Towards a Statistical Equilibrium Theory of Wealth Distribution

by

Mishael Milaković

May 2003

Submitted to the Graduate Faculty of Political and Social Science of the New School University in partial fulfillment of the requirements for the degree of Doctor of Philosophy.

Dissertation Committee:

Dr. Duncan Foley
Dr. Salih Neftci
Dr. Lance Taylor

To my parents, Dušanka and Ivan.

Preface

My dissertation is based on three essays, unpublished to date, that have been written in a format suitable for publication in professional journals. It is not unheard of in our field to literally bind up three such essays and preface them with a more or less coherent story that explains why they are fitting pieces in a larger scheme. I have decided against such practice in favor of a monograph style, which, on one hand, should make the presentation of my distributional theory more accessible to readers who are not familiar with the concept of statistical equilibrium.

On the other hand, the monograph style will hopefully result in a more elegant presentation of the main ideas. Since all three essays are closely related but specifically geared towards journal publication, the presentation would be repetitive and even redundant if the essays were simply placed in sequential order. I believe nobody enjoys reading three introductions that explain the same concepts three times over, differing only in length and the degree of detail.

At the core of my dissertation is a theoretical model that explains the power law distribution of wealth as the statistical equilibrium outcome of a system with many heterogeneous agents who face an aggregate wealth constraint with respect to the average rate of growth. The second essay, a co-authored effort with Carolina Castaldi, deals with the empirical calibration of the theoretical model, including the estimation of the power law exponent for the years 1996–2002 from a named subset of the wealthiest individuals in the United States. The third essay reviews the formal concept underlying statistical equilibrium analysis, known as the *maximum entropy principle*, with the aim of extending the statistical equilibrium theory of distribution beyond the power law tail. The vast majority of individuals in an economy—who often account for only half of total wealth—are usually described by a Gamma law, which has different economic implications from the viewpoint of statistical equilibrium analysis.[1]

I have rearranged and modified some parts of these essays in order to best represent the involved ideas as a coherent body of thought. Quite obviously I cannot claim to have provided a general or complete theory (if there ever is such a thing) of wealth distribution in statistical equilibrium—hence the "Towards" in the dissertation title. Yet I would like to believe that the fundamental or most elementary principles of distributional theory in statistical equilibrium have been worked out here.

[1] The essay titles are, in the order mentioned here, *A Statistical Equilibrium Model of Wealth Distribution*; *Turnover Activity and Wealth Mobility*; and *Do We All Face the Same Constraints?*

Acknowledgments

I would like to thank my supervisor Duncan Foley, as well as Salih Neftci and Lance Taylor, not only for serving on my committee but for educating and supporting me in many ways during the long course of my studies and dissertation work. It is more than an empty phrase that without them my work could not have been undertaken—for what it is worth, I would like to express my deepest gratitude to them.

I am also grateful to Giovanni Dosi for his support and kind invitation to Scuola Superiore Sant' Anna in Pisa, Italy, where I made substantial progress towards the completion of my dissertation.

The company of Nelson Barbosa, Josh Bivens, Giulio Bottazzi, Dirk Jenter, Ilfan Oh, Angelo Secchi, and Toshihiko Udagawa has been a privilege that provided me with an unique environment of inspiration and intellectual stimuli.

A special thanks goes to Carolina Castaldi for being my co-author and friend, to Ljubiša Grujčić for keeping me straight, and to my parents for unconditionally supporting me to the best of their abilities from Day One.

Table of Contents

List of Figures

List of Tables

Chapter 1

Brief Taxonomy and Stylized Facts of Wealth Distribution

The economic sources of wealth are income, inheritance, and the revaluation of assets or liabilities. Savings are a theoretical accounting tool, essentially describing the mediation from income flows to the stock of wealth.

The economic uses of wealth are expressed in the composition of wealth portfolios and break down into five broad categories: (i) cash and savings accounts; (ii) financial assets like bonds, stocks, and their derivatives; (iii) real estate, held either for investment purposes or in the form of owner-occupied housing; (iv) retirement provisions like pension accounts and life insurance plans; and (v) stakes in private or unincorporated businesses.

How, or whether, human capital might influence an agent's flow of income or her investment decisions will both remain unanswered questions here. Our concept of wealth only requires the sources and uses of wealth to fulfill an accounting identity in order to be logically consistent. Hence we will not theorize about causal relationships among the involved accounting categories in the subsequent analysis. One implication of such an approach to the concept of personal wealth is that we cannot make any statements about human capital wealth in particular, nor about the *individual* destinies of agents in general. It is rather the distribution of the measured accounting level of wealth that we want to explain.

Unfortunately, personal wealth data cannot be measured with high accuracy. Surveys of household wealth, personal estate tax—and in some countries that levy them—wealth tax data, individual investment income data, as well as independent estimates on a subset of very wealthy named individuals (e.g. compiled by *Fortune* and *Forbes Magazine* in the US or *The Sunday Times* in the UK) all suffer from shortcomings that are discussed and summarized in [7]. Nevertheless, a clear qualitative picture emerges from household survey data regarding the composition of wealth. According to Wolff [46, 47], the composition of US marketable wealth[1] has remained fairly stable over the last two decades, with roughly two thirds of household wealth stemming from owner-occupied housing (30 percent gross value), other real estate (15 percent gross value), and business equity (20 percent). Financial securities make up 15 percent of total household wealth, matched in

[1] Common accounting definitions of wealth, all of which exclude the concept of human capital wealth, are given in Appendix A.

size by the share of total deposits—i.e. checking and saving accounts, time deposits, money market funds, and CDs—and retirement accounts. The remaining share in total wealth, amounting to roughly 5 percent, is composed of various assets, including net equity in personal trusts, precious metals, royalties, jewelry, antiques etc. Individually, none of these items are of significant magnitude.

