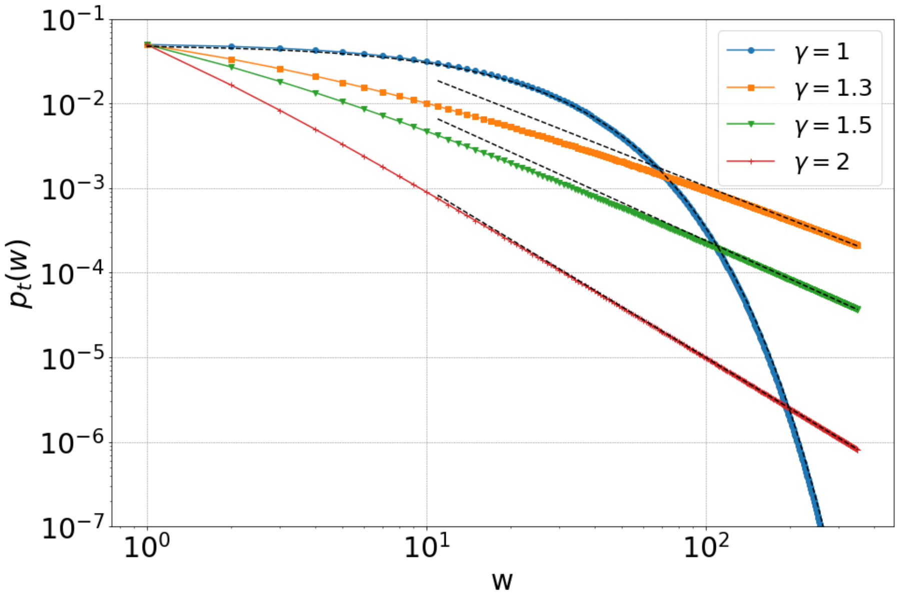

Figure 3.3: Calculation of mass function $p_t(\omega)$ for the feedback model with $\omega_0 = 1$ at time $t = 3$ using the sum form (3.18) for $\gamma = 1$, 1.3, 1.5, 2 up to $\omega = 350$. For $\gamma = 1$ we find an exponential distribution approximation fits well and for $\gamma > 1$ we see evidence of the power law $p_t(\omega) = \alpha/\omega^\gamma$ for $\omega \in [\omega_m, \omega_M]$, see black dashed lines.

Figure 3.4 shows that (3.26) can give a reasonable approximation for $\mathbb{E}[W_t|W_0 = w_0]$. Specifically Figure 3.4 shows the results for $\gamma = 1.1$, initial distribution $W_0 \sim \text{Bin}(10{\cdot}N, 1/N)$, arbitrary slightly skewed fitness $\eta \sim \text{Gamma}(9, 2)$, $N = 1024$ agents at iteration $n = 10^6$, 10^7, 10^8, 10^9 .

To approximate continuous time $t \in \mathbb{R}$ at discrete time iteration $n \in \{0, 1, 2, \ldots\}$ we numerically solve for t in the following

$$\sum_{i=1}^{N} I_n(i) = \sum_{i=1}^{N} \mathbb{E}[W_t(i)|W_0(i) = \omega_0(i)] \approx \sum_{i=1}^{N} \left(\frac{1}{\omega_0(i)^{\gamma-1}} - (\gamma - 1)\eta_i t \right)^{-\frac{1}{\gamma-1}}$$

where $\omega_0(i)$ and η_i is the initial wealth and the fitness respectively of agent i. In other words we solve for t where the total wealth of the system at iteration k is equal to the total wealth of the continuous time approximations (3.26) of the expected wealth of agents. We could do this approximation for the feedback model without fitness by using (3.9) instead of (3.26).

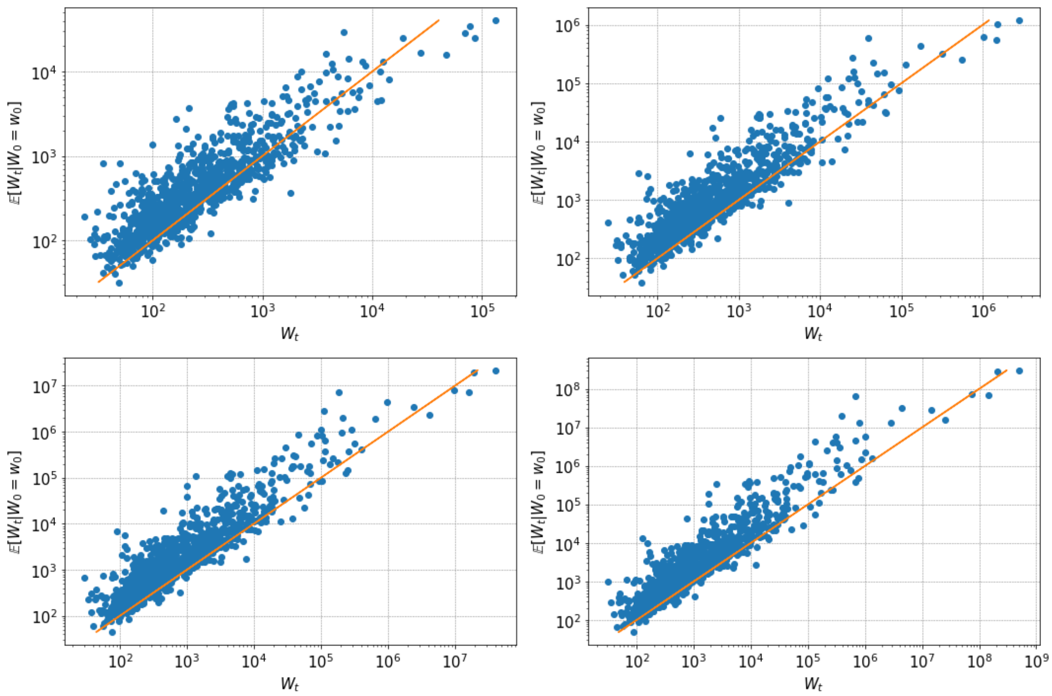

Figure 3.4: W_t versus $\mathbb{E}[W_t|W_0=\omega_0]$ (3.26) for fitness feedback model with fitness $\alpha \sim \text{Gamma}(9,2)$ and initial distribution $W_0 \sim \text{Bin}(10 \cdot N, 1/N)$, $\gamma = 1.1$ and $N = 2^{10} = 1024$ agents. Top left, top right, bottom left and bottom right corresponds to iteration $n = 10^6$, 10^7, 10^8, 10^9 . Lines correspond to $y = x$ for comparison.

3.8 Discussion

We have presented theoretical results of the balls in bins process with feedback also referred to as a non-linear generalised Pólya urn model for the general $N \in \mathbb{N}$ numbers of bins/agents case. The primary feedback function studied was W^γ with extension to more general ηW^γ with fitness η . We outlined the technique of exponential embedding to view the discrete time feedback model in continuous time. The master equation was used to find the explosion time and proof of monopoly for $\gamma > 1$, an approximation for the expectation of wealth given initial conditions, specific probability mass functions for all agents initially with one unit of wealth, $w_0 = 1$, probability mass functions for general $\gamma \neq 0$ in terms of summations and prediction of the power law exponent $\beta = \gamma - 1$ for $\gamma > 1$. We showed a few figures illustrating that the simulations back up the theory.

There are several aspects of the model that could be analysed in more detail. For instance exploring further via simulations the convergence of the power law exponent relationship $\beta = \gamma - 1$ for different numbers of agents. Related to this we could analyse more generally how fast the rank ordering of agents changes over time for different γ and how the model scales for different numbers of agents. One could also investigate further how to attach fitness to agents. What fitness means in terms of a quality like intelligence or productivity and how to find it's distribution would be inquiries for further study. Finally one could think of more complicated feedback functions and methods to break up the monopoly of the model such as a fluctuating γ or additional random factors.

Future work would include running the model with regards to UK data and comparing the fits to US data [142]. However as mentioned in [142] it is difficult to understand how the parameters of the model relate to the real world. A potentially more intuitive approach to modelling wealth is considered in the next chapter.

Chapter 4

Non-linear Kesten Process

4.1 Introduction

This chapter contains our work found in [60] and centres around an agent based model for positive wealth generalising the Kesten process [88], see Chapter 2, Section 2.6. The inspiration for the model is based on an empirical relationship between wealth and its rate of return. We find, from UK wealth data, evidence that the rate of return has a rough power law relationship meaning that agents with higher wealth are non-linearly more likely to have a higher return. We shall see from simulations evidence that a power law emerges in the model along with a crossover point leading to super-exponential growth and complete inequality. We find that this model, like the balls in bins process, leads to unrealistically high inequality over long times. However over shorter time periods the wealth biased returns could be a reason for the increasing inequality seen generally since the 1980s.

4.2 Model

We consider independent agents (representing households), whose wealth at discrete time $n \in \{0, 1, 2, \dots\}$ (representing years) is denoted by $W_n > 0$. The focus of this model is wealth growth rather than exchange, and we model the dynamics of positive wealth only, keeping track of bankruptcy events after which we reset the wealth value of the agent (see Section 4.5 for details). We assume that the wealth of an agent over the time period n to $n + 1$ changes via two mechanisms: **returns on existing wealth**, where $R_{n+1} \in \mathbb{R}$ denotes the corresponding rate of return (ROR), and **savings** $S_{n+1} \geq 0$, resulting for example from excess earnings which are independent of the current wealth of an agent (see Section 4.4.3 for details).

This leads to the recursion

$$W_{n+1} = W_n(1 + R_{n+1}) + S_{n+1} \quad \text{with initial condition} \quad W_0 > 0\,. \tag{4.1}$$

Here the RORs R_n and savings S_n are independent random variables. It is commonly accepted that RORs depend monotonically on wealth [54, 14, 50], and we assume the following power-law form,

$$R_{n+1} = \alpha_{n+1} W_n^{\gamma-1} \quad \text{for some } \gamma \geq 1\,, \tag{4.2}$$

where $\alpha_n \in \mathbb{R}$ are i.i.d. random variables from some fixed probability distribution, and with small probability can also take negative values. The very simple choice (4.2) is consistent with empirical data for the UK presented in Section 4.4.1. We are not claiming that this is the best or most detailed model for RORs, which have been observed in some cases to exhibit an intermediate plateau rather than a strict increase as a function of W_n (see e.g. Figure 2 in [50]). But our aim here is to capture the most essential features in a simple model that can also be analysed analytically, and it is of course possible for simulations to replace (4.2) by different functions. We find that a non-central t distribution (see Appendix C.5 for details) provides a good match with data for α_n, which is discussed in Section 4.4, Figure 4.3.

Substituting (4.2) in (4.1) gives the recursion

$$W_{n+1} = W_n + \alpha_{n+1} W_n^{\gamma} + S_{n+1}\,. \tag{4.3}$$

With $\gamma > 1$ we refer to (4.3) as a **non-linear Kesten process**. We now summarise theoretical results of (4.3) for different γ values.

$\gamma = 1$. See Chapter 2 Section 2.6 for a more in depth summary of the mathematical properties of this case. Here we have $R_n = \alpha_n$ and $W_{n+1} = (1 + \alpha_{n+1})W_n + S_{n+1}$. The stationary version of this linear model has been introduced and studied by Kesten [88], and the non-stationary asymptotic growth case is more recently discussed in [74]. It is easy to see that the asymptotic behaviour of W_n is dominated by the exponential $e^{n \log|1+\alpha_n|}$, and we present details on the analysis of both cases in Section 2.6. In the stationary case with $\mu := \mathbb{E}[\log|1 + R_n|] < 0$, the model is known to exhibit power-law tails in the limiting distribution, but for wealth dynamics the non-stationary case of asymptotic growth is most relevant, which occurs for $\mu > 0$. Following results in [74], the asymptotics, see Section 2.6, is given by a **log-normal distribution** such that to leading exponential order [1]

$$W_n \asymp W_0 \exp\left(\mu n + \sqrt{n\nu^2} Z\right) \quad \text{as } n \to \infty\,, \tag{4.4}$$

where $\nu^2 := \mathrm{Var}[\log|1+R_n|]$ and $Z \sim \mathcal{N}(0,1)$ is a standard Gaussian. The rigorous version of this result is subject to further reasonable and mild regularity assumptions on the distributions of parameters (see Theorem 2(i) in [74]), and the leading order behaviour is independent of the savings S_n. Since (4.3) is linear in W_n, the model also has a natural scale invariance for the units of wealth (see discussion in [23]), and the initial condition W_0 enters (4.4) as a simple multiplicative constant.

$\gamma > 1$**.** To our knowledge the non-linear model has not been studied before. Details are given in Section 4.3, where we find asymptotic super-exponential growth to leading order,[2]

$$W_n \asymp \left(W_0 e^D\right)^{\gamma^n} \quad \text{as } n \to \infty\,, \tag{4.5}$$

where D is given by a convergent series depending on the distribution of α_n and the initial behaviour of the process. Again, we focus on the non-stationary case with $W_0 e^D > 1$. In contrast to the linear case, we see that the asymptotics depend in a strong, non-linear way on the initial conditions and early dynamics of the process. Therefore there is no central limit theorem on the logarithmic scale that leads to (4.4), and we are not able to predict the asymptotic scaling distribution of W_n. But numerical results presented in Section 4.5 show that the model exhibits power-law tails with realistic shapes on relevant time scales.

For realistic initial conditions and parameters the dynamics follows initially an exponential growth regime, and super-exponential growth sets in when the dominant term in yearly gains in equation (4.3) changes from W_n to $\alpha_{n+1} W_n^\gamma$ (additive savings again do not influence the asymptotic behaviour). This means that the returns from wealth in a single year become of the same order or higher than current wealth, which happens for values around

$$W_n \approx \alpha_{n+1}^{-1/(\gamma-1)}. \tag{4.6}$$

Billionaire return data in Figure 4.1 below indeed confirm that RORs of around 100% or more can be achieved. From numerical results in Section 4.5 we see that this crossover leads to a two-tailed structure of the distribution of W_n similar to what

[1] For a reminder the symbol $\asymp$ means that $W_n = W_0 \exp\left(\mu n + \sqrt{n\nu^2} Z + o(\sqrt{n})\right)$ as $n \to \infty$, with Bachmann-Landau (or little o) notation such that $o(a_n)/a_n \to 0$ for all positive sequences $(a_n : n \in \mathbb{N})$.

[2] Likewise $W_n \asymp \left(W_0 e^D\right)^{\gamma^n}$ means $W_n = \left(W_0 e^D\right)^{\gamma^n + o(\gamma^n)}$.

we see in the data in Figure 1.1, and we think this feature of the model provides a promising explanation for this effect. Since we find in the next section that γ is close to 1, (4.6) is very sensitive to the value of the random variable α_{n+1} (which is raised to a large power), leading to a broad crossover region. While this crossover is a realistic feature seen in data from the UK and other countries ([145], but notably not in the USA, see online Appendix of [145]), the non-linearity also implies that the model is not scale invariant and coefficients will heavily depend on the currency unit.

We further find empirically that α_n is mostly positive with a heavy tail, but negative values are possible, see Figure 4.3 of Section 4.4.2, and thus W_n may become negative. Since our dynamics (4.1) are not built to describe agents in debt, we replace W_n with one of three replacement mechanisms discussed in Section 4.5.1. We note that bankruptcy events where agents' losses exceed their current wealth are realistic and do occur, but in this paper we focus on modelling the dynamics of agents with positive wealth.

We also note that both, the non-stationary linear and super-linear models, exhibit **monopoly**, where the wealth fraction of the richest agent in a system of N independent agents tends to 1 as time $n \to \infty$. This behaviour is well known for distributions with heavy tails (see e.g. Table 3.7 in [69]), which include the log-normal distribution in the linear case (4.4), and is only more pronounced in the super-linear model with heavier tails. We present related numerical results for the Gini coefficient and the top 1% wealth share in simulations, both tending to 1 in the long-time limit. While of course this extreme limit is not realistic currently, inequality measures are well known to increase since the 1980s (see summary in Chapter 1, Section 1.6). This is consistent with understanding current wealth distributions as transient behaviour of our model, which leads to monopoly if $\gamma \geq 1$ remains unchanged over time. Of course we can only parametrise our model over the current range of wealth values, and in order to get more realistic forecasts for future wealth distributions, we would have to include also the lifetime and inheritance dynamics for agents and the role of external influences (such as war or other catastrophies). The simplified model we present here explains how current wealth distributions can arise naturally from generic initial conditions, and we discuss possible refinements for further study in Section 4.6.

4.3 Non-Linear Kesten Process Theory

We now present the theory of the non-linear Kesten process in more detail. The wealth returns R_n at time n are assumed to have the following relationships

$$R_{n+1} = \frac{W_{n+1} - W_n - S_{n+1}}{W_n}, \quad R_{n+1} = \alpha_{n+1} W_n^{\gamma-1} \tag{4.7}$$

where α_n is drawn from some i.i.d. RV and $S_n \geq 0$ is wealth independent savings. We then combine (4.7) to analyse the following non-linear process with $\gamma > 1$ as in (4.3):

$$W_{n+1} = W_n + \alpha_{n+1} W_n^{\gamma} + S_{n+1} \,.$$

We can re-write (4.3) in the same form as a Kesten process

$$W_{n+1} = A_{n+1}(W_n) W_n + S_{n+1}, \qquad \text{where } A_{n+1}(W_n) = 1 + \alpha_{n+1} W_n^{\gamma-1}$$

where now the prefactor on the W_n is non-linearly dependent on W_n. Thus, as mentioned, we refer to (4.3) as a **non-linear Kesten process**. However we note that unlike with a strict Kesten process we can allow savings to be zero.

We will see empirical analysis for fitting the distribution of α_n, which could be negative, in Figure 4.3 of Section 4.4.1. However here assume for simplicity that $\alpha_n > 0$, which implies that W_n is increasing and strictly positive for all $n \geq 0$. Negative values of α_n will lead to bankruptcy events as $n \to \infty$ which we shall consider in Section 4.5. Taking logarithms of (4.3) leads to

$$\begin{aligned} \log W_{n+1} &= \log\left(W_n + \alpha_{n+1} W_n^{\gamma} + S_{n+1}\right) \\ &= \gamma \log W_n + \log\left(\alpha_{n+1} + 1/W_n^{\gamma-1} + S_{n+1}/W_n^{\gamma}\right) \\ \text{so that} \quad X_{n+1} &= \gamma X_n + B_{n+1}, \end{aligned} \tag{4.8}$$

where $X_n := \log W_n$ and $B_{n+1} := \log(\alpha_{n+1} + 1/W_n^{\gamma-1} + S_{n+1}/W_n^{\gamma})$. Now using (2.20) we get

$$X_n = \gamma^n \Big(X_0 + \sum_{k=1}^{n} B_k \gamma^{-k} \Big) \quad \text{so that} \quad \frac{X_n}{\gamma^n} \xrightarrow{d} X_0 + D \quad \text{as } n \to \infty, \tag{4.9}$$

where $D := \sum_{k=1}^{\infty} B_k \gamma^{-k}$. Since $W_n > 0$ is increasing with n and α_n are i.i.d., B_k are bounded random variables, so $D \in (0, \infty)$ is a well defined random variable since

$\gamma > 1$. Thus, as $n \to \infty$, this implies to leading exponential order[3]

$$\frac{X_n}{\gamma^n} \simeq X_0 + D \quad \text{so that} \quad W_n \asymp \left(W_0 e^D\right)^{\gamma^n} . \tag{4.10}$$

We can see from (4.10) that W_n exhibits super-exponential growth[4].