Central to our analysis, however, is not the relatively unchanging composition of total wealth but the composition of wealth by wealth class. The very wealthy hold most of their wealth in financial assets and investment real estate while the not-so-wealthy hold theirs primarily in the form of owner-occupied housing, deposits, and pension and life insurance plans [7, 41]. To give a numerical illustration, Wolff [46] calculates that in 1989 the top one percent of US wealth holders had 52 percent of their wealth invested in investment real estate and unincorporated businesses, 29 percent in traditional financial securities, 11 percent in liquid assets and only 8 percent in owner-occupied housing. In contrast, the bottom 80 percent of households held 63 percent of their wealth in the form of owner-occupied housing, 21 percent in the form of liquid assets, 10 percent in real estate and business equity, and only 6 percent in traditional financial assets. Folbre [12] estimates that in the same year the richest one percent of the US population held 45 percent of all nonresidential real estate, 62 percent of all business assets, 49 percent of all publicly held stock, and 78 percent of all bonds. The richest 10 percent of families held 80 percent of all nonresidential real estate, 91 percent of all business assets, 85 percent of all stocks, and 94 percent of all bonds. Except for a slight deviation in the figure for bonds held by the top one and top ten percent of households, Wolff's data are identical to Folbre's. In addition, he shows that the bottom 90 percent of wealth holders account for 64 percent of all principal residences, 55 percent of the value of life insurances, 40 percent of deposits, and 38 percent of the value of pension accounts [46]. In the UK, evidence from estate data confirms the qualitative picture observed in the US [41].

The relevant stylized fact for our model will be the pronounced difference of portfolio compositions between the very wealthy and the rest. Different households are subject to different economic processes that govern their possibilities of accumulating personal wealth. We want to argue that the vast majority of households engages in a 'life cycle' type of saving in order to provide them with housing and financial claims that will ensure their economic viability beyond working age. Hence, their wealth will be roughly proportional to earned income, describing an *additively* driven process designed to realize a return in the distant future. In contrast, the very wealthy accumulate their riches mainly by re-investing returns in financial assets and speculative real estate. Wealthy households seek to realize returns throughout their lifetime, thereby accumulating wealth in *multiplicative* fashion.

Thus it is not surprising that we observe two different functional forms that describe the distribution of wealth. The upper tail of wealth distributions displays remarkable regularity in the functional form of a power law, typically covering the richest three to five percent of households and sometimes accounting for over half of total wealth [1, 10, 30, 42, 45]. The riches of the remaining 95 percent of the population with positive wealth—the 'left part' of the distribution—are typically Gamma distributed [6, 10, 35].

Walrasian theory cannot explain this regularity endogenously since wealth enters exogenously in the form of endowments. Markets will not change the distribution of wealth because exchange takes place exclusively at equilibrium prices, which ensures that the value of a chosen consumption bundle will equal the value of the endowment.

Economic models based on intertemporal maximization plans of heterogeneous agents also have difficulties reproducing the observed distribution of wealth, particularly when it comes to the upper tail.[2] In contrast, models from probability theory [4, 29, 31, 32, 37] provide insights into why a variable should be distributed according to a power law—but they often lack a clear relationship to economic theory.[3] We will develop an alternative probabilistic theory of the power law distribution of wealth in Chapter 3—a probabilistic theory that is firmly rooted in the "economics" of the stylized facts we just reviewed. Before doing so, we study some elementary properties of the statistical equilibrium methodology in the next chapter, which will also prove useful in Chapter 6, where we discuss extensions of the theory to account for the Gamma distribution of wealth.

[2] Quadrini and Rios-Rull [36] provide a survey of the literature.
[3] See Brock [3] and Gabaix [15] for critical assessments.

Chapter 2

The Concept of Statistical Equilibrium and Some Useful Formal Results

Market economies consist of a large number of heterogeneous agents whose interactions produce aggregate consequences—possibly unintended and regularly unforseen—that feed back into agents' behavior and the environment they interact in. The vast amount of information in such a complex system does not allow us to explain the distribution of wealth by tracing the microscopic fate of all agents. The concept of statistical equilibrium [13, 14] acknowledges this difficulty from the start and consequently curbs its methodological ambition to more modest levels, being content with describing the statistical properties of aggregate outcomes as a probability distribution of economic agents over possible outcomes.

The mathematical formalism underlying statistical equilibrium is known as the *maximum entropy principle.* Building on entropy concepts from statistical mechanics and information theory, Jaynes [21] generalized the principle of entropy maximization into a theory of probabilistic inference that has found numerous applications across the natural and social sciences [23]. Based on the premise of exclusively incorporating knowledge that has been given to us and scrupulously avoiding probabilistic statements that would imply more information than we actually have, the maximum entropy principle derives probability distributions from known moment constraints. Virtually all known distributions—discrete as well as continuous—can be derived from the maximum entropy principle [22].