Crossover

We note a particular aspect of model (4.3) which we call the **crossover**. This happens when the model starts to exhibit super-exponential growth and occurs when the $\alpha_{n+1} W_n^\gamma$ term starts to dominate the W_n term in (4.3) corresponding to returns greater than 1. Solving

$$\begin{aligned} \alpha_{n+1} W_n^\gamma &\geq W_n \ \text{ we have} \\ W_n &\geq \frac{1}{\alpha_{n+1}^{1/(\gamma-1)}} \end{aligned} \tag{4.11}$$

and we find super-exponential growth happens for wealth values greater than (4.11). We note that the α_n are random so (4.11) is not an exact point but would correspond to a region for typical values of α_n.

4.4 Empirics

Before moving on to the simulations of the non-linear Kesten process (4.3) we undertake some key empirical analysis to parametrise the model. We calculate returns on wealth, R_n, and the prefactor, α_n, and make statistical fits on these variables. Although savings do not evolve with wealth as mentioned above, they are correlated with initial wealth values of an agent as part of their social status or fitness. To infer this dependence, we look at UK income and expenditure data for the year 2016 [3, 2].

4.4.1 Returns R_n

As in Section (4.3) from (4.3) we rearrange to find the ROR as

[3] So that $W_n = \left(W_0 e^D\right)^{\gamma^n + o(\gamma^n)}$ as $n \to \infty$ where o is again the Bachmann-Landau or little o notation.

[4] Formally W_n exhibits super-exponential growth when $\lim_{n\to\infty} \frac{W_n}{c^n} = \infty$ for all $c > 0$.

$$R_{n+1} = \frac{W_{n+1} - W_n - S_{n+1}}{W_n} \approx \frac{W_{n+1} - W_n}{W_n} \quad \text{for billionaires.} \tag{4.12}$$

For wealthy agents, wealth gain is to a large extent dominated by returns on wealth, so that $W_{n+1} - W_n \gg S_{n+1}$ and savings can typically be ignored. The ROR is then simply given by the wealth growth rate, which we will use to compute R_n for billionaires, while we include savings to estimate ROR from survey data for other agents.

As mentioned previously, fairly recent work [54, 14, 50] has suggested an increasing wealth dependence on returns. We also find empirical evidence for this from WAS as summarised in Figure 4.1, and assume a simple power-law relationship as in (4.2) which is roughly consistent with the data.[5] According to this relationship we have

$$\mathbb{E}[R_{n+1}|W_n] = \mu W_n^{\gamma-1} \ , \quad \text{where} \quad \mu = \mathbb{E}[\alpha_{n+1}]\,. \tag{4.13}$$

We fit the power-law exponent γ and the prefactor μ as shown in Figure 4.1. From Figure 4.2 we see evidence that returns are independent across time and the variance of returns is proportional to the square of the mean returns as wealth increases with the particular fit,

$$\mathrm{var}(R_{n+1}|W_n) \approx 0.57\,\mathbb{E}[R_{n+1}|W_n]^2\,. \tag{4.14}$$

Such a quadratic scaling relationship of mean and variance is common in multiplicative processes, and consistent with our model assumption (4.2), as is explained in Appendix C.3.

Note that the apparent structure in percentile returns data in Figure 4.1 for individual years does not constitute reliable information in our view, since the variation of the points is artificially decreased due to our numerical procedure as explained in Appendix C.4. Viewing all years as a combined dataset, we find an increasing wealth dependence of RORs consistent with a simple power-law relationship, which also matches well with data for billionaires. In the next subsection we present a method to estimate a reasonable value of the power-law exponent γ so that both, WAS and billionaire return data, can be modelled well with our assumption on returns (4.2).

[5] However we note the relationship may be more complex and that returns are likely a more complicated function of wealth with a flattening of returns before the very rich, see Figure 2 in [50].

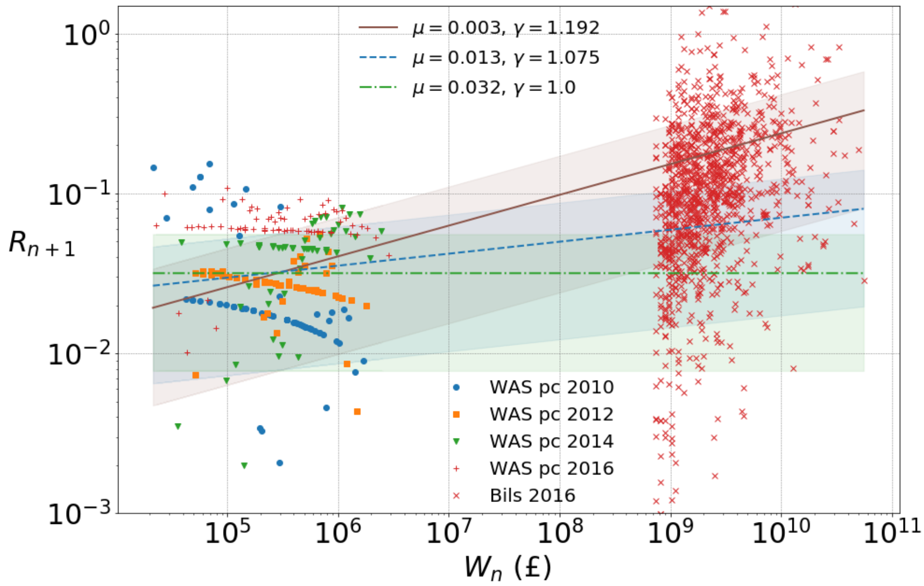

Figure 4.1: Percentile ROR using WAS data [116] for the years 2010, 2012, 2014 and 2016, and ROR for individual billionaires for 2016 [1]. Power law fits according to (4.13) to the cluster of WAS ROR data combined over all four time periods, leads to $\mu \approx 0.003$, $\gamma \approx 1.192$ (with both parameters free) and to $\mu \approx 0.013$ with chosen $\gamma = 1.075$ (justified below in Figure 4.3). We also include $\gamma = 1$ for comparison, leading to $\mu \approx 0.032$, i.e. an average ROR of about 3%. Respective shaded regions are one standard deviation around the power fit means (4.13) as explained in Appendix C.3.

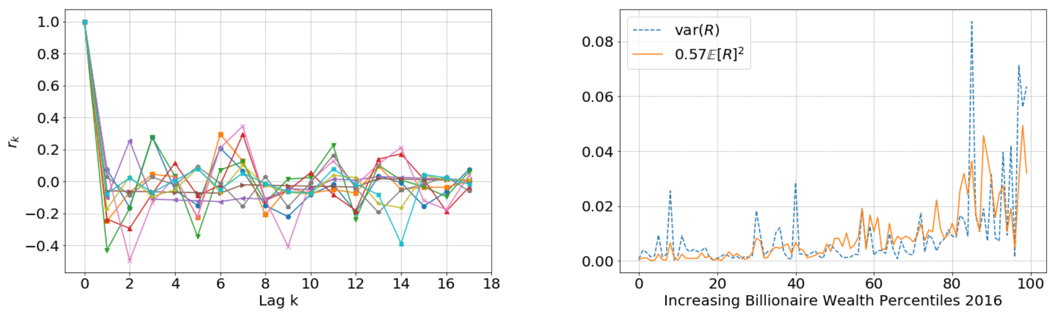

Figure 4.2: Left: autocorrelation of a sample of billionaire ROR indicating independence in returns. Right: average annual billionaire returns from 2008-2016 Forbes list [1], showing mean and variance relationship for increasing wealth percentiles as in (4.14).

4.4.2 Fitting α_n

With (4.3) we have in analogy to (4.12)

$$\alpha_{n+1} = \frac{W_{n+1} - W_n - S_{n+1}}{W_n^\gamma} \approx \frac{W_{n+1} - W_n}{W_n^\gamma} \quad \text{for billionaires.} \tag{4.15}$$

As illustrated in Figure 4.3, we choose the power-law exponent $\gamma = 1.075$, such that the return data from the WAS and billionaires can be best explained with a single power law of the form (4.2). We fit the distribution of the α_n (which we assume to be i.i.d.) with a shifted and scaled non-central t-distribution (nct), i.e. we take

$$\alpha_n \sim \mathrm{nct}(k, c, l, s)\,.$$

This distribution has four parameters: $k > 0$ represents the degrees of freedom controlling the heaviness of the tail, $c \in \mathbb{R}$ is the centrality that controls the skewness of the distribution, $l \in \mathbb{R}$ is the shift and $s > 0$ is the scale, see Appendix C.5 for details.

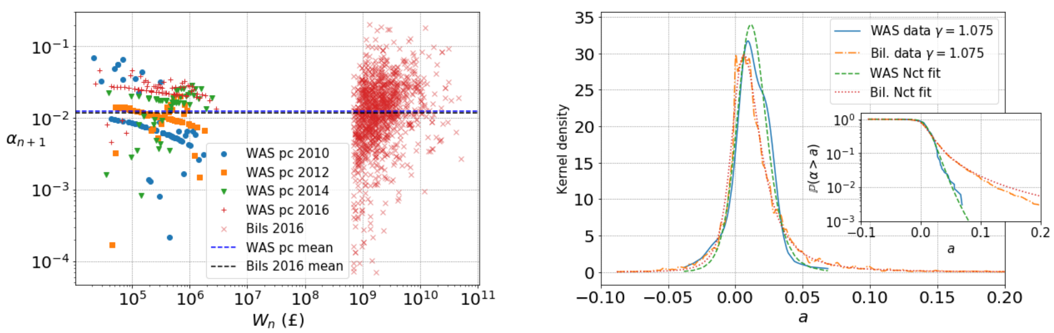

Figure 4.3: Left: α_{n+1} (4.15) for WAS data percentiles [116] for four time periods along with 2016 billionaire data plotted against wealth W_n. We choose $\gamma = 1.075$ so that the means of WAS and billionaire data essentially agree (dotted lines). Right: Kernel density of α_{n+1} for WAS data and 2016 billionaire data as seen in the left Figure. Inset: corresponding empirical tails $\mathbb{P}(\alpha_n > a)$ on logarithmic scale. Dotted green and red lines provide fits by the non-central t-distribution (nct) to WAS and billionaires with respective nct parameter fits $k \approx 6.03$, $c \approx 0.0573$, $l \approx -0.00575$, $s \approx 0.0112$ and $k \approx 2.01$, $c \approx 0.941$, $l \approx -0.00156$, $s \approx 0.0112$.

We find that, while the bulk of the distributions of α_n agree well, the billionaire data lead to heavier tails than WAS data. Again, our method of extracting returns from WAS data leads to decreased fluctuations, and therefore we use the parameter values corresponding to billionaire data in simulations in Section 4.5.

4.4.3 Savings S_n

We recall that in our model (4.1) savings S_n represent all contributions to wealth growth that are independent of the current wealth of an agent. They do not evolve with increasing wealth and only contribute additive noise, which does not influence the long-time behaviour of the dynamics. However, we need to estimate savings and their correlation with (initial) wealth to run simulations, and in particular in order to extract empirical RORs from wealth data using (4.12), which determine the statistics of the crucial parameter α_n. [86] presents evidence for recent years in the US, that income and salary are positively correlated with wealth.

We estimate savings by equivalised disposable income after expenditure for increasing deciles of median wealth using ONS data sources [2, 3]. Equivalised disposable income is household size adjusted income available for spending after tax and deductions, and by expenditure we summarise costs that do not contribute to wealth, such as buying food or paying rent. We fit the dependence on wealth w with a logistic function

$$S(w) = \frac{\kappa_1}{1+\kappa_2 w^{\kappa_3}} \quad \text{with parameters } \kappa_1, \kappa_2 > 0 \text{ and } \kappa_3 < 0\,. \tag{4.16}$$

This is illustrated in Figure 4.4, where we show data on equivalised disposable income, household expenditure and give the fitted parameter values for (4.16).

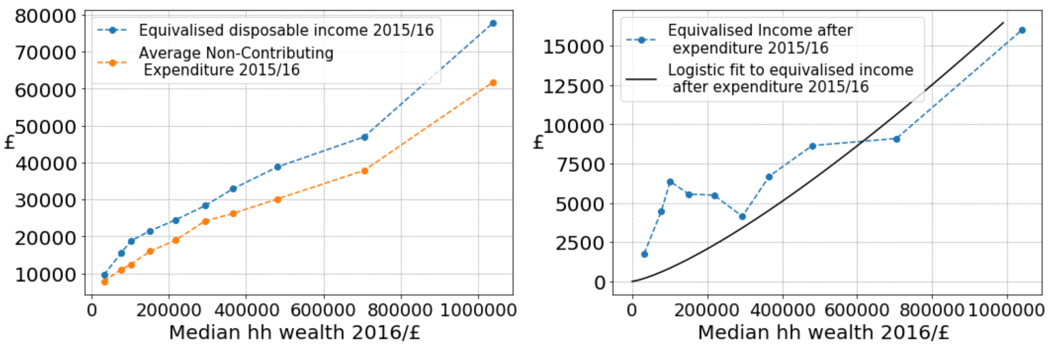

Figure 4.4: Left: Plot of equivalised disposable income and average household expenditure for 2015/16 against 2016 median wealth deciles. Right: Fit of the logistic function (4.16) to equivalised disposable income after expenditure, where we choose $\kappa_1 = 10^6$ and fit $\kappa_2 = 4.13 \cdot 10^9$ and $\kappa_3 = -1.308$. ONS data sources used can be found in [2, 3].

We used (4.16) as an estimate for additive contributions to wealth growth when calculating percentile returns in Figure 4.1, see Appendix C.4 and in simulations in Section 4.5.2 as a function of initial wealth $w = W_0$. Note that the logistic fit levels off at $\kappa_1 = 10^6$ for large values of w which is an arbitrary cap of 10^6 GBP on wealth independent savings. For most rich households, contributions to wealth growth significantly beyond this scale are in the form of wealth returns. It is important to note that none of our results are sensitive to the choice of parameters κ_1, κ_2 and κ_3, since savings only really play a role in parameter estimation or simulations on the scales shown in Figure 4.4.

4.5 Simulation Results

For all simulations presented in this section we use i.i.d. $\alpha_n \sim \mathrm{nct}(k, c, l, s)$ with parameters

$$k = 2.008 \ , \quad c = 0.941 \ , \quad l = -0.00156 \quad \text{and} \quad s = 0.0112 \ , \tag{4.17}$$

corresponding to data from individual billionaires which represent our best estimate of fluctuations for individual households for $\gamma = 1.075$ see Figure 4.3. We do, however, experiment with changing γ values in which case we multiply the α_n by a positive constant to keep the mean at the same level. This is explained further in Section 4.5.1.

4.5.1 Generic Initial Conditions with Zero Savings

To investigate the general properties and dependence on initial conditions of our model over longer time horizons, we consider the following four different initial conditions each with mean 10000:

I.1 $W_0 = 10000$ (BLUE •)

I.2 $W_0 \sim 5000 + \mathrm{Exp}(1/5000)$ (ORANGE ■)

I.3 $W_0 \sim \mathrm{Exp}(1/10000)$ (GREEN ▼)

I.4 $W_0 \sim \mathrm{Pareto}(5000, 2)$ (RED +)

In other words, in **I.1** all agents start with initial wealth 10000, in **I.2** agents get 5000 plus an exponentially distributed random amount with mean 5000, in **I.3** initial wealth is drawn from an exponential with mean 10000 and in **I.4** it is Pareto distributed with parameters $x_m = 5000$ and exponent $\beta = 2$.

It is also possible in our simulations for the wealth $W_n(i)$ of an agent i to become negative. In this case we choose one of the following replacements for $W_n(i)$:

R.1 replace with a proportion of the agent's previous positive wealth value $pW_{n-1}(i) > 0$ such that p is uniformly chosen from $(0, 1]$;

R.2 replace with the agent's previous positive wealth value $W_{n-1}(i) > 0$;

R.3 replace with wealth $W_n(j) > 0$ of another uniformly chosen agent j.

We can think of **R.1** as the agent losing a random proportion of wealth, **R.2** as no change in the agent's wealth and **R.3** as the agent being removed from the system and being replaced uniformly with another agent with positive wealth. We note that **R.3** is a simple approximation to resampling the agent's wealth from the current wealth distribution. We focus here on simulations with the more realistic compromise mechanism **R.1**. In Appendix C.6.1 we will present simulation results for the more extreme replacement mechanisms **R.2** and **R.3** which lead to similar results, confirming that our model is not very sensitive on the choice of the replacement mechanism.

For each initial distribution we run the simulations iteratively using (4.3) for $N = 10^6$ independent agents and **zero savings** $S_n = 0$ with parameters in (4.17) and replacement mechanism **R.1**. We choose zero savings for convenience in this section, to isolate the effect of the multiplicative dynamics which is dominant in generating the wealth distribution in this model, see Section 4.3. Results for empirical tail distributions at times $n = 10, 100, 200$ and 300 are presented in Figure 4.5, using the colour code indicated in **I.1-I.4**. We also show standard inequality measures (see Chapter 1 Section 1.6 for the definitions), the Gini coefficient g and the top one percent income share $s_{0.01}$ for $\gamma = 1.075$ up to time $n = 300$ in top left and right of Figure 4.7. We see that all initial conditions eventually lead to monopoly, and for intermediate times power-law tails emerge in the wealth distribution. Due to the crossover (4.6) to super-exponetial growth, a two-tailed structure emerges for large times and wealth values.

In Figure 4.6 we show for comparison empirical tails for $\gamma = 1.19$ with $\alpha_n \sim 0.23 \cdot \mathrm{nct}(k, c, l, s)$, and for $\gamma = 1$ with $\alpha_n \sim 2.5 \cdot \mathrm{nct}(k, c, l, s)$, so that average ROR values are well approximated with different fits for $\mu = \mathbb{E}[\alpha_{n+1}]$ (4.13) as shown in Figure 4.1. For $\gamma = 1$ we also compute the two inequality measures g and $s_{0.01}$ up to $n = 400$, see bottom left and right of Figure 4.7 which shows the independence of initial conditions and slower progression towards monopoly. For the higher value of γ we see that the crossover sets in earlier at more realistic wealth values around 10^7 with a two-tailed structure with quite realistic power-law tails (cf. Figure 1.1). For the linear model with $\gamma = 1$ we see no crossover and can fit the distribution for large times well by a log-normal distribution in accordance with (4.4). In this case there is also no noticeable difference between distributions originating from different initial conditions as we have seen in Figure 4.7. This is also illustrated in Figure 4.8, where we also see a clear dependence of final wealth values on initial conditions in the non-linear case with $\gamma > 1$.