Let us denote the number of theoretically admissible values of our variable of interest x by $i = 1, \ldots, z$; then the maximum entropy principle prescribes to maximize (informational) entropy $H \equiv -\sum_i p_i \log p_i$ subject to the natural constraint $\sum_i p_i = 1$ and $m < n$ observed moment constraints $\sum_i p_i g_k(x_i) = \overline{g_k}$ for all $k = 1, \ldots, m$. Applying Lagrange's multiplier technique yields probability distributions of the generic form $p_i = Z^{-1}(\lambda_1, \ldots, \lambda_m) \exp(-\lambda_1 g_1(x_i) - \ldots - \lambda_m g_m(x_i))$, where $Z(\lambda_1, \ldots, \lambda_m) \equiv \sum_i \exp(-\lambda_1 g_1(x_i) - \ldots - \lambda_m g_m(x_i))$ is the *partition function* that normalizes the distribution and $\lambda_1, \ldots, \lambda_m$ are the Lagrange multipliers chosen so as to satisfy the moment constraints, which is the case when $\overline{g_k} = -\partial \log Z / \partial \lambda_k$ for all $k = 1, \ldots, m$. Concavity of the objective function and linear (or a convex set of) constraints ensure that the resulting probability distribution is unique and attained at a global entropy maximum, while the exponential form of the generic distribution admits only positive probabilities so that we do not have to incorporate non-negativity constraints explicitly into the variational problem.

The continuous case proceeds almost analogously because the Euler-Lagrange equation of the Calculus of Variations [5] tells us that the solution to an extremal problem of the form $\int \mathrm{d}x F[x, f(x), f'(x)]$, where F is a known function, corresponds to $\partial F/\partial f(x) - \mathrm{d}[\partial F/\partial f'(x)]/\mathrm{d}x = 0$. Since the Lagrangian of the continuous maximum entropy program does not involve $f'(x)$ the constrained maximization problem can be solved with the regular method of Lagrange multipliers. In passing to the continuous case as the limit of the discrete case, however, we encounter a technical difficulty: in order to keep the entropy results invariant with respect to a change of variables we need to introduce an "invariance measure" $m(x)$ in the objective function, which now becomes $H^c \equiv -\int_R \mathrm{d}x f(x) \log[f(x)/m(x)]$, where R is the support and $f(x)$ the probability density of x, and $m(x)$ is proportional to the limiting density of discrete points. If we transform x, say, by measuring wealth in euros instead of dollars, both $f(x)$ and $m(x)$ transform in the same way to the effect that all parameters of the maximum entropy distribution remain unchanged.

In the absence of any moment constraints the continuous maximum entropy distribution will be proportional to $m(x)$ so that we can interpret it as the prior distribution expressing 'complete' ignorance. The question which prior best represents such a state of complete ignorance has divided probability theorists for over two centuries now, and we do not have the ambition to add anything to this discourse.[1] We notice in passing two ways around the philosophical dilemma of continuous maximum entropy distributions. First, if it is sensible to express the variable of interest as a ratio we avoid the problem of choosing a particular prior altogether. Second, in the spirit of Laplace, discrete entropy maximization prescribes to regard the uniform distribution as the best representation of a state of 'complete' ignorance. A uniform prior will work even in the continuous case but only on a finite support. For an infinite interval, Kapur [22] argues that taking a constant prior "[...]means using an improper prior distribution which will be justified if the posterior maximum entropy probability distribution is a proper distribution." In spite of the measure issue, the general properties and intuition of entropy maximization carry over from the discrete case. Maximizing H^c subject to $\int_R \mathrm{d}x f(x) = 1$ and $\int_R \mathrm{d}x g_k(x) f(x) = \overline{g_k}$ for all $k = 1, \ldots, m$ leads to a probability density $f(x) = Z^{-1}(\lambda_1, \ldots, \lambda_m) m(x) \exp(-\sum_k \lambda_k g_k(x))$ with the partition function $Z(\lambda_1, \ldots, \lambda_m) \equiv \int_R \mathrm{d}x\, m(x) \exp(-\sum_k \lambda_k g_k(x))$.

But making sound probabilistic statements from limited information does not exhaust the methodological scope of statistical equilibrium. According to Jaynes' *concentration theorem* [21] the distribution of maximum

[1] The interested reader may consult a series of articles by Jaynes [19, 21, 20] that deal extensively with questions of what the appropriate choice of $m(x)$ should be.

entropy is not only 'most likely' in the combinatorial sense of being achievable in the largest number of ways—but the overwhelming majority of possible distributions compatible with our constraints will have entropy very close to the maximum.[2] Thus inference from observed constraints to resulting frequency distribution becomes exceptionally robust *and vice versa*: suppose our variable of interest is distributed with a specific functional form; then the concentration theorem assures us of the extreme improbability that constraints other than those implied by the maximum entropy principle are responsible for the observed outcome, and probability distributions and aggregate constraints become two sides of the same coin.

Statistical equilibrium presents a potent tool for the analysis of large complex systems, but its descriptive and predictive ability is contingent on the model incorporating all relevant constraints that produce the observable regularities in the system. We will argue in Chapter 3 why decentralized investment activity of wealthy households, who are constrained by the growth rate of wealth, leads to a power law distribution of wealth. Recent evidence on the UK distribution of wealth [10] illustrates that the power law distribution covers roughly the top 5% of households while the remaining 95% of the population supposedly show exponentially distributed wealth. A multiplicative constraint alone cannot explain the observed regularity across all

[2] Jaynes' concentration theorem can even quantify 'the overwhelming majority' of possible cases because it shows that (twice the number of observations multiplied by) the entropy difference among feasible distributions will be distributed over the set of feasible distributions as Chi-squared with degrees of freedom equal to the number of theoretically possible outcomes minus the number of constraints minus one.

households and we should determine, at least in principle, which constraint will result in an exponential distribution.