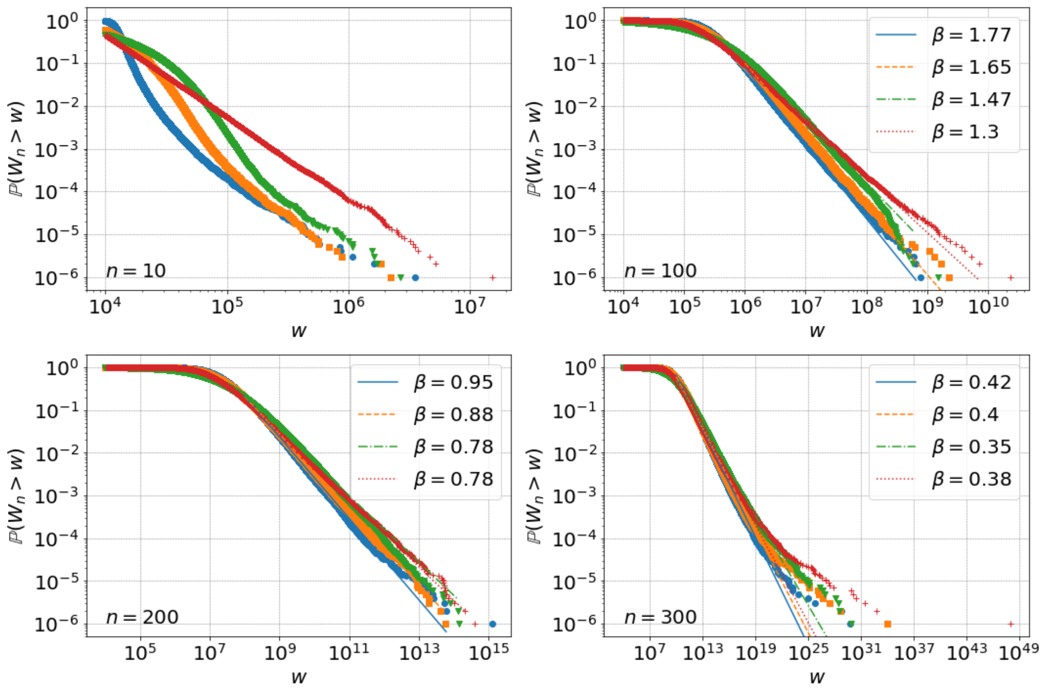

Figure 4.5: Empirical tails for simulation (4.3) with $N = 10^6$ agents, zero savings $S_n = 0$, $\alpha_n \sim \mathrm{nct}(k, c, l, s)$ with $\gamma = 1.075$, fitted parameters in (4.17), the four initial conditions **I.1**-**I.4** with colour code, and replacement mechanism **R.1**. Power law fits show heavier tails with exponents β decreasing with increasing times $n = 10, 100, 200$ and 300.

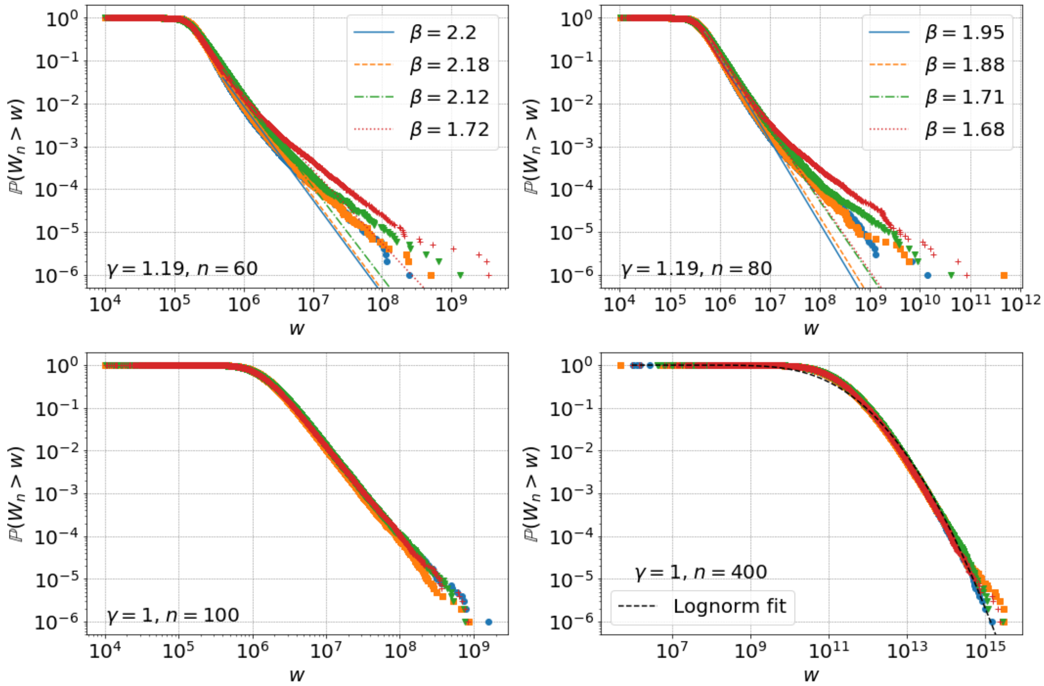

Figure 4.6: Top left and right: empirical tails for simulation (4.3) with $N = 10^6$ agents, zero savings $S_n = 0$, $\alpha_n \sim 0.23 \cdot \mathrm{nct}(k, c, l, s)$ with fitted parameters (4.17) but with $\gamma = 1.19$ for the four initial conditions with respective colour coding **I.1**-**I.4**, replacement mechanism **R.1** and power law fits with exponent β. Bottom left and right: empirical tails for simulations as in top row, but with $\gamma = 1$, $S_n = 0$, $\alpha_n \sim 2.5 \cdot \mathrm{nct}(k, c, l, s)$, with lognormal fit at $n = 400$.

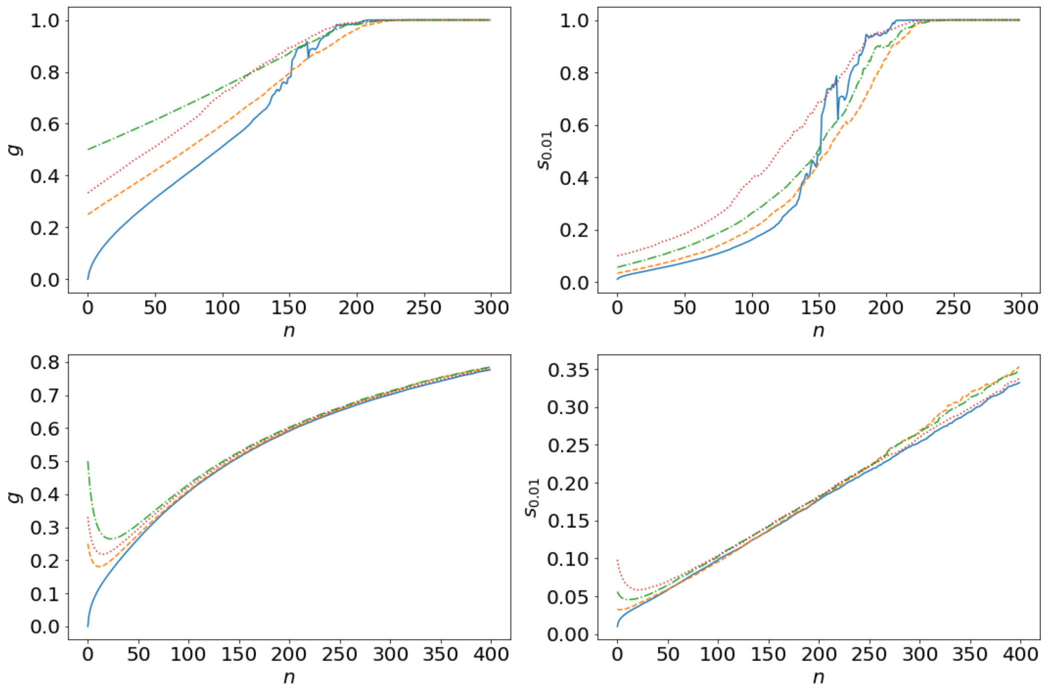

Figure 4.7: Gini, g, and top 1% wealth shares, $s_{0.01}$, for simulation (4.3) with $N = 10^6$ agents, zero savings $S_n = 0$, $\alpha_n \sim \mathrm{nct}(k, c, l, s)$ with fitted parameters in (4.17), $\gamma = 1.075$ for top left and right and $\alpha_n \sim 2.5 \cdot \mathrm{nct}(k, c, l, s)$, $\gamma = 1$ for bottom left and right. The four initial conditions with respective colour coding **I.1**-**I.4** are used with replacement mechanism **R.1**.

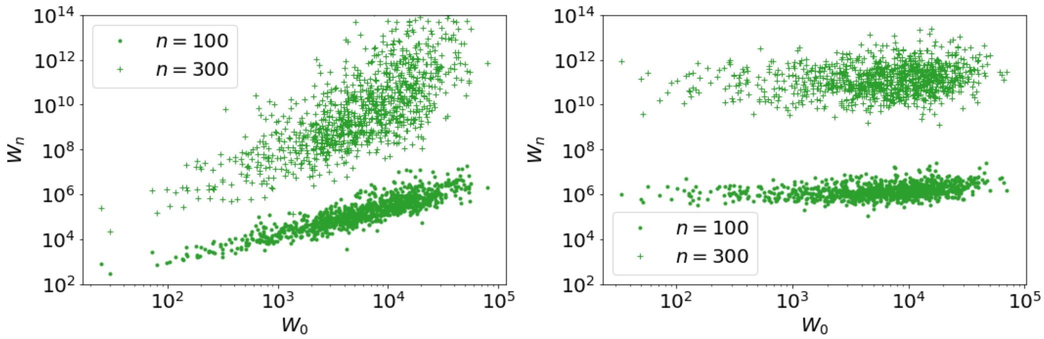

Figure 4.8: W_0 versus W_n for 1000 randomly chosen agents for simulation (4.3) with $N = 10^6$ agents, zero savings $S_n = 0$ with fitted parameters in (4.17), left $\alpha_n \sim \mathrm{nct}(k, c, l, s)$, $\gamma = 1.075$ and right $\alpha_n \sim 2.5 \cdot \mathrm{nct}(k, c, l, s)$ and $\gamma = 1$. We use initial conditions with colour coding **I.3** and replacement mechanism **R.1**. We see a clear dependence on initial conditions for $\gamma > 1$, and essentially no dependence for $\gamma = 1$.

4.5.2 Realistic Initial Conditions

In this section we simulate a realistic scenario for the UK, with $N = 23 \cdot 10^6$ households, initial conditions W_0 extracted from the UK wealth distribution in 2008, and with fixed savings $S_n = S(W_0)$ as given in (4.16) of Section 4.4.3. Figure 4.9 shows the empirical tail of the resulting wealth distribution at times $n = 0, 2, 4, 6, 8, 10, 20$ and 50, after simulating (4.3) with $S_n = S(W_0)$, $\gamma = 1.075$, $\alpha_n \sim \text{nct}(k, c, l, s)$ with fitted parameters in (4.17) and replacement mechanism **R.1**. Figure C.4, in Appendix C.6.2, shows empirical tails for the other two replacement mechanisms **R.2**, **R.3** which lead to very similar results. The number of agents (N) is a rough estimate for the number of households in the UK with positive wealth in 2016. Time n corresponds to the number of years after 2008, so for example $n = 8$ corresponds to 2016. Again we can see increasing inequality, see Figure C.5 in Appendix C.6.2, with the decreasing power-law exponent β. We also see that the average returns for the simulations, see Figure 4.10, matches roughly the shape of the RORs seen in Figure 4.1.

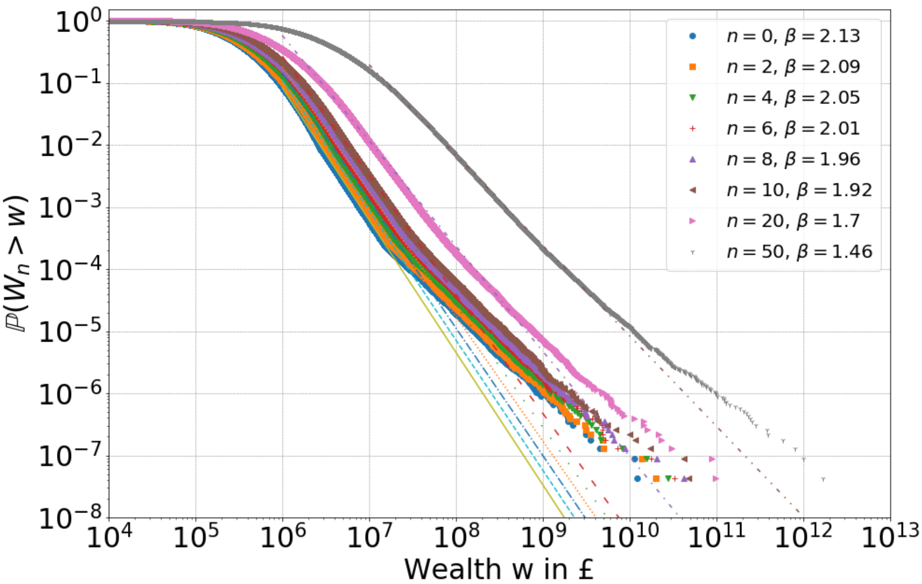

Figure 4.9: Empirical tails for simulation (4.3) with $N \approx 23 \cdot 10^6$ agents, replacement mechanism **R.1**, $\gamma = 1.075$, fixed savings $S_n = S(W_0)$ (4.16), $\alpha_n \sim \text{nct}(k, c, l, s)$ with fitted parameters in (4.17) for 2008 initial conditions. Fit values for a power-law tail exponent β decrease from the initial value 2.13.

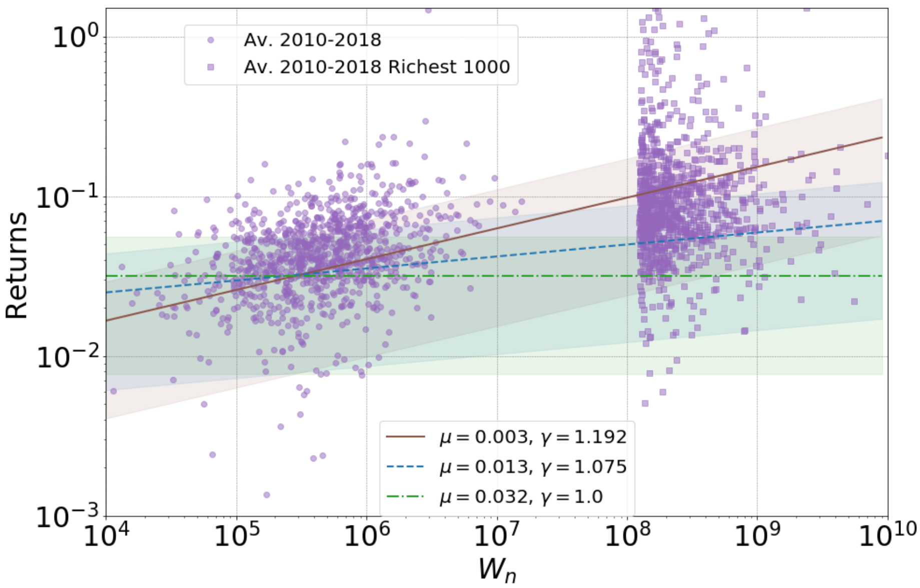

Figure 4.10: Average return over 2010-2018 for 1000 randomly chosen agents and 1000 richest agents in 2018 against agents wealth in 2018 for simulation (4.3) with $N \approx 23 \cdot 10^6$ agents, replacement mechanism **R.1**, $\gamma = 1.075$, fixed savings $S_n = S(W_0)$ (4.16), $\alpha_n \sim \mathrm{nct}(k, c, l, s)$ with fitted parameters in (4.17) for 2008 initial conditions. The power fits (straight lines) for $\mathbb{E}[R_{n+1}|W_n] = \mu W_n^{\gamma-1}$, are the three fits to the real world data from Figure 4.1, along with one standard deviation error region.

Comparing Figure 4.9 to Figure 1.1 we see that the two-tailed structures differ slightly: While the heavier tail for billionaires with a power-law exponent of about $\beta = 1$ is shifting but well preserved, the stability of the lighter power-law tail for millionaires is not well represented in our simulation. This is because we deliberately chose a simple model assuming that average ROR follows a monotone power law with wealth. While this is largely consistent with data, the survey data for RORs show some plateau behaviour for millionaires clearly visible in Figure 4.1, which has also been suggested for other countries, see Figure 2 of [50]. This may be related to the changing wealth composition of the very rich [45].

4.6 Discussion

The model defined by the iterative equation (4.1) represents a generic evolution of household wealth, based on the well motivated assumption that wealth exchange between households does not play an important role but is governed primarily by wealth returns. The particular form (4.3) of a non-linear Kesten process has been motivated by inferring empirically that RORs increase with household wealth, and that this relationship is consistent with a simple power law with exponent γ as in (4.2), see also Figure 4.1. We want to stress that the qualitative results and main features of our model do not depend on this particular choice, which we have taken for simplicity and in order to study the effect of the non-linearity with a single parameter. We have seen from theory and simulations that the asymptotic dynamics of the model (4.3) and the resulting tail of the wealth distribution is dominated by the exponent γ. For the linear case with $\gamma = 1$ the RORs do not depend on wealth, and it is known that wealth grows asymptotically with a lognormal distribution (see Section 2.6), which does not correspond to power-law tails seen in real data as in Figure 1.1. As demonstrated by our main results, the non-linear model with $\gamma > 1$ exhibits power-law tails from generic initial conditions, including even perfect equality or light tailed exponential distributions, see Section 4.5.1. It also leads to a two-tailed structure resulting from a crossover (4.6) to super-exponential growth for the richest households.

We now summarise the most important theoretical features and differences of the linear ($\gamma = 1$) and the non-linear ($\gamma > 1$) non-stationary Kesten process (4.3):

- for all $\gamma \geq 1$, including the linear case, the model exhibits **monopoly**[6], i.e. for N independent households the wealth fraction of the richest household increases with time and asymptotically approaches 1[7];

- the linear model is **ergodic**[8], in the sense that the asymptotic exponential growth rate of household wealth does not depend on the initial condition W_0. The latter only enters as a multiplicative factor and the model is **scale invariant**, i.e. wealth can be measured in units of W_0 in a dimensionless way;

- the non-linear model is **not ergodic**[9], i.e. the asymptotic exponential growth rate depends on W_0 and the early dynamics. It is also **not scale invariant**, and the non-linearity on the right hand side leads to a **critical scale** (4.6) where wealth gain per year can exceed current wealth, which is observed in data for the richest households.

[6] A less strict version than in the balls in process with feedback seen in Chapter 2 and 3.

[7] Nevertheless, realistic levels of inequality can of course be achieved on intermediate timescales.

[8] In particular the log returns $\log(W_{n+1}/W_n)$ tend to an ergodic process as $n \to \infty$. See [120] for more discussion on transforming non-ergodic processes to ergodic ones in an economics/wealth context.

[9] In particular the log returns $\log(W_{n+1}/W_n)$ explode as $n \to \infty$ and is thus non-ergodic.

Moreover, we would like to stress that our model is phenomenological and not built from first principles, since we simply assume an empirically motivated non-linear relationship between ROR and current wealth. Therefore the model lacks a natural scale invariance and the parameter α_n is not universal, but depends on the units of measurement (the currency) and will vary between different countries/economic areas. On the other hand, the non-linearity induces a crossover scale that can be a possible explanation for an apparent two-tailed structure in the data. This is an important aspect of our model and although analysed in Chapter 1 should be investigated further. While not present in data from the USA, the two-tailed structure has been observed [145] for several countries which have a less liberal economic system and put more emphasis on social equality. Related political measures such as taxation then lead to a more even wealth distribution and a lighter power-law tail for rich households including millionaires, while the richest in society distribute their wealth globally and can escape such measures, leading to a heavier tail for billionaires.