It is well known that an arithmetic mean constraint in the maximum entropy program leads to the exponential distribution [14, 21, 22], while a logarithmic mean constraint results in a power law [22, 23]. But what happens in a system simultaneously constrained by arithmetic and logarithmic means? Suppose wealth x can take on a continuum of values on $R = [0, \infty)$ and that all households would face additive as well as multiplicative constraints. Then the statistical equilibrium distribution of wealth will be given by the solution to the following maximum entropy program,

$$\max_{f(x)} H \equiv -\int_R \mathrm{d}x f(x) \log f(x) \tag{2.1}$$

subject to

$$\int_R \mathrm{d}x f(x) = 1\,, \tag{2.2}$$

$$\int_R \mathrm{d}x f(x) x = \bar{x}\,, \tag{2.3}$$

$$\int_R \mathrm{d}x f(x) \log x = \overline{\log x}\,. \tag{2.4}$$

The associated Lagrangian is $L \equiv H - \lambda(\int_R \mathrm{d}x\, f(x) - 1) - \mu(\int_R \mathrm{d}x\, f(x)x - \bar{x}) - \nu(\int_R \mathrm{d}x\, f(x) \log x - \overline{\log x})$ and implies the first order condition $\partial L / \partial f(x) = -(\log f(x) + 1) - \lambda - \mu x - \nu \log x = 0$. Solving for $f(x)$ and determining the partition function from the natural constraint (2.2) we have

$$f(x) = \frac{e^{-\mu x} x^{-\nu}}{\int_R dx\, e^{-\mu x} x^{-\nu}} . \tag{2.5}$$

The gamma function is defined as $\Gamma(a) = \int_R dx\, e^{-x} x^{a-1}$; by letting $z = \mu x$, so that $dz = \mu dx$, we can express the partition function $\int_R dx\, e^{-\mu x} x^{-\nu}$ as $\mu^{\nu-1} \int_R dz\, e^{-z} z^{-\nu} = \mu^{\nu-1}\Gamma(-\nu+1)$. Finally, we define $\gamma = -\nu + 1$ and thus obtain the well-known gamma distribution for x

$$f(x) = \frac{\mu^{\gamma}}{\Gamma(\gamma)} e^{-\mu x} x^{\gamma-1} . \tag{2.6}$$

If only (2.3) was prescribed, the maximum entropy distribution would be

$$f(x) = \mu\, e^{-\mu x} , \tag{2.7}$$

while if (2.4) alone was given, the resulting distribution would be

$$f(x) = \frac{\nu - 1}{m^{1-\nu}} x^{-\nu} , \tag{2.8}$$

with $m > 0$ as the minimum value of x.

Notice that the Gamma probability density reduces to the exponential density if $\gamma = 1$ because the definition of γ then implies $\nu = 0$ such that there is no multiplicative constraint in the first place. Thus we see how pronounced the difference among maximum entropy distributions will be

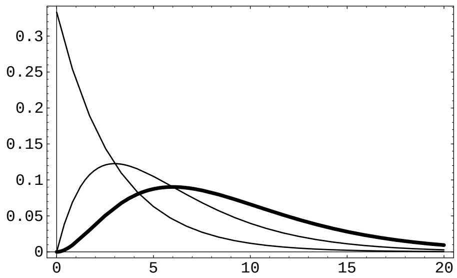


Figure 2.1: Exponential probability density function with scale $\mu = 1/3$, and hump-shaped Gamma density functions with identical scale $\mu = 1/3$ and different shape parameters $\gamma = 2$ and $\gamma = 3$ (thick curve).

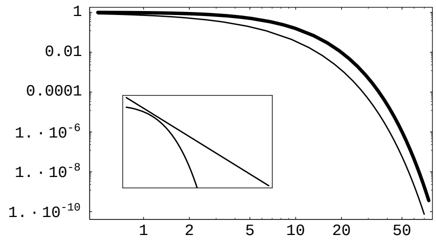


Figure 2.2: Inverse cumulative distribution functions of the exponential and Gamma distribution (thick curve). The inset shows the tail behavior of the inverse Gamma vs. the inverse power law; both graphs are on double-logarithmic scale.

depending on whether the constraints (2.3) and (2.4) are simultaneously or separately present. Figure 2.1 illustrates the marked difference among the respective probability *densities*. Yet the *cumulative* distributions of the Gamma and exponential laws, shown in Figure 2.2, are very similar.

In contrast, the behavior of the inverse cumulative Gamma distribution will be governed by the exponential term for large values of x so neither the Gamma nor the exponential law are capable of producing the 'fat' tail of a power law distribution, as illustrated in the inset of Figure 2.2.

We are now in a position to tackle the power law distribution of wealth in the next chapter, with slightly more rigorous attention to detail as well as with more economic content in the maximum entropy formalism.

Chapter 3

Statistical Equilibrium Theory of the Power Law Distribution of Wealth

We recall from Chapter 1 that wealth consists of the current value of assets a household owns minus the current value of liabilities it services, and that the economic sources of wealth are income, inheritance, and the revaluation of assets or liabilities. In the case of very wealthy households that make up the power law tail, income mostly flows from financial assets, rents, and business operations. Savings are mediating between income and asset acquisition. The economic uses of wealth are expressed in the composition of wealth and lead to the notion of a household's *wealth portfolio.* A complex set of market interactions determines the value of different components in the portfolios and, thereby, the distribution of wealth.