Other interesting generalisations to make the model more realistic include dynamics for negative wealth, a realistic treatment of bankruptcy events and also household lifetime and fragmentation over longer time periods, or a household dependence of the parameter α_n reflecting variations in 'fitness' to generate returns from investment. Also, mechanisms of household interaction possibly via a general redistribution or taxation procedure could be included and could lead to interesting effects on the dynamics similar to recent work in [19]. But the aim of this chapter was to introduce a simple model, that can explain the main features of wealth distribution and dynamics, and how they can be explained by a non-linear wealth dependent rate of return.

Conclusion

This thesis, by means of statistical analysis and mathematical modelling, attempted to provide further insight into aspects of the wealth distribution and key mechanisms for how the distribution is generated. We have focused to a large extent on the power law, a feature we see in the upper tail of the wealth distribution. A significant change in power law exponent between the UK survey rich and rich lists was found in alignment with previous research [145]. We saw that survey bias could account for this change but other reasons such as a difference in measurement between survey and rich list may be a more realistic explanation [7]. Future work would be needed to be certain there is only one power law in the tail. As well as analysing various methods for fitting the power law we also fitted the entire distribution of both negative and positive UK wealth with a mixture distribution. For positive wealth we found that an exponential as well as lognormal distribution is a good fit before the power law and that the κ-generalised distribution [32] fitted reasonably to the whole of positive wealth. However there are potentially other distributions that would also fit positive wealth just as well and it would be further study to compare them. Also, different methods of fitting the positive wealth distributions such as MLE would be useful as a comparison to the NLLS method used in Chapter 1. We focused our study on UK wealth data between 2008-2016. It would be interesting to see how the distribution evolves over a longer period of time and to compare the distribution between countries, however this is limited by data availability [121].

Our wealth modelling, specifically for positive wealth, was classified into the general themes of hierarchy, exchange, feedback and multiplicative processes. Various models of these types were analysed in Chapter 2. We found that the Pareto distribution as well as the κ-generalised distribution can be arranged in a hierarchy suggesting there are some stabilising forces that the structure of hierarchies provide to the economy. We then considered exchange models, originating in the econophysics literature, where agents repeatedly exchange with each other. Under certain conditions when the amount exchanged is fixed proportionally to the agent or the rate of exchange differs between agents we find power laws can emerge. Finally the balls in bins feedback model and various multiplicative processes were analysed from the stand point of both discrete and continuous time. The multiplicative process models were characterised by repeated random multiplications. The Kesten process, a linear model, is known to give a power law stationary distribution under certain conditions by purely random effects [88]. We also showed several SDEs can produce power law stationary distributions, including the κ-generalised distribution, by solving the stationary Fokker Planck equation. Many of the models in Chapter 2 have not been fitted adequately to data. This would be necessary to show the models can fit more of the distribution than just the power law.

Models generating non-stationary power law distributions for growing wealth were found by adding a non-linear rich gets richer power law term. This was true for both the balls in bins process or feedback model analysed in Chapter 3 and the non-linear Kesten process examined in Chapter 4. Both these models although dependent on initial conditions, produce power law tails and extreme inequality in the long time limit for all initial conditions for a critical parameter $\gamma > 1$. We note the balls in bins feedback model produces a stronger monopoly than the non-linear Kesten process. In the feedback model richer agents gained more wealth with probability proportional to W^γ with $\gamma > 1$ and at each iteration there is a dependence between all agents. In the non-linear Kesten model richer agents were more likely to gain more wealth through returns proportional to $W^{\gamma-1}$ with $\gamma > 1$ and at each iteration all agents run independently of each other.

In Chapter 3 the master equation under the backdrop of exponential embedding was used to find some new results of the balls in bins process with feedback in continuous time for a general $N \in \mathbb{N}$ agents. We found that there is an 'explosion' time which proves monopoly and a means of calculating the expected wealth of an agent over time. We found time dependent probability mass functions for the case with wealth packet of one unit entering the system and an initial condition of one unit per agent including a general sum solution for $\gamma \neq 0$ that can be numerically calculated. Assuming a power law solution to the process an exact relationship $\beta = \gamma - 1$ where $\gamma > 1$ and β is the power law exponent of the tail distribution was also found. How the process scales with more agents, changes with differing wealth packets and initial distributions and how the rank ordering of the poorest to richest agents changes throughout the process are several questions for further study. We slightly extended the feedback model to include fitness, which we think of as something like intelligence or productivity. We found higher values of fitness leads to higher expected wealth. Specifying exactly what fitness is and then finding what distribution it follows from data would be subject to future work. More elaborate feedback functions could also be analysed and simulated in the feedback model. We also ran several simulations that gave evidence for the theoretical results in the chapter.

In Chapter 4 our original non-linear Kesten process was run from a realistic 2008 UK wealth initial distribution for the number of agents approximating the number of UK households. We found that running the model till 2016 produces higher inequality but not at an unrealistic level. From the UK data for 2008-2016, we saw in Chapter 1 that the Gini coefficient appeared to remain roughly constant with fairly fluctuating power law tails. However due to data errors there may in fact have been a rise in inequality in alignment with the model [7]. Thus we conclude that the non-linear Kesten process may be useful over short or intermediate time scales but over longer time scales where inequality even decreases new additions to the model would need to be included. For instance γ values could be fitted for multiple years or a different returns function entirely may be found and used instead of the power law relationship in the model. Although this model included a replacement of agents if they go into debt further addition of shocks could be added. We also ran the model under general initial conditions showing that power laws and extreme inequality appear even for equal initial conditions. Key theoretical results of the non-linear Kesten process also included the non-linear dependence of initial conditions and the critical region after which super-exponential growth sets in. These results were also backed up by simulations.

There is always a conflict between having a model simple enough to be analysed and simulated and complex enough to take into account the reality. Population changes, household size and type, age, technological changes and shocks such as recessions, war and pandemics are all factors that can effect the distribution. However this thesis focused on analysing only one general factor encapsulated by the phrase *the rich get richer*. We showed that this effect often characterised by a non-linear factor in mathematical models is enough to explain power laws and high inequality both seen in the wealth distribution.

References

[1] Forbes rich list data. `https://github.com/abresler/forbesListR`. (Cited on pages 12, 28, 29, 89, 117, 126, 127, and 136.)

[2] ONS expenditure data. `https://www.ons.gov.uk/peoplepopulationandcommunity/personalandhouseholdfinances/expenditure/datasets/detailedhouseholdexpenditurebydisposableincomedecilegroupuktable31`. (Cited on pages 87, 91, and 136.)

[3] ONS income data. `https://www.ons.gov.uk/peoplepopulationandcommunity/personalandhouseholdfinances/incomeandwealth/datasets/householddisposableincomeandinequality`. (Cited on pages 87, 91, and 136.)

[4] Times 2020 UK rich list data. `https://www.thetimes.co.uk/sunday-times-rich-list#TableFullRichList`. (Cited on page 136.)

[5] UK data service. `https://ukdataservice.ac.uk/`. (Cited on page 11.)

[6] Abergel, F., Chakrabarti, B. K., Chakraborti, A., Deo, N., and Sharma, K. (2019). *New perspectives and challenges in econophysics and sociophysics*. Springer. (Cited on page 41.)

[7] Advani, A., Bangham, G., and Leslie, J. (2021). The UK's wealth distribution and characteristics of high-wealth households. *Fiscal Studies*, 42(3-4):397–430. (Cited on pages 1, 3, 13, 33, 35, 101, and 103.)

[8] Alvaredo, F., Atkinson, A. B., and Morelli, S. (2016). The challenge of measuring UK wealth inequality in the 2000s. *Fiscal Studies*, 37(1):13–33. (Cited on pages 5 and 6.)

[9] Alvaredo, F., Atkinson, A. B., and Morelli, S. (2018). Top wealth shares in the UK over more than a century. *Journal of Public Economics*, 162:26–47. (Cited on pages 1, 4, 32, 33, and 34.)

[10] Arthur, W. B. et al. (1994). *Increasing returns and path dependence in the economy.* University of Michigan Press. (Cited on pages 3 and 45.)

[11] Athreya, K. B. and Karlin, S. (1968). Embedding of urn schemes into continuous time markov branching processes and related limit theorems. *The Annals of Mathematical Statistics*, 39(6):1801–1817. (Cited on page 66.)

[12] Atkinson, A. B. (2018). Wealth and inheritance in Britain from 1896 to the present. *The Journal of Economic Inequality*, 16(2):137–169. (Cited on page 32.)

[13] Atkinson, A. B. and Croome, D. R. (1972). *Unequal shares: wealth in Britain.* London: Allen Lane. (Cited on page 4.)

[14] Bach, L., Calvet, L. E., and Sodini, P. (2016). Rich pickings? risk, return, and skill in the portfolios of the wealthy. (Cited on pages 83 and 88.)

[15] Bell, T., Brewer, M., Corlett, A., Hale, S., Handscomb, K., Judge, L., Leslie, J., Marshall, J., Murphy, L., Shah, K., Smith, J., Slaughter, A., and Thwaites, G. (2022). Inflation nation, putting spring statement 2022 in context. *Resolution foundation.* (Cited on page 1.)

[16] Benaroya, H., Han, S. M., and Nagurka, M. (2005). *Probability models in engineering and science*, volume 192. CRC press. (Cited on page 67.)

[17] Benhabib, J. and Bisin, A. (2018). Skewed wealth distributions: theory and empirics. *Journal of Economic Literature*, 56(4):1261–91. (Cited on pages 32 and 51.)

[18] Benhabib, J., Bisin, A., and Luo, M. (2017). Earnings inequality and other determinants of wealth inequality. *American Economic Review*, 107(5):593–97. (Cited on page 51.)

[19] Berman, Y., Peters, O., and Adamou, A. (2020). Wealth inequality and the ergodic hypothesis: Evidence from the United States. *Available at SSRN 2794830.* (Cited on pages 56 and 100.)

[20] Bingham, N. H., Goldie, C. M., Teugels, J. L., and Teugels, J. (1989). *Regular variation.* Number 27. Cambridge university press. (Cited on pages 8 and 9.)

[21] Black, F. and Scholes, M. (2019). The pricing of options and corporate liabilities. In *World Scientific Reference on Contingent Claims Analysis in Corporate Finance: Volume 1: Foundations of CCA and Equity Valuation*, pages 3–21. World Scientific. (Cited on page 57.)

[22] Blundell, R. W., Dias, M. C., Cribb, J., Joyce, R., Waters, T., Wernham, T., and Xu, X. (2022). Inequality and the covid crisis in the United Kingdom. Technical report, IFS Working Paper. (Cited on page 1.)

[23] Bouchaud, J.-P. and Mézard, M. (2000). Wealth condensation in a simple model of economy. *Physica A: Statistical Mechanics and its Applications*, 282(3-4):536–545. (Cited on pages 2, 37, 55, 57, and 84.)

[24] Chakrabarti, B. K., Chakraborti, A., Chakravarty, S. R., and Chatterjee, A. (2013). *Econophysics of income and wealth distributions*. Cambridge University Press. (Cited on page 41.)

[25] Champernowne, D. G. (1953). A model of income distribution. *The Economic Journal, Vol. 63, No. 250*. (Cited on pages 37, 38, and 49.)

[26] Champernowne, D. G. (1974). A comparison of measures of inequality of income distribution. *The Economic Journal , Dec., 1974, Vol. 84, No. 336, pp. 787-816*. (Cited on page 32.)

[27] Chatterjee, A., Chakrabarti, B. K., and Manna, S. (2004). Pareto law in a kinetic model of market with random saving propensity. *Physica A: Statistical Mechanics and its Applications*, 335(1-2):155–163. (Cited on page 43.)

[28] Chen, C.-N., Tsaur, T.-W., and Rhai, T.-S. (1982). The Gini coefficient and negative income. *Oxford Economic Papers*, 34(3):473–478. (Cited on page 32.)

[29] Cirillo, R. (2012). *The Economics of Vilfredo Pareto*. Routledge. (Cited on pages 10 and 11.)

[30] Clark, R. M., Cox, S., and Laslett, G. M. (1999). Generalizations of power-law distributions applicable to sampled fault-trace lengths: model choice, parameter estimation and caveats. *Geophysical Journal International*, 136(2):357–372. (Cited on page 35.)

[31] Clauset, A., Shalizi, C. R., and Newman, M. E. (2009). Power-law distributions in empirical data. *SIAM review*, 51(4):661–703. (Cited on pages 11, 21, 23, 24, and 30.)

[32] Clementi, F. and Gallegati, M. (2016). *Distribution of Income and Wealth*. Springer. (Cited on pages 17 and 101.)

[33] Clementi, F., Gallegati, M., and Kaniadakis, G. (2009). A κ-generalized statistical mechanics approach to income analysis. *Journal of Statistical*

Mechanics: Theory and Experiment, 2009(02):P02037. (Cited on pages 5, 41, and 61.)

[34] Clementi, F., Gallegati, M., Kaniadakis, G., and Landini, S. (2016). κ-generalized models of income and wealth distributions: A survey. *The European Physical Journal Special Topics*, 225(10):1959–1984. (Cited on pages 13, 16, and 18.)

[35] Collamore, J. F. and Vidyashankar, A. N. (2013). Tail estimates for stochastic fixed point equations via nonlinear renewal theory. *Stochastic Processes and their Applications*, 123(9):3378–3429. (Cited on page 54.)

[36] Cormen, T. H., Leiserson, C. E., Rivest, R. L., and Stein, C. (2022). *Introduction to algorithms*. MIT press. (Cited on page 129.)

[37] Cowell, F. A. (1988). Inequality decomposition: three bad measures. *Bulletin of Economic Research*, 40(4):309–312. (Cited on page 32.)

[38] Cox, J. (1975). Notes on option pricing i: Constant elasticity of variance diffusions. *Unpublished note, Stanford University, Graduate School of Business*. (Cited on page 58.)

[39] Cramer, C. (2003). Does inequality cause conflict? *Journal of International Development: The Journal of the Development Studies Association*, 15(4):397–412. (Cited on page 32.)

[40] Daffin, C. (2009). Wealth in Great Britain: main results from the wealth and assets survey 2006/08. *Newport: Office for National Statistics*. (Cited on pages 4 and 5.)

[41] D'Agostino, R. B. and Stevens, M. A. (1987). *Goodness-of-fit Techniques*. New York: Marcel Dekker. (Cited on page 7.)

[42] Dagsvik, J. K. and Vatne, B. H. (1999). Is the distribution of income compatible with a stable distribution? Technical report, Discussion Papers. (Cited on page 17.)

[43] Dagum, C. (1990). A model of net wealth distribution specified for negative, null and positive wealth. a case study: Italy. In *Income and wealth distribution, inequality and poverty*, pages 42–56. Springer. (Cited on pages 13 and 18.)

[44] Dagum, C. and Zenga, M. (2012). *Income and Wealth Distribution, Inequality and Poverty: Proceedings of the Second International Conference on Income*

Distribution by Size: Generation, Distribution, Measurement and Applications, Held at the University of Pavia, Italy, September 28–30, 1989. Springer Science & Business Media. (Cited on pages 17 and 32.)

[45] Desjardins, J. (2018). Visual Capitalist. Federal reserve survey of consumer finances (2016), `https://www.visualcapitalist.com/chart-assets-make-wealth/`. (Cited on page 98.)

[46] Diamond, J. et al. (1987). The worst mistake in the history of the human race. *Discover*, 8(5):64–66. (Cited on page 32.)

[47] Dorfman, R. (1979). A formula for the Gini coefficient. *The review of economics and statistics*, pages 146–149. (Cited on page 33.)

[48] Dragulescu, A. and Yakovenko, V. M. (2000). Statistical mechanics of money. *The European Physical Journal B-Condensed Matter and Complex Systems*, 17(4):723–729. (Cited on page 41.)

[49] Drinea, E., Frieze, A., and Mitzenmacher, M. (2002). Balls and bins models with feedback. In *SODA*, volume 2, pages 308–315. Citeseer. (Cited on pages 37, 63, and 64.)

[50] Ederer, S., Mayerhofer, M., and Rehm, M. (2020). Rich and ever richer? differential returns across socioeconomic groups. *Journal of Post Keynesian Economics*, pages 1–19. (Cited on pages 83, 88, and 98.)

[51] Emanuel, D. C. and MacBeth, J. D. (1982). Further results on the constant elasticity of variance call option pricing model. *Journal of Financial and Quantitative Analysis*, pages 533–554. (Cited on page 58.)

[52] Eskew, H. L. and Lawler, K. S. (1994). Correct and incorrect error specifications in statistical cost models. *The Journal of Cost Analysis*, 11(1):105–123. (Cited on pages 21 and 22.)

[53] Evans, A. J. E. and Baxendale, T. (2012). Heterogeneous entrepreneurship and the sunday times rich list. *REVISTA PROCESOS DE MERCADO*, pages 411–428. (Cited on page 5.)

[54] Fagereng, A., Guiso, L., Malacrino, D., and Pistaferri, L. (2020). Heterogeneity and persistence in returns to wealth. *Econometrica*, 88(1):115–170. (Cited on pages 83 and 88.)

[55] Faloutsos, M., Faloutsos, P., and Faloutsos, C. (2011). On power-law relationships of the internet topology. In *The Structure and Dynamics of Networks*, pages 195–206. Princeton University Press. (Cited on page 11.)

[56] Feller, W. (1968). *An Introduction to Probability Theory and Its Applications*, volume 1. Wiley, 3rd edition. (Cited on page 70.)

[57] Feller, W. (1971). *An Introduction to Probability Theory and Its Applications*, volume 2. Wiley, 2nd edition. (Cited on page 48.)

[58] Fix, B. (2018). Hierarchy and the power-law income distribution tail. *Journal of Computational Social Science*, 1(2):471–491. (Cited on pages 38 and 39.)

[59] Forbes, S. Author's repository of saved data. `https://github.com/saf92/PLOS-ONE-Kesten`. (Cited on page 136.)

[60] Forbes, S. and Grosskinsky, S. (2022). A study of UK household wealth through empirical analysis and a non-linear kesten process. *PloS one*, 17(8):e0272864. (Cited on pages v, 1, and 82.)

[61] Foss, S., Korshunov, D., Zachary, S., et al. (2011). *An introduction to heavy-tailed and subexponential distributions*, volume 6. Springer. (Cited on pages 7, 8, and 9.)

[62] Furstenberg, H. and Kesten, H. (1960). Products of random matrices. *The Annals of Mathematical Statistics*, 31(2):457–469. (Cited on page 48.)

[63] Gabaix, X. (2009). Power laws in economics and finance. *Annu. Rev. Econ.*, 1(1):255–294. (Cited on pages 1 and 51.)

[64] Gabaix, X. (2016). Power laws in economics: An introduction. *Journal of Economic Perspectives*, 30(1):185–206. (Cited on pages 9, 11, and 36.)

[65] Gardiner, C. W. et al. (1985). *Handbook of stochastic methods*, volume 3. Springer Berlin. (Cited on pages 54, 55, and 65.)

[66] Georgiou, N., Kiss, I. Z., and Scalas, E. (2015). Solvable non-Markovian dynamic network. *Physical Review E*, 92(4):042801. (Cited on page 17.)

[67] Gere, I., Kelemen, S., Toth, G., Biró, T. S., and Neda, Z. (2021). Wealth distribution in modern societies: collected data and a master equation approach. *Physica A: Statistical Mechanics and its Applications*, page 126194. (Cited on page 57.)

[68] Gibrat, R. (1931). *Les Inégalités Économiques*. Paris: Libr. Recl. Sirey. (Cited on pages 37 and 49.)

[69] Goldie, C. M. and Klüppelberg, C. (1998). Subexponential distributions. *A practical guide to heavy tails: statistical techniques and applications*, pages 435–459. (Cited on pages 8 and 85.)

[70] Gontis, V., Alaburda, M., et al. (2011). Interplay between positive feedbacks in the generalized cev process. *Physica A: Statistical Mechanics and its Applications*, 390(8):1393–1401. (Cited on pages 58, 59, and 60.)

[71] Grusky, D. and Hill, J. (2018). *Inequality in the 21st century: A reader*. Routledge. (Cited on page 32.)

[72] Hickey, J. and Davidsen, J. (2019). Self-organization and time-stability of social hierarchies. *PloS one*, 14(1):e0211403. (Cited on page 38.)

[73] Hill, B. M. (1975). A simple general approach to inference about the tail of a distribution. *The annals of statistics*, pages 1163–1174. (Cited on page 25.)

[74] Hitczenko, P., Wesołowski, J., et al. (2011). Renorming divergent perpetuities. *Bernoulli*, 17(3):880–894. (Cited on pages 52, 53, 83, and 84.)

[75] Hodges, J. L. (1958). The significance probability of the Smirnov two-sample test. *Arkiv för Matematik*, 3(5):469–486. (Cited on page 29.)

[76] Hogg, R. V., McKean, J., and Craig, A. T. (2005). *Introduction to mathematical statistics*. Pearson Education. (Cited on page 24.)

[77] Hong, L. (2012). A remark on the alternative expectation formula. *The American Statistician*, 66(4):232–233. (Cited on page 14.)

[78] Inoua, S. (2021). Beware the Gini index! a new inequality measure. *arXiv preprint arXiv:2110.01741*. (Cited on pages 32 and 33.)

[79] Ispolatov, S., Krapivsky, P. L., and Redner, S. (1998). Wealth distributions in asset exchange models. *The European Physical Journal B-Condensed Matter and Complex Systems*, 2(2):267–276. (Cited on pages 42 and 43.)

[80] Jefferies, J. (2005). The UK population: past, present and future. In *Focus on people and migration*, pages 1–17. Springer. (Cited on page 4.)

[81] Jessen, H. A. and Mikosch, T. (2006). Regularly varying functions. *Publications de L'institut Mathematique*, 80(94):171–192. (Cited on pages 8 and 9.)

[82] Jordà, Ò., Knoll, K., Kuvshinov, D., Schularick, M., and Taylor, A. M. (2019). The rate of return on everything, 1870–2015. *The Quarterly Journal of Economics*, 134(3):1225–1298. (Cited on pages 4, 138, and 139.)

[83] Kaniadakis, G. (2001). Non-linear kinetics underlying generalized statistics. *Physica A: Statistical mechanics and its applications*, 296(3-4):405–425. (Cited on page 16.)

[84] Karatzas, I. and Ruf, J. (2016). Distribution of the time to explosion for one-dimensional diffusions. *Probability Theory and Related Fields*, 164(3):1027–1069. (Cited on page 58.)

[85] Karlin, S. and Taylor, H. (1975). A first course in stochastic processes, new york: Academic. (Cited on page 64.)

[86] Keister, L. A. and Lee, H. Y. (2017). The double one percent: Identifying an elite and a super-elite using the joint distribution of income and net worth. *Research in Social Stratification and Mobility*, 50:1–12. (Cited on page 91.)

[87] Kemp, S. (2016). *Was Communism Doomed?: Human Nature, Psychology and the Communist Economy.* Springer. (Cited on page 32.)

[88] Kesten, H. (1973). Random difference equations and renewal theory for products of random matrices. *Acta Mathematica 131, 207-248.* (Cited on pages 3, 37, 51, 52, 82, 83, and 102.)

[89] Kesten, H. and Spitzer, F. (1984). Convergence in distribution of products of random matrices. *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete*, 67(4):363–386. (Cited on page 48.)

[90] Khanin, K. and Khanin, R. (2001). A probabilistic model for the establishment of neuron polarity. *Journal of Mathematical Biology*, 42(1):26–40. (Cited on pages 45, 46, 64, and 69.)

[91] Klein, L. R. (1983). Some laws of economics. *Bulletin of the American Academy of Arts and Sciences*, 36(4):23–45. (Cited on page 10.)

[92] Kononovicius, A. and Ruseckas, J. (2015). Nonlinear GARCH model and 1/f noise. *Physica A: Statistical Mechanics and its Applications*, 427:74–81. (Cited on page 59.)

[93] Korom, P., Lutter, M., and Beckert, J. (2017). The enduring importance of family wealth: Evidence from the Forbes 400, 1982 to 2013. *Social science research*, 65:75–95. (Cited on page 5.)

[94] Kumamoto, S.-I. and Kamihigashi, T. (2018). Power laws in stochastic processes for social phenomena: An introductory review. *Frontiers in Physics*, 6:20. (Cited on pages 36, 55, 57, and 59.)

[95] Levy, M. and Soloman, S. (1998). Of wealth power and law: the origin of scaling in economics. *Racah Institute of Physics, Hebrew University*. (Cited on page 49.)

[96] Linetsky, V. and Mendoza, R. (2010). Constant elasticity of variance (cev) diffusion model. *Encyclopedia of Quantitative Finance*. (Cited on page 58.)

[97] Lydall, H. F. and Tipping, D. (1961). The distribution of personal wealth in Britain. *Bulletin of the Oxford University Institute of Economics & Statistics*, 23(1):83–104. (Cited on pages 5 and 6.)

[98] Malevergne, Y., Pisarenko, V., and Sornette, D. (2011). Testing the Pareto against the lognormal distributions with the uniformly most powerful unbiased test applied to the distribution of cities. *Physical Review E*, 83(3):036111. (Cited on pages 9 and 20.)

[99] Mandelbrot, B. (1960). The Pareto-Lévy law and the distribution of income. *International Economic Review, Vol. 1, No. 2 (May, 1960), pp. 79-106*. (Cited on page 10.)

[100] Mandelbrot, B. (1963). New methods in statistical economics. *Journal of political economy*, 71(5):421–440. (Cited on page 9.)

[101] Mandelbrot, B. B. and Hudson, R. L. (2010). *The (mis) behaviour of markets: a fractal view of risk, ruin and reward*. Profile books. (Cited on page 5.)

[102] McDonald, J. and Snooks, G. D. (1985). Statistical analysis of domesday book (1086). *Journal of the Royal Statistical Society: Series A (General)*, 148(2):147–160. (Cited on page 5.)

[103] McLaughlin, E., Hanley, N., Greasley, D., Kunnas, J., Oxley, L., and Warde, P. (2014). Historical wealth accounts for Britain: progress and puzzles in measuring the sustainability of economic growth. *Oxford Review of Economic Policy*, 30(1):44–69. (Cited on page 4.)

[104] Miotto, J. M., Kantz, H., and Altmann, E. G. (2017). Stochastic dynamics and the predictability of big hits in online videos. *Physical Review E*, 95(3):032311. (Cited on pages 11 and 54.)

[105] Mohtadi, H. and Ruediger, S. (2011). The heavy tail in finance: a survey. *forthcoming in Econometrics: New Developments (Nova Science Publications).* (Cited on page 58.)

[106] Montgomery, D. C., Peck, E. A., and Vining, G. G. (2021). *Introduction to linear regression analysis*. John Wiley & Sons. (Cited on page 21.)

[107] Moore, M. and Tambini, D. (2018). *Digital dominance: the power of Google, Amazon, Facebook, and Apple*. Oxford University Press. (Cited on page 47.)

[108] Murphy, R. (2016). *The joy of tax*. Random House. (Cited on page 5.)

[109] Nair, J., Wierman, A., and Zwart, B. (2022). *The Fundamentals of Heavy Tails*, volume 53. Cambridge University Press. (Cited on pages 8 and 9.)

[110] Newman, M. C. (1993). Regression analysis of log-transformed data: Statistical bias and its correction. *Environmental Toxicology and Chemistry: An International Journal*, 12(6):1129–1133. (Cited on pages 21 and 22.)

[111] Newman, M. E. (2005). Power laws, Pareto distributions and Zipf's law. *Contemporary physics*, 46(5):323–351. (Cited on pages 1, 30, and 36.)

[112] Oksendal, B. (2013). *Stochastic differential equations: an introduction with applications*. Springer Science & Business Media. (Cited on page 54.)

[113] Oliveira, R. (2004). *Preferential attachment (Phd Thesis)*. New York University. (Cited on pages 45, 46, 63, and 66.)

[114] Oliveira, R. I. (2009). The onset of dominance in balls-in-bins processes with feedback. *Random Structures & Algorithms*, 34(4):454–477. (Cited on pages 2, 45, 46, 47, 63, 69, and 74.)

[115] ONS (2009). Main results from the wealth and assets survey 2006/08. (Cited on pages 6 and 136.)

[116] ONS (2018). Wealth and assets survey's. See previous versions for granulated data., `https://www.ons.gov.uk/peoplepopulationandcommunity/personalandhouseholdfinances/incomeandwealth/datasets/totalwealthwealthingreatbritain/july2006tojune2016`. (Cited on pages 12, 15, 17, 18, 19, 20, 25,

26, 27, 28, 29, 30, 33, 34, 89, 90, 117, 120, 121, 122, 123, 124, 125, 127, 136, 138, and 139.)

[117] Pareto, V. (1896). *Cours d'économie politique: professé à l'Université de Lausanne*, volume 1. F. Rouge. (Cited on page 5.)

[118] Pemantle, R. (2007). A survey of random processes with reinforcement. *Probability surveys*, 4:1–79. (Cited on pages 2, 37, and 45.)

[119] Perret, S. (2020). Why did other wealth taxes fail and is this time different. *Wealth Tax Commission. Evidence Paper*, 6. (Cited on page 1.)

[120] Peters, O. (2019). The ergodicity problem in economics. *Nature Physics*, 15(12):1216–1221. (Cited on page 99.)

[121] Piketty, T. (2013). *Capital in the 21st Century*. Harvard University Press. (Cited on pages 1, 32, and 101.)

[122] Redner, S. (1990). Random multiplicative processes: an elementary tutorial. *American Journal of Physics*, 58(3):267–273. (Cited on page 50.)

[123] Reed, W. J. and Hughes, B. D. (2002). From gene families and genera to incomes and internet file sizes: Why power laws are so common in nature. *Physical Review E*, 66(6):067103. (Cited on page 36.)

[124] Reed, W. J. and Jorgensen, M. (2004). The double pareto-lognormal distributiona new parametric model for size distributions. *Communications in Statistics-Theory and Methods*, 33(8):1733–1753. (Cited on page 17.)

[125] Rockoff, H. (2008). Great fortunes of the gilded age. Technical report, National Bureau of Economic Research. (Cited on page 47.)

[126] Rudin, W. et al. (1976). *Principles of mathematical analysis*, volume 3. McGraw-hill New York. (Cited on page 66.)

[127] Saez, E. and Zucman, G. (2020). The rise of income and wealth inequality in america: Evidence from distributional macroeconomic accounts. *Journal of Economic Perspectives*, 34(4):3–26. (Cited on pages 4 and 33.)

[128] Sahlins, M. and Graeber, D. (2017). *On kings*. Hau Books. (Cited on page 32.)

[129] Sandford, C. (1980). The diamond commission and the redistribution of wealth. *Brit. JL & Soc'y*, 7:286. (Cited on page 5.)

[130] Santos, M., Coelho, R., Hegyi, G., Néda, Z., and Ramasco, J. (2007). Wealth distribution in modern and medieval societies. *The European Physical Journal Special Topics*, 143(1):81–85. (Cited on page 11.)

[131] Scalas, E., Garibaldi, U., and Donadio, S. (2006). Statistical equilibrium in simple exchange games i. *The European Physical Journal B-Condensed Matter and Complex Systems*, 53(2):267–272. (Cited on pages 42 and 43.)

[132] Seber, G. A. F. and Wild, C. J. *Non-Linear Regression.* (Cited on pages 14 and 23.)

[133] Seneta, E. (1976). *Regularly varying functions*, volume 508. Springer. (Cited on page 8.)

[134] Simkin, M. V. and Roychowdhury, V. P. (2011). Re-inventing willis. *Physics Reports*, 502(1):1–35. (Cited on page 36.)

[135] Simon, H. A. (1957). The compensation of executives. *Sociometry*, 20(1):32–35. (Cited on pages 36 and 38.)

[136] Smirnoff, N. (1939). Sur les écarts de la courbe de distribution empirique. *Matematicheskii Sbornik*, 48(1):3–26. (Cited on page 29.)

[137] Sornette, D. and Cont, R. (1997). Convergent multiplicative processes repelled from zero: power laws and truncated power laws. *Journal de Physique I*, 7(3):431–444. (Cited on page 51.)

[138] Stanley, H. E. and Mantegna, R. N. (2000). *An introduction to econophysics.* Cambridge University Press, Cambridge. (Cited on page 41.)

[139] Stirzaker, D. and Grimmett, G. (1992). Probability and random processes. (Cited on pages 65 and 68.)

[140] Tang, D. (2019). Efficient algorithms for modifying and sampling from a categorical distribution. *arXiv preprint arXiv:1906.11700.* (Cited on pages 128 and 131.)

[141] Tsionas, E. G. (2002). Bayesian inference in the noncentral Student-t model. *Journal of Computational and Graphical Statistics.* (Cited on page 139.)

[142] Vallejos, H. A., Nutaro, J. J., and Perumalla, K. S. (2018). An agent-based model of the observed distribution of wealth in the United States. *Journal of Economic Interaction and Coordination*, 13(3):641–656. (Cited on pages 36, 46, 63, 76, 77, and 81.)

[143] Van der Vaart, A. W. (2000). *Asymptotic statistics*, volume 3. Cambridge university press. (Cited on page 7.)

[144] Van Kampen, N. G. (1992). *Stochastic processes in physics and chemistry*, volume 1. Elsevier. (Cited on page 65.)

[145] Vermeulen, P. (2018). How fat is the top tail of the wealth distribution? *Review of Income and Wealth*, 64(2):357–387. (Cited on pages 6, 27, 30, 31, 85, 100, and 101.)

[146] Yakovenko, V. M. and Rosser Jr, J. B. (2009). Colloquium: Statistical mechanics of money, wealth, and income. *Reviews of modern physics*, 81(4):1703. (Cited on pages 2, 5, 16, 37, 41, 42, and 43.)

[147] Yule, G. U. (1925). A mathematical theory of evolution, based on the conclusions of Dr. J. C. Willis, F.R.S. *Philosophical Transactions of the Royal Society of London. Series B, Containing Papers of a Biological Character, Vol. 213 (1925), pp. 21-87.* (Cited on page 49.)

[148] Zeng, W. S. and Tang, S. Z. (2011). Bias correction in logarithmic regression and comparison with weighted regression for nonlinear models. *Nature Precedings*, pages 1–1. (Cited on page 22.)

[149] Zhu, T. (2009). Nonlinear Pólya urn models and self-organizing processes. *University of Pennsylvania Dissertation*. (Cited on page 63.)

Appendix A

Fitting the Tail of UK Wealth

This appendix concerns results from Chapter 1.

A.1 Remarks on the Methods of Fitting

To fit the wealth distribution we have used several methods in Chapter 1. For the exponential and lognormal fits to positive wealth we used the method of estimating parameters with the mean and median respectively. For the remaining lognormal parameter and the κ-gen parameter fits to positive wealth we used NLLS on the empirical tail. We also use NLLS on the empirical tail for the adapted Weibull distribution to negative wealth and as one of the methods to fit the power law to the top 10% WAS data richest. As noted in Section 1.4.6 NLLS could not fit the rich lists due to the numerical algorithm not optimising. This could be due to irregularities in the rich lists and their small sizes. The remaining methods we use to fit the power law are OLS and MLE. We note that both OLS and MLE have analytical solutions however this is not the case in general for NLLS. Thus for NLLS numerical methods are required and we implement this with the Python least_squares package from the scipy optimize module. We run the least_squares with method set to Levenberg-Marquardt which is the standard method setting. When using the algorithm it is important to choose appropriate initial conditions otherwise the algorithm does not converge to a solution.

In this appendix we provide parameter estimates for fits to the empirical tail of the wealth distribution using WAS data [116] and Forbes rich list data [1]. We refer to the empirical tail by the N points $(w_i, \mathbb{P}_N(W > w_i))_{i=1}^{N}$. Throughout a parameter k is estimated by $\hat{k}$.

A.2 Positive Wealth Tail Fits

Here we fit positive wealth. Technically we sometimes fit non-negative wealth but we don't emphasise the distinction.

Exponential

The mean for the positive wealth distribution can be estimated using the following alternative expectation formula

$$\mu_+ := \mathbb{E}[W_+] = \int_0^\infty \mathbb{P}(W_+ > w)\mathrm{d}w.$$

We estimate μ_+ from the data using numerical integration[1]. An exponential distribution $W_+ \sim \mathrm{Exp}(1/\mu_+)$ has density

$$f_{W^+}(w) = 1/\mu_+ \exp\left(-\frac{1}{\mu_+}w\right)$$

for w, $\mu_+ > 0$ with mean and variance equal to μ_+ and μ_+^2 respectively. Then the tail of the exponential distribution for positive wealth $w > 0$ is

$$\mathbb{P}(W > w) = \mathbb{P}(W \geq 0) \exp\left(-\frac{1}{\mu_+}w\right).$$

Table A.1 shows estimates from the empirical tail WAS data for $\mathbb{P}(W \geq 0)$ and μ_+ and the cost (1.17) of the fits for all $w_i > 0$. $\mathbb{P}(W \geq 0)$ is estimated by $\mathbb{P}(W \geq \tilde{w}_0)$ where $\tilde{w}_0 \geq 0$ is the smallest data point greater than or equal to 0.

Year	$\hat{\mathbb{P}}(W \geq 0)$	$\hat{\mu}_+$	Cost
2008	0.83	336276	11.15
2010	0.85	356144	8.43
2012	0.84	382654	10.58
2014	0.83	424329	6.87
2016	0.84	485223	5.55

Table A.1: Estimates of $\mathbb{P}(W \geq 0)$ (2 decimal places) and positive wealth exponential empirical tail fits of μ_+ (nearest whole number) with the cost (1.17) (2 decimal places) to WAS data for biennial years 2008-2016.

[1] Specifically using the trapezoidal rule using Python's Scipy integrate.trapezoid module.

Lognormal

The density of a lognormal distribution $W_+ \sim \text{Lognorm}(k, s)$ is

$$f_{W_+}(w) = \frac{1}{ws\sqrt{\pi}} \exp\left(-\frac{(\log w - k)^2}{2s^2}\right)$$

for $w, s > 0$ and $k \in (-\infty, \infty)$. We can find the median from the tail as follows

$$m_+ := \text{Med}[W_+] = w \text{ s.t. } \mathbb{P}(W > w) = \frac{\mathbb{P}(W \geq 0)}{2}.$$

We estimate the median $\hat{m}_+$ from the data by using $\hat{\mathbb{P}}(W \geq 0)$ from Table A.1 and then finding the closest tail value by absolute value to $\hat{\mathbb{P}}(W \geq 0)/2$. Thus

$$\hat{m}_+ = \arg\min_{w>0} \left| \mathbb{P}(W > w) - \frac{\hat{\mathbb{P}}(W \geq 0)}{2} \right|$$

Then for a lognormal it can be shown that $k = \log m_+$. We can estimate the wealth distribution with lognormal tail for $w > 0$ as

$$\mathbb{P}(W > w) = \mathbb{P}(W \geq 0)\mathbb{P}_{\text{Lognorm}(k,s)}(W > w)$$

We estimate the remaining parameter $\hat{s}$ by NLLS (1.16). Table A.2 shows these parameter fits along with the cost (1.17) of the fits for $w_i > 0$ for WAS data.