The underlying complexity introduces an enormous amount of information, making it impractical to model the distribution of wealth by tracing the fate of individual portfolios. We can, however, observe a well-defined average growth in the whole economy that constrains the growth of individual portfolios. Each portfolio has a characteristic *return factor*, corresponding to a portfolio's gross return over a given period. This cumulative return factor can be thought of as a combination of different rates of return accruing to the different uses of the portfolio. Under the assumption that at some initial point in time we start out with an egalitarian distribution of wealth, where each household enjoys the same level of wealth, it follows that return factors and wealth levels will be proportional. Differences in the return factors that each of the portfolios achieve are thus responsible for differences in wealth.

We take the position that differences in return factors are first and foremost the result of decentralized investment activity per se, and not of individual skill and ability, nor the inheritance of a family dynasty. That does not mean factors like human capital wealth, inheritance, personal ability, lucky streaks (or, for that matter, losing streaks) are excluded from the analysis. All such factors are part of the general environment leading to the statistical equilibrium outcome.[1] To paraphrase the methodological view

[1] Though individual determinants of a household's fortune cannot be separately identified in our model, all such factors are at least in principle included in the characteristic return factor.

underlying the field of "complexity science"—as popularized in [2, 44]—our model considers aggregate properties of the economy as being caused by the very process of aggregation.

Competitive markets ensure a tendency towards a uniform rate of return for activities with the same risk. We interpret the uniformity in the sense that returns to wealth will be different in absolute terms while tending to be proportional to the size of the wealth portfolio, so that the *rate* of return is independent of the size of the portfolio.[2]

Since different activities bear different risks, however, individual portfolios will ultimately experience different realizations of risky prospects. The realization of a portfolio's return factor will depend on the number of *turnovers* that occur. A turnover reflects a household's decision to change the composition of its portfolio, by either changing the weights of existing components or by including components previously not held.

Formally, the model builds on Jaynes' [21] *maximum entropy program* that we introduced in the previous chapter, and on Foley's [13] economic interpretation of the program as a *statistical equilibrium* of markets. When the number of wealthy households is large, combinatorial factors can lead to statistical regularities in the distribution of wealth. The wealth distribution that can be achieved in the largest number of ways while satisfying the ag-

[2] Wolff [46, 47] documents systematic differences in return factors between the rich and the rest. We justify the assumption of uniform rates of return by pointing out that we are only concerned with the upper tail of the wealth distribution, where households display similar compositions of their portfolios, primarily investing in financial assets and real estate.

gregate growth constraint is the *statistical equilibrium* or *maximum entropy wealth distribution*[3]

An arithmetic mean constraint in the maximum entropy program leads to the (Gibbsian) exponential distribution.[4] Basically, our model establishes the power law distribution as the outcome of the maximum entropy program under a logarithmic mean constraint. A logarithmic scale expresses proportionality; the idea that intervals of proportionate extent are responsible for the emergence of power laws dates back to Champernowne's [4] model of income distribution. More recently, Levy and Solomon [29] have developed a generalization of Champernowne's Markov chain model. They demonstrate that a power law emerges from a less restrictive stochastic process, only requiring it to be multiplicative—even if the process is not stationary or if the transition probabilities of the process change over time. Starting from a stochastic difference equation for wealth $w_i(t+1) = \gamma w_i(t)$, where the multiplicative factor γ has an arbitrary distribution $P(\gamma)$ with finite support, Levy and Solomon [29] prove that the ergodic distribution of w will converge to a power law.[5]

The key to their proof lies in the logarithmic scale of wealth, so that

---

[3]It should be understood implicitly by now that generic terms like 'wealth distribution' or 'households' strictly refer to the upper tail of the distribution.

[4]Kapur and Kesavan [22, 23] present numerous applications of the maximum entropy program under different constraints taken from the natural and social sciences. Foley [13] provides an economic example where the Gibbsian exponential distribution (of commodity prices) arises from an arithmetic mean constraint (such that excess demand for commodities equals zero).

[5]The ratios of the midpoints of Champernowne's intervals are rates of return, very similar to Levy and Solomon's [29] multiplicative factor γ.

the particular shape of $P(\gamma)$ will not influence the ergodic distribution of wealth as a power law. Instead of assuming an arbitrary distribution of return factors, our model treats all return factors as equally likely and then determines the distribution of wealth that can be achieved in the largest number of ways while meeting the aggregate growth constraint. But it does not matter whether we assume an arbitrary distribution, or whether we assume return factors to be equally likely and then mix them in the most disorganized fashion: in both models, the power law distribution depends on the logarithmic scale of wealth.

Where our model differs from theirs, however, is in the exponent that characterizes the statistical properties of the power law distribution. As far as the distribution of wealth is concerned, we can interpret the magnitude of the characteristic exponent as a measure of inequality: the greater the exponent in absolute value, the more equal the distribution of wealth; the closer to unity the exponent of the cumulative distribution function, the more unequal the distribution [1, 25, 42].[6] Economic policy aimed at influencing the degree of inequality would have to ask which economic forces determine the characteristic exponent of the wealth distribution. Levy and Solomon's [29] characteristic exponent depends on an exogenous lower bound of the distribution. From the viewpoint of economic theory, an arbitrary lower bound carries little in the way of relevant information. In addition to a minimum

[6] If the characteristic exponent is less or equal to unity, the power law distribution is degenerate because it has infinite mean. Correspondingly, the *density* function will be degenerate if the characteristic exponent is smaller or equal to two.

wealth level, the statistical equilibrium exponent depends on the aggregate rate of growth in wealth portfolios, as well as on the average number of turnovers that occur during the period.