Year	$\hat{m}_+$	$\hat{s}$	Cost
2008	257057	1.01	5.63
2010	271638	1.04	9.69
2012	292347	1.05	7.83
2014	289257	1.15	3.25
2016	328702	1.16	3.21

Table A.2: Positive wealth lognormal empirical tail fits to positive WAS data of m_+ (nearest whole number) and s (2 decimal places) with cost (1.17) (2 decimal places) to WAS data for biennial years 2008-2016.

κ-generalised

With the generalised exponential function $\exp_\kappa(w)$:

$$\exp_\kappa(w) = (\sqrt{1 + \kappa^2 w^2} + \kappa w)^{1/\kappa}$$

the κ-generated (κ-gen) distribution has the following density

$$f_{\kappa\text{-gen}}(w) = \frac{\alpha}{\beta}\left(\frac{w}{\beta}\right)^{\alpha-1}\frac{\exp_\kappa\left(-\left(\frac{w}{\beta}\right)^\alpha\right)}{\sqrt{1+\kappa^2\left(\frac{w}{\beta}\right)^{2\alpha}}}$$

with tail function

$$\mathbb{P}_{\kappa\text{-gen}}(W > w) = \exp_\kappa\left(-\left(\frac{w}{\beta}\right)^\alpha\right)$$

for $w \geq 0$, $\alpha, \beta > 0$ (here β is **not** the power law exponent) and $\kappa \in [0, 1)$. For positive wealth we want to estimate

$$\mathbb{P}(W > w) = \mathbb{P}(W \geq 0)\mathbb{P}_{\kappa\text{-gen}}(W > w)\,.$$

We fit $\mathbb{P}(W \geq 0)$ as previous with Table A.1 and the κ-gen parameters α, β and κ with NLLS (1.16). The κ-gen distribution is a regularly varying distribution (1.6) with

$$\mathbb{P}_{\kappa\text{-gen}}(W > w) \simeq \frac{1}{\beta(2\kappa)^{1/\kappa}}\frac{1}{w^{\alpha/\kappa}}.$$

Thus when fitting we can approximate the κ-gen power law exponent as

$$\hat{\beta}_p = \frac{\hat{\alpha}}{\hat{\kappa}} \tag{A.1}$$

where $\hat{\alpha}$ and $\hat{\kappa}$ are fits of the parameters α and β respectively. Table A.3 shows the fits for the κ-gen parameters along with the cost for WAS data for $w_i > 0$.

Year	$\hat{\alpha}$	$\hat{\beta}$	$\hat{\kappa}$	$\hat{\beta}_p = \hat{\alpha}/\hat{\kappa}$	Cost
2008	1.13	341405	0.50	2.27	1.26
2010	1.11	350901	0.51	2.19	1.93
2012	1.10	381672	0.52	2.14	2.13
2014	1.03	418226	0.50	2.07	1.87
2016	1.00	480765	0.44	2.24	1.34

Table A.3: Positive wealth κ-gen empirical tail fits to positive WAS data [116] for α (2 decimal places), β (nearest whole number), κ (2 decimal places), the regularly varying power law exponent β_p (2 decimal places) and cost (1.17) (2 decimal places) for biennial years 2008-2016.

Plots

Figure A.1 shows the empirical tail fits to biannual 2008-2016 WAS data for positive wealth with the exponential, lognormal and κ-gen distributions.

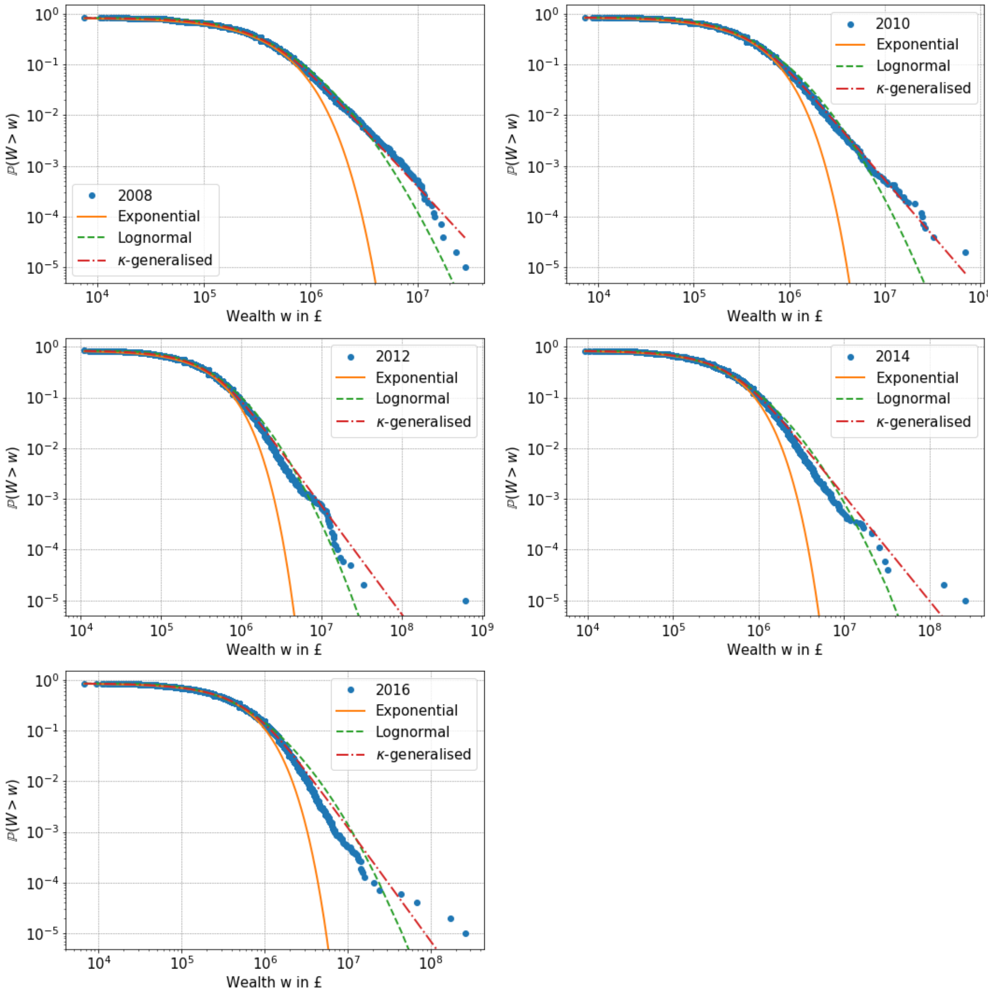

Figure A.1: Empirical tail distribution of positive UK household wealth for biennial years 2008-2016 with exponential, lognormal and κ-gen fits to WAS data [116]. Fits for exponential found in Table A.1, for lognormal in Table A.2 and for κ-gen in Table A.3 .

A.3 Debt Tail Fits

Weibull

The density of the Weibull is defined as

$$f_{\text{Weib}}(|w|) = \frac{s}{\lambda}\left(\frac{|w|}{\lambda}\right)^{s-1} \exp\left(-\left(\frac{|w|}{\lambda}\right)^{s}\right)$$

where we assume $w < 0$ and take $s, \lambda > 0$. Then for wealth $w < 0$ we have debt $|W|$ with the following tail

$$\mathbb{P}(|W| > |w|) = \mathbb{P}(W < 0) \exp\left(-\left(\frac{|w|}{\lambda}\right)^{s}\right).$$

We fit the parameter $\mathbb{P}(W < 0) = 1 - \mathbb{P}(W \geq 0)$ using the fit of $\mathbb{P}(W \geq 0)$ from Table A.1. For the remaining parameters s and λ we use NLLS (1.16). Table A.4 shows the empirical tail fits for the adapted Weibull distribution to WAS data and the cost (1.17) for debt.

Year	$\hat{\mathbb{P}}(W < 0)$	$\hat{s}$	$\hat{\lambda}$	Cost
2008	0.17	0.23	100	$1.6 \cdot 10^{-5}$
2010	0.15	0.34	516	$2.0 \cdot 10^{-6}$
2012	0.16	0.24	106	$4.7 \cdot 10^{-6}$
2014	0.17	0.31	439	$1.2 \cdot 10^{-6}$
2016	0.16	0.28	261	$4.2 \cdot 10^{-6}$

Table A.4: Debt adapted Weibull empirical tail fits for θ (2 decimal places), s (2 decimal places), λ (nearest whole number) and cost (1.17) (2 significant figures) for WAS biennial years 2008-2016 [116].

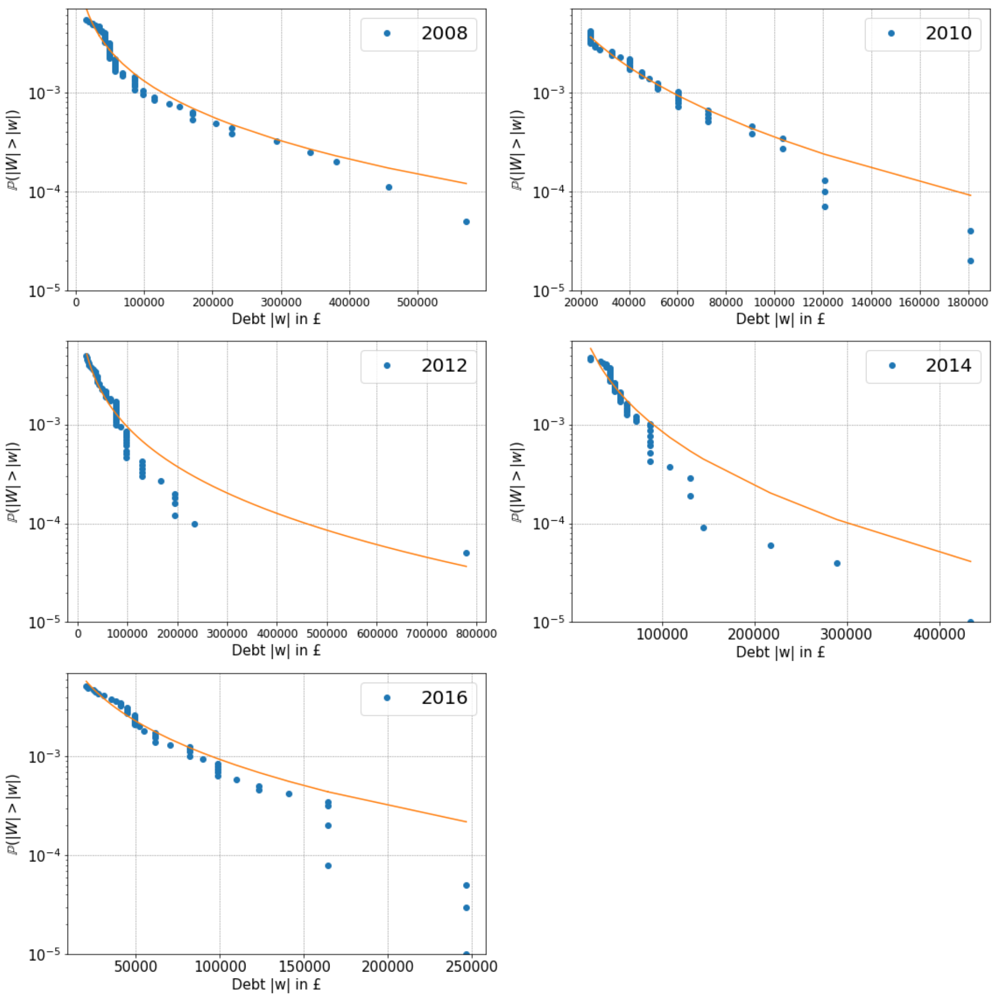

Figure A.2: Empirical tail distribution of UK household wealth debt from WAS data [116] for biennial years 2008-2016 with Weibull fit. Parameter fits found in Table A.4.

A.4 Survey Wealth Power Law Tail Fits

The power law for the positive tail is defined

$$\mathbb{P}(W > w) = \frac{\alpha}{w^{\beta}}, \quad \text{for } w > w_m > 0\,.$$

We choose w_m as the tail value for the top 10% of richest households:

$$w_m : \mathbb{P}(W > w_m) = 0.1\,.$$

We estimate w_m as

$$\hat{w}_m = \arg\min_w |\mathbb{P}(W > w) - 0.1| \ .$$

We fit the power law with three methods: OLS (where we do not correct for potential bias) (1.27), MLE (1.31), (1.33) and NLLS (1.30) for $w > w_m$ to the WAS biennial 2008-2016 data. We also, as discussed above, fit the κ-gen power law exponent (A.1). See Tables A.5 and A.6 for parameter fits and Figure A.3 for plots of the empirical tail fits for biennial top 10% richest WAS data 2008-2016.

Year	$\hat{w}_m$	$\hat{\alpha}_{\text{OLS}}$	$\hat{\alpha}_{\text{MLE}}$	$\hat{\alpha}_{\text{NLLS}}$
2008	783413	250265412470	64648898927	84931716366
2010	814914	579244022874	39911333772	81025881213
2012	909526	3604068491633	604161877226	362201752682
2014	1041327	6851243114176	608035578326	549962767634
2016	1205244	61327974512444	3105708814824	1593163542759

Table A.5: Power law fits for for WAS data biennial years 2008-2016 [116] for w_m and α (nearest whole number) with OLS (1.27), MLE (1.31), (1.33) and NLLS (1.30).

Year	$\hat{\beta}_{\text{OLS}}$	$\hat{\beta}_{\text{MLE}}$	$\hat{\beta}_{\text{NLLS}}$	$\hat{\beta}_{\kappa\text{-gen}}$	Cost_{OLS}	Cost_{MLE}	$\text{Cost}_{\text{NLLS}}$	$\text{Cost}_{\kappa\text{-gen}}$
2008	2.10	2.01	2.02	2.27	0.010	0.010	0.007	0.007
2010	2.15	1.98	2.01	2.19	0.023	0.037	0.016	0.015
2012	2.27	2.15	2.11	2.14	0.026	0.026	0.019	0.037
2014	2.29	2.13	2.11	2.07	0.031	0.038	0.021	0.054
2016	2.42	2.22	2.17	2.24	0.044	0.043	0.026	0.051

Table A.6: Power law fits for β (2 decimal places) for WAS data biennial years 2008-2016 [116] with OLS (1.27), MLE (1.31), (1.33) and NLLS (1.30) and the regularly varying power law exponent (A.1) for κ-gen as previously given in Table A.3 and costs (1.17) (3 decimal places).

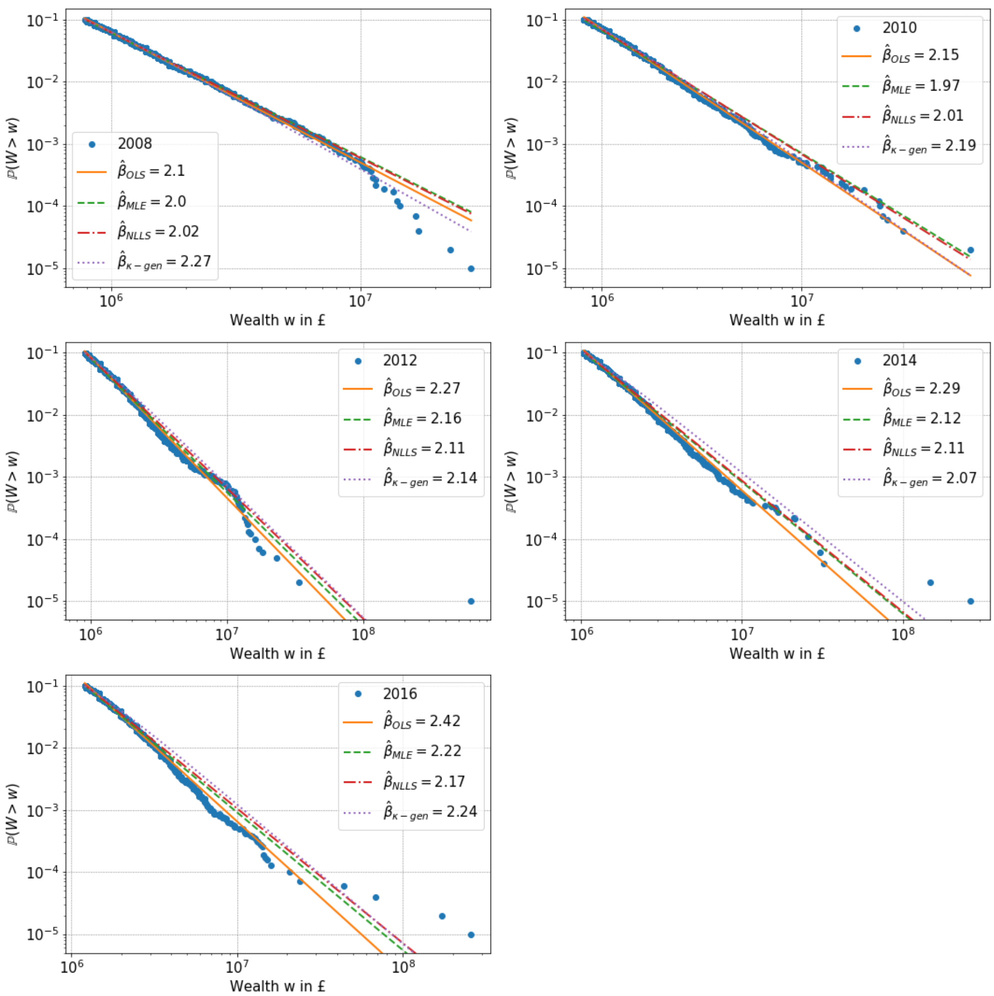

Figure A.3: Empirical tail distribution of positive UK household wealth greater than w_m (1.23) for WAS data biennial years 2008-2016 [116] with κ-gen fit and with power law fits using OLS (1.27), MLE (1.31), (1.33) and NLLS (1.30). Parameter fits found in Tables A.5 and A.6.

A.5 Survey Wealth and Forbes Rich List Power Law Empirical Tail Fits

Two power laws in the tail could be approximated with the following

$$\mathbb{P}(W > w) = \begin{cases} \dfrac{\alpha_1}{w^{\beta_1}} & \text{for } 0 < w_{1,m} < w \leq w_{2,m} \\ \dfrac{\alpha_2}{w^{\beta_2}} & \text{for } w > w_{2,m} \end{cases}$$

with all parameters in $\mathbb{R}_{>0}$ and $\beta_1 \neq \beta_2$ where we fit α_2 and β_2 to the rich lists. We note that we could not find a numerical NLLS solution to α_2 and β_2 due to the small sample size (under 60) and potential irregularities in the data. The parameter fits for $w_{1,m}$, α_1 and β_1 are the same as in Tables A.5 and A.6. The costs and parameter fits for the rich lists, $w_{2,m}$, α_2 and β_2, are seen in Tables A.7 and A.8 for OLS and MLE. We estimate $w_{2,m}$ as the smallest wealth value in the rich list. See Figure A.4 for the two power law fits for OLS and MLE for the WAS and rich lists for biennial years 2008 to 2016.

Year	$\hat{w}_{2,m}$	$\hat{\alpha}_{2,\mathrm{OLS}}$	$\hat{\alpha}_{2,\mathrm{MLE}}$
2008	439189189	3479000	83601
2010	430107527	528233	518398
2012	439253269	47162738	74779
2014	404040404	614937	4362
2016	603566529	616326	118734

Table A.7: Power Law Fits for $w_{2,m}$ and α_2 (nearest whole number) with OLS (1.27), MLE (1.31), (1.33) to Forbes rich list biennial 2008-2016 data [1].

Year	$\hat{\beta}_{2,\mathrm{OLS}}$	$\hat{\beta}_{2,\mathrm{MLE}}$	$\mathrm{Cost}_{2,\mathrm{OLS}}$	$\mathrm{Cost}_{2,\mathrm{MLE}}$
2008	1.44	1.27	$1.42 \cdot 10^{-13}$	$2.33 \cdot 10^{-13}$
2010	1.35	1.35	$1.48 \cdot 10^{-13}$	$1.80 \cdot 10^{-13}$
2012	1.55	1.24	$5.10 \cdot 10^{-14}$	$1.60 \cdot 10^{-13}$
2014	1.33	1.09	$1.55 \cdot 10^{-12}$	$8.69 \cdot 10^{-13}$
2016	1.30	1.23	$7.34 \cdot 10^{-13}$	$5.52 \cdot 10^{-13}$

Table A.8: Power Law Fits for β_2 (2 decimal places) with OLS (1.27), MLE (1.31), (1.33) and costs (1.17) (3 decimal places) to Forbes rich list biennial 2008-2016 data [1].

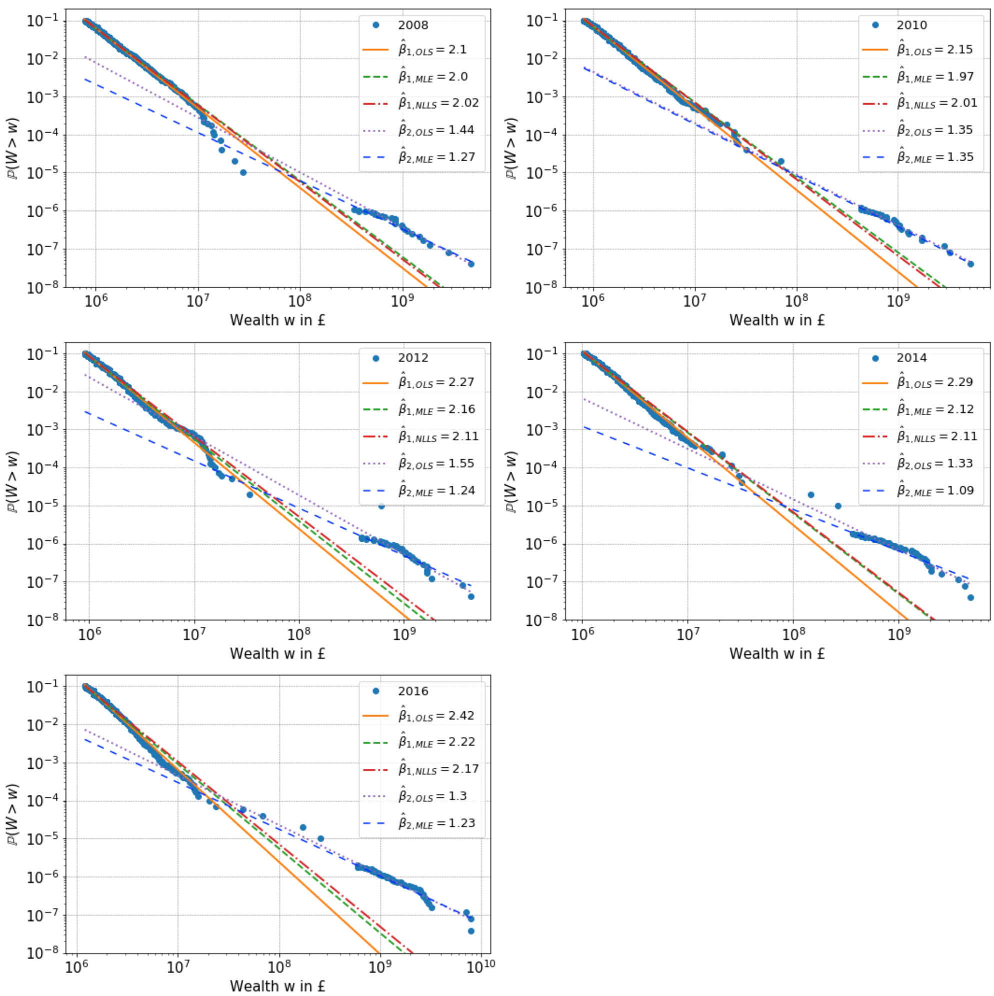

Figure A.4: Empirical tail distribution of positive UK household wealth greater than w_m (1.23) for biennial years 2008-2016 with power law fits using OLS (1.27), MLE (1.31), (1.33) and NLLS (1.30) to WAS [116] and OLS and MLE fits to Forbes rich lists [1]. Parameter fits found in Tables A.5, A.6, A.7 and A.8.

Appendix B

Binary Tree Search

In the balls in bins process with feedback in Chapter 3 we consider a collection of N agents with agent j having wealth $W_n(j)$ and rate $r_n(j) = f(W_n(j)) > 0$ at time $n \in \mathbb{N}$. At each time n a wealth packet $\omega_p > 0$ is given to an agent with the following probability update rule given $\mathbf{W}_n = \{W_n(1), W_n(2), \ldots, W_n(N)\}$:

$$\mathbb{P}(W_{n+1}(j) = W_n(j) + \omega_p | \mathbf{W}_n) = \frac{r_n(j)}{\sum_{i=1}^{N} r_n(i)} . \tag{B.1}$$

Agent	Rate	Cumulative Rate
1	$r_n(1)$	$r_n(1)$
2	$r_n(2)$	$r_n(1) + r_n(2)$
$\vdots$	$\vdots$	$\vdots$
j	$r_n(j)$	$\sum_{i=1}^{j} r_n(i)$
$\vdots$	$\vdots$	$\vdots$
N	$r_n(N)$	$R := \sum_{i=1}^{N} r_n(i)$

Table B.1: Rates

We want to code (B.1). We note that this is a more general problem of sampling from a categorical distribution (also called an empirical distribution) which is a common problem in probabilistic programming [140].

Define $R := \sum_{i=1}^{N} r_n(i)$. A simple algorithm to calculate which agent j should have its wealth $W_n(j)$ updated at time n by ω is to first uniformly at random choose a value p between 0 and R inclusive of 0 and exclusive of R. Then we iteratively go through the cumulative rates and find the first j such that $\sum_{i=1}^{j-1} r_n(i) \leq p < \sum_{i=1}^{j} r_n(i)$ for $j = 1, 2, \ldots, N$ where $\sum_{i=1}^{0} r_n(i) := 0$. The probability that p will lie in this interval will then be as in (B.1):

$$\frac{\sum_{i=1}^{j} r_n(i) - \sum_{i=1}^{j-1} r_n(i)}{R} = \frac{r_n(j)}{R}$$

This is illustrated in Algorithm 1.

Algorithm 1 Calculate which agent j to update (3.1) at time n

-Choose p uniformly at random between 0 and R inclusive of 0 and exclusive of R: $p \sim \mathcal{U}[0, R)$
-Let $j \leftarrow 1$, $u \leftarrow r_n(j)$
while $p > u$ **do**
  $j \leftarrow j + 1$, $u \leftarrow u + r_n(j)$
end while
$W_{n+1}(j) \leftarrow W_n(j) + \omega_p$ and update $r_{n+1}(j) = f(W_{n+1}(j))$

Algorithm 1 will take time complexity $\mathcal{O}(N)$ as there are N rates. Therefore if we run the update rule for $n_u \in \mathbb{N}$ iterations this will take

$$T_1(N, n_u) = \mathcal{O}(n_u N).$$

We can improve on this time complexity by using a binary tree search, see for example Chapter 12 of [36]. We make the assumption that there are $N = 2^M$ agents where $M \in \mathbb{Z}_{\geq 0}$. Figure B.1 illustrates the layout of the binary tree. Each node of the tree has a position and holds a value. The top node of the tree has position 0 and holds value 0, the node directly below has position 1 and holds value R. The position 1 node's value is then divided into two and the left node below has position 2 or 10 in binary and holds the first half of R: $\sum_{i=1}^{N/2} r_n(i)$ whilst the right node below has position 3 or 11 in binary and holds the second half of R: $\sum_{i=N/2+1}^{N} r_n(i)$. From here we repeat this process of dividing the node values in two until we reach the leaves at the bottom of the tree which hold the rates. Every time we go in position down and left from a node we shift the digits of the binary number of that node and add a 0 to get the new position and when we go down and right we do the same but instead add a 1. In C++ this is done using the shift operator <<. To go down from a node left we use the shift operator << only and to go down from a node right we use the shift operator and add 1: $(<<) + 1$.

There are $(M+1)+1 = M+2 = \log_2(N)+2$ layers of the binary tree and $\sum_{i=0}^{M} 2^i + 1 = 2^{M+1}$ nodes. We can see now that the position 0 node with value 0 is merely a dummy node so the number of nodes works out to a power of two for convenience and is not necessary. The rate of the j^{th} agent is in the tree position $\sum_{i=0}^{M-1} 2^i + 1 + j = 2^M + j = N + j$.

Code B.1 returns the binary tree of rates in vector form with the i^{th} place of the vector holding the i^{th} value of the tree after inputting the rates and M. It is built by first inputting the bottom layer of the tree and then building to the top of the tree by the new parent node above adding the two child nodes below.

Code B.2 returns the position in the tree and the corresponding agent to update (3.1) after inputting the binary tree vector outputted from Code B.1, M and $p \sim \mathcal{U}[0, R)$. It finds the position (poss) by moving down left or right in the binary tree and comparing p to the node in the bottom left from the current position (position k, value R_tree[k]) while adjusting p by taking away the comparison node value (R_tree[k]) if the position moves right. It ends at the final rates layer of the tree and outputs this tree position (pos) and the rates position (j) in a vector r which is just the tree position take away N as mentioned above.

Code B.3 returns the updated binary tree which will be used in the next iteration with δ added to the rate $r_{n+1}(j) = r_n(j) + \delta$ where $\delta = f(W_n(j)+\omega) - f(W_n(j))$. It inputs the binary tree outputted in Code B.1, M, the tree position outputted in Code B.2 and δ. First δ is added to the rate $r_n(j)$ with corresponding tree position and then by using the forward shift operator (>>) which knocks off the end bit from the binary tree position, δ is added to each parent going up the tree finishing with adding δ to the position 1 node with updated value $R + \delta$.

Algorithm 2 shows how to run the update rule (3.1) $n_u \in \mathbb{N}$ times with the binary tree search using Code B.1, B.2 and B.3. First the rates tree is initialised using Code B.1 which is $\mathcal{O}(N)$ time complexity then within the for loop Code B.2 and B.3 are utilised each of order $\mathcal{O}(M) = \mathcal{O}(\log_2(N))$. Thus the total time complexity for Algorithm 2 is

$$T_2(N, n_u) = \mathcal{O}(N) + \mathcal{O}(n_u \log_2(N))$$

which is a substantial time gain to Algorithm 1 for a large agent number N. We noted before that N is to be a power of two: $N = 2^M$. However if we wanted to have N_1 agents that is not a power of two we can introduce $N - N_1$ dummy agents with zero rates so we still in a sense have N agents. As zero rates does not effect the cumulative sums of rates only the N_1 agents will be updated with the update rule (3.1).

We note that a binary search tree is not the most efficient tree in order to sample from a categorical distribution. A Huffman tree is more efficient as it minimises the average number of branches searched instead of the maximum number of branches in the tree as in binary search [140]. Details of sampling from a categorical distribution as well as updating probabilities (3.1) are found in [140].

Algorithm 2 Algorithm to run update rule (3.1) n_u times using binary tree search

Initialise wealth vector $\mathbf{W} = [W_0(0), W_0(1), \ldots, W_0(N)]$, ω and δ
Initialise binary tree of rates ($r_0(i) = f(W_0(i))$) R_tree using Code B.1
for $i = 1, 2, \ldots, n_u$ **do**
  -Set $R =$R_tree[1]
  -Take $p \sim \mathcal{U}[0, R)$
  -Find tree position (poss) and rate position (j) using Code B.2
  -Set before $\leftarrow \mathbf{W}[j]$
  -$\mathbf{W}[j] \leftarrow \mathbf{W}[j] + \omega_p$
  -Set after $\leftarrow \mathbf{W}[j]$
  -Calculate $\delta = f(\text{after}) - f(\text{before})$
  -Update R_tree using Code B.3
end for

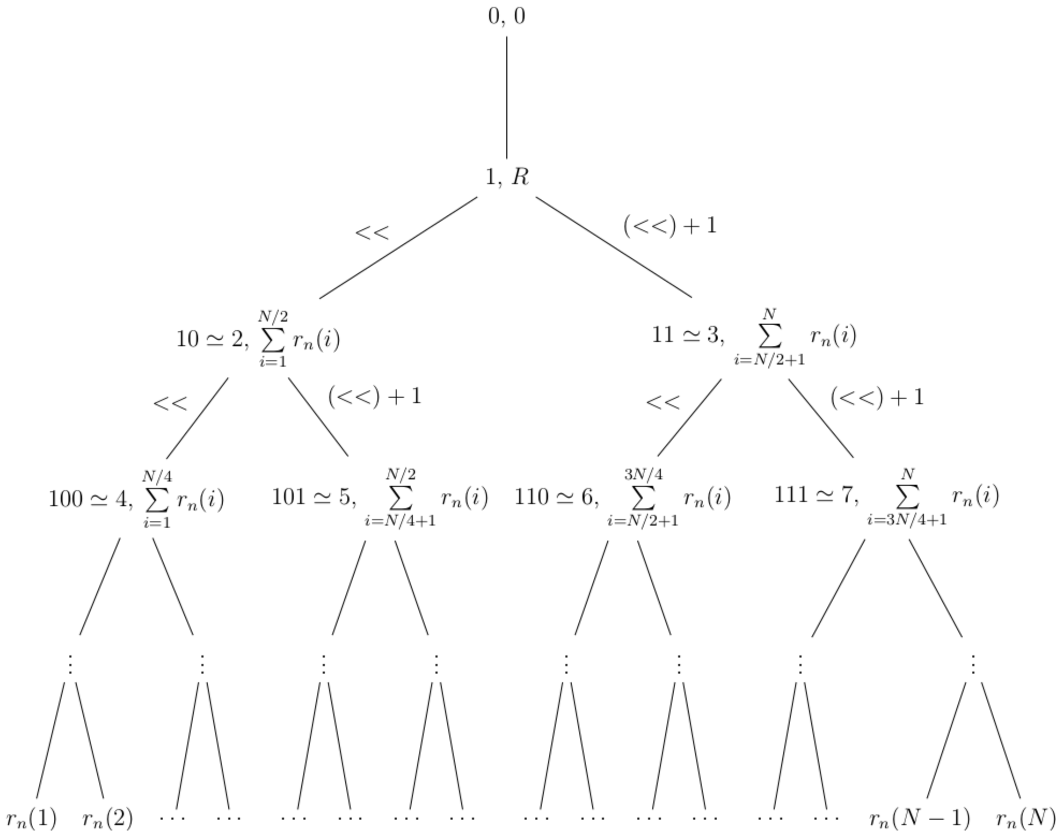

Figure B.1: Binary Search Tree. Rates are at the bottom layer. Written on the nodes are a, b where a is the node position and b is the node value. We also write the binary number a_2 for the positions a with the $\simeq$ indicating their equivalence: $a_2 \simeq a$.[1]

[1] We note here $\simeq$ has a different meaning to the asymptotic relation as in Chapter 1 .

Code B.1: C++ function to set up binary tree with rates, see Figure B.1, as a vector.

```
vector<double> r_tree(vector<double> rates, long int M){
double N=pow(2.0,M); // no. of agents
double T=pow(2.0,M+1); // size of tree
N= (long int) N;
T= (long int) T;
vector<double> R_tree(T);

for (long int i=0; i<N; i++){ // fill in bottom layer of tree
R_tree[N+i]=rates[i];
}

for(long int i=N-1; i>0; i--){ // remaining nodes filled with
// values of the two (child) nodes below
R_tree[i] = R_tree[(i<< 1)] + R_tree[(i<<1) +1];
}

return R_tree;
}
```

Code B.2: C++ function to find the position of rate in binary tree.

```
vector<long int> find_position(vector<double> R_tree, long int M,
double p){
double N=pow(2.0,M);
N= (long int) N;
long int pos=1; // tree position: start at node 1 with value R
long int k=2; // position of left child to parent at node at pos
double check = R_tree[k]; // value of left child to parent at pos
vector<long int> r;

for(long int i=1; i<=M; i++){
if(p<check){
pos=pos<<1; // go left in tree
if(k<N){ // stop at last layer to calculate check
k=k<<1;
check = R_tree[k];
}}
else{
p=p-check;
pos=(pos<<1)+1; // go right in tree
if(k<N){// stop at last layer to calculate check
k=k+1;
k=k<<1;
check = R_tree[k];
}}}

long int j=pos-N;
r.push_back(pos); // final tree position
r.push_back(j); // position of rate

return r;
}
```

Code B.3: C++ function to update the binary tree.

```
vector<double> r_tree_update(vector<double> R_tree, long int M,
long int pos, double delta){

for(long int i=1; i<=M+1; i++){
R_tree[pos] = R_tree[poss]+delta;
pos=pos>>1; // goes to position of parent node
}

return R_tree;
}
```

Appendix C

Theory and Empirics for Fitting the Non-Linear Kesten Model

This appendix is concerned with results in Chapter 4.

C.1 Data Sources

Here we list the data sources used in Chapter 4:

1. Biannual wealth and asset survey (WAS) data 2008-2016 from the Office for National Statistics (ONS) [116]

2. Forbes rich lists [1]

3. Times rich list data - extracted from Times online newspaper 2019, 2020 and 2021. See [59] for this data and for the current list see [4].

4. ONS household income, salary and expenditure data [3, 2]

Full details on the data can be found at the author's repository [59].

C.2 Tail of UK Wealth

Here we outline how we extract the empirical tail from wealth survey and rich list data from [116, 1]. For extensive discussion on the wealth and asset survey see [115]. We have wealth survey data in the form $(\tilde{h}_i, \tilde{w}_i)$ for $i = 1, 2, \dots n$ where $\tilde{h}_i \in [0, 1]$ is the cumulative proportion of households and $\tilde{w}_i \in [0, 1]$ is their corresponding cumulative proportion of wealth. Let us assume that the $\tilde{h}_i$ are ordered by increasingly wealthy households with positive wealth. The survey data was in this form to calculate the Gini coefficient from the Lorenz curve defined by the points $(\tilde{h}_i, \tilde{w}_i)$. Let H_T and W_T be the total number of households and the total amount of wealth of all households respectively.