I Economic Foundations

We conceptualize the economy as a set $\mathbf{K} = \{1, \ldots, K\} \subseteq \mathbf{N}$ of *economic activities* or *investment opportunities*. For all $k \in \mathbf{K}$, let $V^k(t)$ denote the time t value of economic activity k, and for all $h \in \{1, \ldots, n\}$, $n < \infty$, let $a_h^k(t)$ denote the *position* of household h in activity k, with the interpretation that $a_h^k(t) > 0$ indicates a long position at time t (k is an *asset*) and $a_h^k(t) < 0$ a short position (k is a *liability*). Obviously, $a_h^k(t) = 0$ allows for the absence of activity k in the portfolio of household h.

The value of the *wealth portfolio* of household h at time t, denoted $w_h(t)$, follows from the household's combination of the K different activities in the economy

$$w_h(t) \equiv \sum_{k \in \mathbf{K}} a_h^k(t) V^k(t) \qquad \forall h \in \{1, \ldots, n\}.$$

Changes in the value of a household's portfolio are either the result of a revaluation of economic activities, or of changes in the behavior of the household—expressed as changes in the household's positions. Traditionally, we think of savings as the principal component determining wealth. Since our model conceptualizes wealth from its uses, savings are implicitly included in the above formulation.

Notice that we are not putting forward a specific theory of portfolio choice here. Instead, our model starts from the weak assumption that we observe a well-defined macroscopic average—the logarithmic mean—and that agents change the composition of their wealth portfolios over time. To measure changes in portfolio composition, we introduce the concept of an average number of *turnovers* in the economy.

Suppose for the moment that there is an 'initial' period t_0, where the portfolio starts out with an amount $w(t_0)$. The fictional device of an initial period serves to conceptualize the value of a wealth portfolio in terms of an average *return factor per turnover* r_h directly proportional to the observed wealth level. A *turnover* describes a change in the household's position during the observational period t_0 to t. Let $T_h(t_0, t^)$ designate the number of elements where $a_h^k(t_0) \neq a_h^k(t)$ for all $k \in \mathbf{K}$, that is to say $T_h(t_0, t)$ gives the number of changes in the composition of household h's portfolio between period t_0 and t. Moreover, let the economy start with an egalitarian distribution of wealth at t_0, where $w_h(t_0) = w_0$ for all $h \in \{1, \ldots, n\}$, and designate $T \equiv \sum_h T(t_0, t)/n$ to be the average number of turnovers in the economy. Suppressing the time index for notational simplicity we have that

$$w_h \equiv w_0 \, {r_h}^T \tag{3.1}$$

so wealth levels and return factors will be directly proportional for all households.

A formulation in terms of return factors allows us to interpret differences in wealth as differences in the returns each portfolio achieves over the period t_0 to t. Though returns in absolute terms will be different they should be proportional to the size of the portfolio if the economy is competitive. In other words, wealthier and poorer portfolios will face the same prospective rates of return, which does not exclude the possibility that different portfolios ultimately experience different realizations of risky prospects. We express proportionality in return factors with a logarithmic scale of wealth, $\log w$.

At the same time, we can also interpret the logarithmic scale as incorporating the growth dynamics of wealth in the sense of a geometric mean.[7] Denote the number of households with wealth w_i by n_i and define $p_i \equiv n_i/n$ to express the logarithmic mean $\overline{\log w}$ as

$$\overline{\log w} = \sum_i p_i(r) \log \left(w_0 \, {r_i}^T\right) = \log w_0 + T \sum_i p_i(r) \log r_i \, . \tag{3.2}$$

A logarithmic mean by itself has no time dimension. We can think of two different time scales, one being the passage of accounting time, the other being the passage of turnovers. The value of portfolios at the end has to be the same, regardless of which time scale we employ. We cannot observe r_i, the return factor per turnover, since in practice we do not know how many turnovers have occurred. What we can observe is the return factor *per year*,

[7]The logarithmic mean is equivalent to a weighted geometric mean where we interpret the weights as probabilities.

denoted R, so that we can express the logarithmic constraint as

$$\overline{\log w} = \log w_0 + T\,\overline{\log r} = \log w_0 + L\,\overline{\log R},$$

where $L = t - t_0$ is the calendar time that has elapsed since the (mythical) initial state. As we will argue in Section IV, the use of two different time scales allows us to resolve the conceptual issue of a 'zero period' and the absolute lapse of time. The growth constraint of aggregate wealth in equation (3.2) thus reads

$$\frac{L}{T}\overline{\log R} = \sum_i p_i(r) \log r_i \,. \tag{3.3}$$

It is important to notice that we are no longer summing over households but over the number of *theoretically possible wealth levels* w_i for all $i \in \mathbf{N}$. In order to ensure that each of the n households is assigned to some wealth level for all $i \in \mathbf{N}$, we have the additional constraint that $\sum_i n_i = n$, or, equivalently

$$\sum_i p_i(r) = 1. \tag{3.4}$$

Except for the notion of a turnover, we are neither making assumptions about the evolution of household behavior nor about the evaluation of economic activities, nor about whether valuation and individual behavior are interdependent. The growth constraint (3.3) reduces the enormous complexity of asset valuation and individual behavior to the observation of a single economy-wide average growth in household wealth. Hence, as it stands so far, our model is drastically under-determined in the sense that we can conceive of a large number of wealth distributions that are consistent with (3.3) and the natural constraint (3.4). Which probability distribution should we choose in the absence of any further information?