Define $\hat{w}_i = (\tilde{w}_{i+1} - \tilde{w}_i)W_T$ and $\hat{h}_i = (\tilde{h}_{i+1} - \tilde{h}_i)H_T$ for $i = 1, 2, \ldots, n-1$. Then $\hat{w}_i$ is the amount of wealth owned by an increasingly rich $\hat{h}_i$ number of households. We have then that $w_i = \dfrac{\hat{w}_i}{\hat{h}_i}$ is the average amount of wealth of increasingly rich $\hat{h}_i$ households and thus w_i is ordered: $w_i \leq w_{i+1}$ for all i.

Therefore the points $(w_i, \tilde{h}_{i+1})$ characterise an approximation to the empirical CDF and the points $(w_i, 1 - \tilde{h}_{i+1})$ give the corresponding approximation to the empirical tail. We plot the approximate empirical tail of positive wealth in Figure 1.1 which are points below £10^8 for the years 2008, 2010, 2012, 2014 and 2016.

We have separate wealth data in the form of rich lists. For rich lists we have data of individuals households wealth w_i for $i = 1, 2, \ldots, R$ where R are the number of households in the rich list. We assume the rich list comprise of the R wealthiest households in the total population of households. Then if w_i are ordered, their empirical CDF and tail are thus the points $(w_i, 1-(R-i)/H_T)$ and $(w_i, (R-i)/H_T)$ respectively. The empirical tail of the rich lists are the points above £10^8 in Figure 1.1 and are matched for corresponding years to the survey data. Note a couple of survey points do also extend past £10^8.

The empirical tails can be found in the same fashion for four components of wealth defined as property, physical, financial and pension in the wealth and asset survey.

C.3 Mean and Variance of Returns

With R_n defined as in (4.2) and α_n i.i.d. from some distribution with $\mu = \mathbb{E}[\alpha_{n+1}]$ and $\sigma^2 = \mathrm{var}(\alpha_{n+1})$ as in Section 4.2 we have

$$\mathbb{E}[R_{n+1}|W_n] = \mathbb{E}[\alpha_{n+1}W_n^{\gamma-1}] = \mathbb{E}[\alpha_{n+1}]W_n^{\gamma-1} = \mu W_n^{\gamma-1}\,, \tag{C.1}$$

$$\mathrm{var}(R_{n+1}|W_n) = \mathrm{var}(\alpha_{n+1}W_n^{\gamma-1}) = \mathrm{var}(\alpha_{n+1})W_n^{2(\gamma-1)} = \frac{\sigma^2}{\mu^2}\mathbb{E}[R_{n+1}|W_n]^2\,. \tag{C.2}$$

Thus the interval I of one standard deviation around the mean of R_{n+1} used in Figure 4.1 is

$$I = \left(\left(1 - \frac{\sigma}{\mu}\right)\mathbb{E}[R_{n+1}|W_n], \left(1 + \frac{\sigma}{\mu}\right)\mathbb{E}[R_{n+1}|W_n]\right) = ((\mu - \sigma)W_n^{\gamma-1}, (\mu + \sigma)W_n^{\gamma-1}) \,. \tag{C.3}$$

C.4 Estimating the ROR

To approximate returns of individual household wealth using the WAS data [116] we use the returns on percentile wealth. For each time period n from the empirical tail of the survey we extract the percentile $w_{i,n}$ such that $\mathbb{P}_N(W_n > w_{i,n}) = p_i$ where $p_i = 1 - i/100$ for $i \in \{1, 2, \ldots, 100\}$. Note we only extract positive percentiles i such that $w_{i,n} > 0$, which excludes the poorest households. Then we substitute $W_n = w_{i,n}$ in (4.12) to calculate RORs of percentiles over each of the five biennial time periods 2008-2016 of the data [116]. For percentile i we have the ROR as

$$r_{i,n+2} = \frac{w_{i,n+2} - w_{i,n} - 2s_{i,n+2}}{2w_{i,n}}, \tag{C.4}$$

where $s_{i,n}$ are the savings in percentile i and time period n (see Section 4.4.3 for details).

Note that these percentile RORs for the ONS survey data [116] plotted in Figure 4.1 only approximate RORs for individual households. Our procedure does not account for households changing percentiles over a time period, leading to reduced fluctuations of the resulting returns data. To compensate for this and also possible effects of the financial crisis from 2008 onwards, we combine all time periods in a single data set to infer system parameters. For billionaires we have individual wealth data across time. We ignore savings to compute returns according to (4.12), and plot these values for 2016 in Figure 4.1.

In order to understand the dependence of ROR on wealth in particular for the UK, it is instructive to consider the different composition of wealth for poorer and richer households. Survey data [116] differentiate four components of wealth: property, physical, financial and pension, and their typical distribution is summarised in Figure C.1, exemplary for 2016 data. Financial and property wealth of the poorest decile have a negative sign (i.e. constitute debt), and the total average wealth in that decile is approximately 0 and not shown in Figure C.1.

The paper 'The rate of return on everything, 1870–2015' [82] provides a comprehensive analysis of average returns across four different types: bills, bonds, equity and housing over 1870-2015. In particular for the period 1980-2015 the average real rate of returns on equity and housing for the UK are 9.11% and 6.81%, respectively (Table 7, p 37 [82]). Therefore, the increasing proportion of property and financial wealth for wealthier households can account for RORs increasing with wealth. This is also confirmed in Figure C.1 (bottom), where we see that ROR (technically ROR with zero savings as it is unclear how to divide savings across components) for physical and pension wealth are largely independent of wealth, while property and financial ROR increase with wealth.

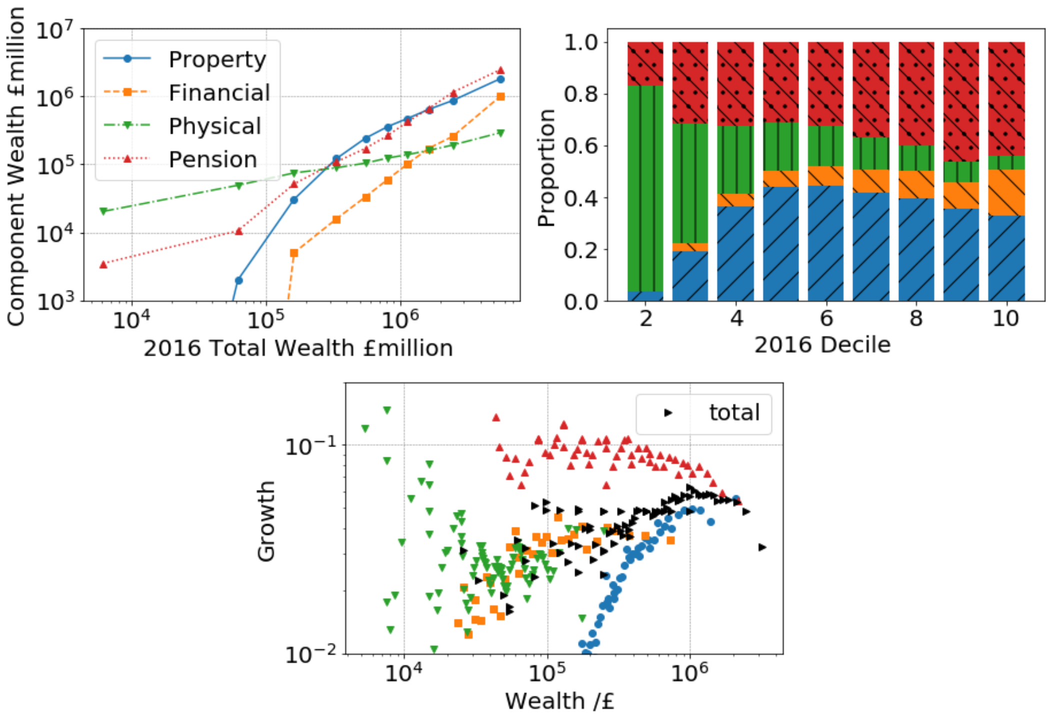

Figure C.1: Absolute wealth by components as a function of total wealth (top left) and wealth proportions by component of positive wealth deciles (top right), both from ONS data [116] from 2016. Bottom: ROR with zero savings (wealth growth) and same colour code averaged over time periods from 2008 to 2016, computed as described in (C.4) from percentile data [116].

C.5 Non-central t Distribution

We fit α_n with a non-central t distribution, see Figure 4.3. The non-central t distribution has been used for fitting stock returns that are both skewed and heavy tailed [141]. We define the non-central t distribution denoted with the random variable U as

$$U = \frac{Z + c}{\sqrt{V/k}}$$

where $Z \sim \mathcal{N}(0, 1)$, $c \in \mathbb{R}$ is the centrality parameter, and $V \sim \chi^2(k)$ with $k \in \mathbb{R}_{>0}$ degrees of freedom. We now define the shifted and scaled non-central t distribution with random variable W such that

$$W = sU + l$$

with the shift parameter $l \in \mathbb{R}$ and scale parameter $s \in \mathbb{R}_{>0}$. We denote W by

$$W \sim \mathrm{nct}(k, c, l, s)\,.$$

C.6 Supplementary Simulation Results

C.6.1 Generic Initial Conditions

Recall the replacement mechanisms (**R.1**)-(**R.3**) in case of bankruptcy events:

R.1 replace with a proportion of the agent's previous positive wealth value $pW_{n-1}(i) > 0$ such that p is uniformly chosen from $(0, 1]$;

R.2 replace with the agent's previous positive wealth value $W_{n-1}(i) > 0$;

R.3 replace with wealth $W_n(j) > 0$ of another uniformly chosen agent j.

We can see from Figures 4.5 (top left and right), C.2 and C.3 that the empirical tails and inequality measures of the simulations (4.3) evolve similarly in time for the three replacement mechanisms (**R.1**)-(**R.3**) until the system enters the crossover region. Then bankruptcy events become more frequent and relevant for the richest households, leading to significant differences with mechanism **R.3** naturally leading to slowest growth.

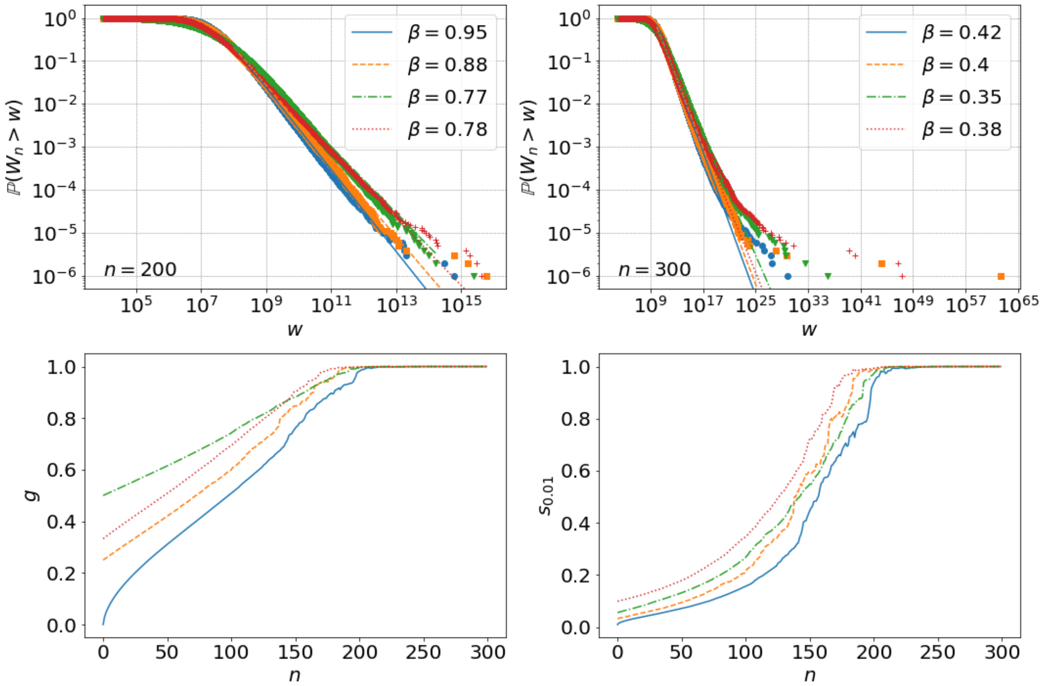

Figure C.2: Simulation (4.3) with $N = 10^6$ agents, zero savings $S_n = 0$, $\alpha_n \sim \text{nct}(k, c, l, s)$ with fitted parameters in (4.17) and $\gamma = 1.075$ for the four initial conditions with respective colour coding **I.1**-**I.4** and replacement mechanism **R.2**. Top left and right show empirical tails at times $n = 200, 300$ and power-law tail fits with exponents β. Bottom left and right show respective Gini, g, and top 1% wealth shares, $s_{0.01}$ up to $n = 300$.

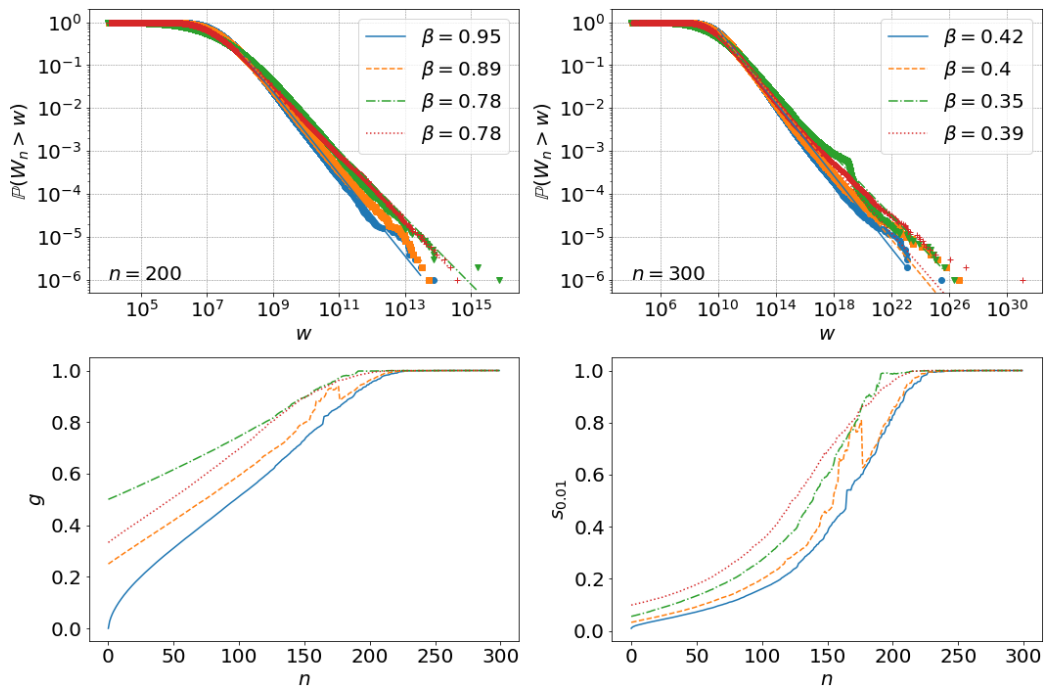

Figure C.3: Simulation (4.3) with $N = 10^6$ agents, zero savings $S_n = 0$, $\alpha_n \sim \text{nct}(k, c, l, s)$ with fitted parameters in (4.17) and $\gamma = 1.075$ for the four initial conditions with respective colour coding **I.1**-**I.4** and replacement mechanism **R.3**. Top left and right show empirical tails at times $n = 200, 300$ and power-law tail fits with exponents β. Bottom left and right show respective Gini, g, and top 1% wealth shares, $s_{0.01}$ up to $n = 300$.

C.6.2 Realistic Initial Conditions

We can see from Figures 4.9, C.4 and C.5 that the three replacement mechanisms (**R.1**)-(**R.3**) give very similar results on wealth distribution and inequality over time n, for the simulations described in the caption of Figure C.4. This is due to the much shorter time horizon compared to our numerical studies of generic initial conditions, and confirms that the choice of replacement mechanism is not crucial over limited time periods.

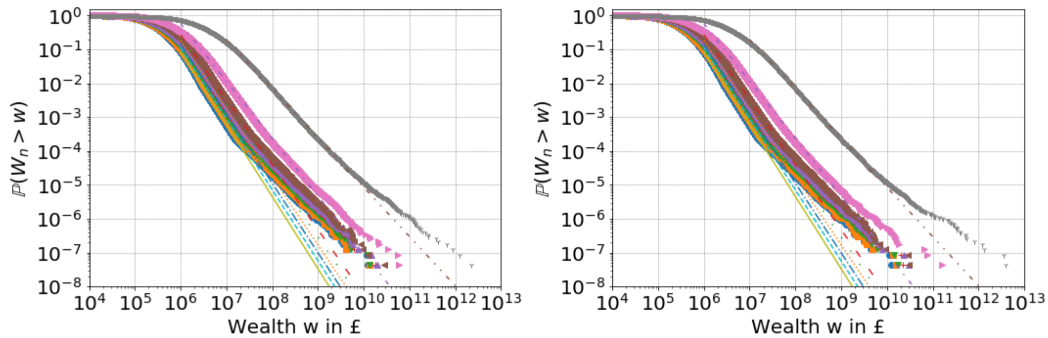

Figure C.4: Empirical tails for simulation (4.3) with $N \approx 23 \cdot 10^6$ agents, replacement mechanisms **R.2** (left) and **R.3** (right), fixed savings $S_n = S(W_0)$ (4.16), $\alpha_n \sim \mathrm{nct}(k, c, l, s)$ with fitted parameters in (4.17) and $\gamma = 1.075$ for 2008 initial conditions at times $n = 0, 2, 4, 6, 8, 10, 20$ and 50. Power law fits with exponents β decreasing from $\beta = 2.13$ at $n = 0$ to $\beta = 1.45$ at $n = 50$.

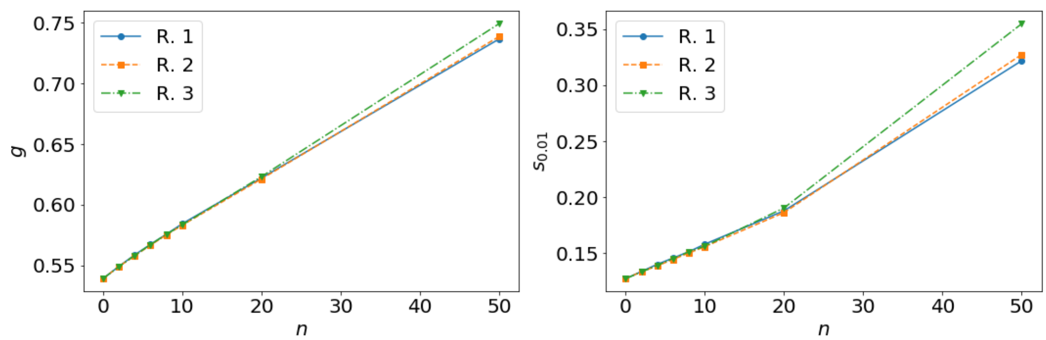

Figure C.5: Gini, g, (left), top 1% wealth shares, $s_{0.01}$, (right) for simulation (4.3) with $N \approx 23 \cdot 10^6$ agents, fixed savings $S_n = f(W_0)$, $\alpha_n \sim \mathrm{nct}(k, c, l, s)$ with fitted parameters (4.17) and $\gamma = 1.075$ with rough 2008 initial conditions and replacement mechanisms **R.1**-**R.3**.