II Maximum Entropy Wealth Distribution

A feasible wealth distribution obeys (3.3) and (3.4). It will clearly remain feasible if we interchange households that enjoy the same wealth level $w_i(t)$ since doing so does not change the distribution. In the absence of any further information, Laplace's "principle of insufficient reason" prescribes to regard each theoretically possible wealth level or return factor as equally likely. Then the likelihood of observing any particular wealth distribution is proportionate to the number of ways that distribution can be achieved by permuting economically indistinguishable households, meaning households that achieve the same return factor in their wealth portfolio.

The number of ways n households can be assigned to C categories, with n_c households assigned to category c is the *multiplicity* of the assignment [13], $M[\{n_c\}] \equiv n!/n_1! \cdots n_c! \cdots n_C!$. Stirling's approximation for large n implies $\ln n! \approx -n + n \ln n$, which upon substitution into the logarithm of the multiplicity yields the *entropy* H of a distribution, $n^{-1} \ln M[\{n_c\}] \approx -\sum_{c=1}^{C} \frac{n_c}{n} \ln \frac{n_c}{n} \equiv H\left[\left\{\frac{n_c}{n}\right\}\right]$. From a statistical point of view, the rationale behind maximizing entropy is that the distribution that can be achieved in the largest number of ways is the most likely distribution to be observed.

The *maximum entropy program* [21] maximizes entropy $H[\{p_i\}]$ subject to the natural constraint (3.4) and a finite number of moment constraints. In our case, we are dealing with the single logarithmic constraint (3.3) that we interpreted as the average growth rate of wealth in the upper tail

$$\max_{\{p_i\}} H[\{p_i(r)\}] = -\sum_i p_i(r) \log p_i(r) \tag{3.5}$$

subject to

$$\sum_i p_i(r) \log r_i = \frac{L}{T} \overline{\log R},$$
$$\sum_i p_i(r) = 1.$$

We can think of the maximum entropy program as assigning a probability distribution based on the premise of using only information we have and strictly avoiding use of any additional information [21, 23].[8] If an entropy-maximizing wealth distribution exists, it is unique [13, 21] because the objective function (3.5) is strictly concave and the constraints define a convex set. The maximum entropy program yields the proof of a similar theorem by Levy and Solomon [29].

[8]It is easily verified that maximizing entropy H subject to the natural constraint results in the uniform distribution—a modern formulation of the principle of insufficient reason.

Theorem 1 (Power laws are logarithmic Boltzmann laws). *For all positive return factors r_i there exists $(\lambda^*, \lambda_0^*) \in \mathbf{R}^2$ such that the optimal solution to the maximum entropy program under a logarithmic growth constraint is a power law distribution of wealth*

$$p_i^*(r) = \frac{r_i^{-\lambda^*}}{Z(\lambda^*)} \tag{3.6}$$

where

$$Z(\lambda^*) \equiv \sum_i r_i^{-\lambda^*} = \exp(\lambda_0^*)$$

is the partition function that normalizes the probability distribution $p_i^*(r)$.

Proof. If return factors are not concentrated in a single point, i.e. for any constant $c > 0$, $r_i \neq c$ for at least one i, the nondegenerate constraint qualification of the optimization program is satisfied and there exists a *characteristic exponent* λ^* and a normalizing multiplier $\lambda_0^* \equiv 1 + \mu^*$ such that $(p_i^*, \lambda^*, \lambda_0^*)$ is a critical point of the associated Lagrangian. This Lagrangian is

$$L[p_i, \lambda, \mu] = H - \lambda \left[\sum_i p_i(r) \log r_i - \frac{L}{T} \overline{\log R} \right] - \mu \left[\sum_i p_i(r) - 1 \right].$$

The first order conditions, which, given the strict concavity of H, are necessary as well as sufficient to characterize the critical point, imply

$$p_i^*(r) = \exp(-\lambda_0^*) \exp(-\lambda^* \log r_i) = \exp(-\lambda_0^*)\, r_i^{-\lambda^*}.$$

From (3.4) we then obtain $\exp(\lambda_0^*) = Z(\lambda^*)$, resulting in the power law $p_i^*(r) = r_i^{-\lambda^*}/Z(\lambda^*)$. □

Theorem 1 says that the most disorderly mixing of return factors leads to a power law distribution of wealth. To paraphrase Foley's [13] metaphor of markets as probability fields over transactions, the statistical equilibrium wealth distribution defines a probability field over return factors from available combinations of investment opportunities. The most decentralized investment activity of households forms the conceptual basis of the maximum entropy distribution of wealth.

The entropy formalism "hesitates" to assign an enormously large return factor to a portfolio because it thereby reduces the degrees of freedom in the remaining assignments of return factors that have to meet the growth constraint. However, statistical equilibrium does by no means exclude the possibility of such extreme outcomes, it merely attaches a very low probability to them according to the power law distribution. While the statistical equilibrium distribution cannot "name" a particular household in the distribution, it specifies an exact functional relationship that describes the fate of all households above the minimum wealth level.

Corollary 1. *The* number *of theoretically feasible return factors r_i does not influence the functional form of the wealth distribution.*

Proof. The functional form of the first order conditions in Theorem 1 is not affected by the number i of feasible return factors. □

III Characteristic Exponent of the Wealth Distribution in Statistical Equilibrium

Using the statistical equilibrium distribution in the growth constraint (3.3), we obtain a parametric solution for λ

$$\frac{L}{T}\overline{\log R} = -\frac{\partial \log Z(\lambda)}{\partial \lambda} = Z(\lambda)^{-1} \sum_i r_i^{-\lambda} \log r_i.$$

Since, however, the characteristic exponent of a power law carries all relevant information about the statistical properties of the distribution, we are interested in an explicit solution for λ. Thus we consider a continuum of possible return factors $r \in W = [r_{\min}, \infty)$, where $r_{\min}$ designates the *minimum return factor* to which the power law distribution applies. The conceptual tool of a 'zero period' relates return factors and wealth levels in a one-to-one correspondence, hence (minimum) wealth levels and (minimum) return factors should be understood as synonyms. We should keep in mind, though, that wealth levels are of different dimensionality than return factors, raising questions about empirical calibration that we take up in Chapter 5.

The "cost" of gaining analytical tractability through a continuous version of the maximum entropy program comes in the form of an additional measure that will keep the continuous entropy measure invariant with respect to a rescaling of variables. We provide the intuition why such a measure would become necessary in Appendix B, where we also derive the general condition for the invariance of the entropy measure under a logarithmic constraint. Moreover, we argue why introducing the new measure does not alter our results qualitatively.[9] Hence, we continue here with the continuous analog to the familiar discrete entropy program. Unless stated otherwise, all results are derived under the following assumption.

Assumption 1 The power law distribution has finite mean, i.e. $\lambda > 2$.

As before, we denote return factors without the time index simply as r. The maximum entropy program then takes the form

$$\max_{f(r)} H\left[f(r)\right] \equiv -\int_W f(r) \log f(r)\, \mathrm{d}r \tag{3.7}$$

subject to

$$\int_W f(r) \log r \,\mathrm{d}r = \frac{L}{T}\overline{\log R} \tag{3.8}$$

$$\int_W f(r)\, \mathrm{d}r = 1. \tag{3.9}$$

Lemma 1 (Continuous wealth distribution). *The continuous statistical equilibrium distribution of wealth remains a power law,*

$$f^*(r) = \frac{\lambda^* - 1}{r_{\min}^{-\lambda^*+1}} r^{-\lambda^*}. \tag{3.10}$$

[9]The simplest—maybe most elegant—argument why we do not have to introduce the measure is that we derive the distribution of return factors: measuring wealth in, say, euros instead of dollars does not affect the scale of return factors.

Proof. From the Euler-Lagrange equation of the calculus of variations we know that the solution to an extremal problem of the form

$$\int F[x, f(x), f'(x)]\,\mathrm{d}x,$$

where F is a known function, corresponds to $\partial F/\partial f(x) - \frac{d}{\mathrm{d}x}\partial F/\partial f'(x) = 0$. The Lagrangian of the continuous maximum entropy program

$$L = H\left[f(r)\right] - \lambda\left(\int_W f(r)\log r\,\mathrm{d}r\right) - \mu\left(\int_W f(r)\,\mathrm{d}r - 1\right)$$

does not involve $f'(r)$, therefore our problem is analogous to the discrete case and reduces to

$$\frac{\partial}{\partial f^*(r)}\left[-f^*(r)\log f^*(r) - \lambda\left(f^*(r)\log r - \frac{L}{T}\overline{\log R}\right) - \mu\left(f^*(r) - 1\right)\right] = 0,$$

where, as usual, $\partial L/\partial\lambda^* = 0$ and $\partial L/\partial\mu^* = 0$ reproduce the constraints. Again, let $\lambda_0^* \equiv 1 + \mu^*$. Then the first order condition with respect to $f^*(r)$ implies $f^*(r) = r^{-\lambda^*}\exp\left(-\lambda_0^*\right)$. As before, the partition function follows from the natural constraint (3.9),

$$Z\left(\lambda^*\right) \equiv \exp\left(\lambda_0^*\right) = \int_{r_{\min}}^{\infty} r^{-\lambda^*}\,\mathrm{d}r = \left.\frac{r^{-\lambda^*+1}}{1-\lambda^*}\right|_{r_{\min}}^{\infty} = \frac{r_{\min}^{-\lambda^*+1}}{\lambda^*-1}.$$

Substitution completes the proof. □

Lemma 1 enables us derive the central proposition of our theory, which identifies the components that determine the characteristic exponent of the statistical equilibrium distribution of wealth.

Proposition 1 (Characteristic exponent in statistical equilibrium). *The characteristic exponent of the statistical equilibrium distribution obeys*

$$\lambda^* = \left(\frac{L}{T} \overline{\log R} - \log r_{\min} \right)^{-1} + 1. \tag{3.11}$$

Proof. We integrate by parts and use L'Hôpital's rule to obtain

$$\int_W r^{-\lambda^*} \log r \, \mathrm{d}r = \frac{r_{\min}^{-\lambda^*+1}}{\lambda^* - 1} \left[\log r_{\min} + \frac{1}{\lambda^* - 1} \right],$$

which upon substitution in (3.8) yields (3.11). □

Proposition 1 identifies the three determinants of the characteristic exponent: the average growth rate of wealth, the average number of turnovers, and the minimum wealth level to which the power law distribution applies. One particularly nice feature of the statistical equilibrium theory of wealth distribution is its ability to unify economic concepts like turnover activity and average growth with the earlier result of Levy and Solomon [29] on the minimum wealth level as the determinant of the characteristic exponent. Drăgulescu and Yakovenko [10] show that the empirically observed wealth distribution changes its functional form at a particular wealth level.[10]

[10] According to [10], the wealth distribution changes from an exponential shape to a power law for the top five percent of households in the US as well as in the UK.

Though the statistical equilibrium theory presented here concerns only the upper tail, it is noteworthy that it explicitly allows for the dependence of the distribution on the minimum wealth level at which the nature of the distribution changes. What exactly determines the minimum wealth level from a theoretical point of view, however, remains an open question at this point.

IV Interpretation and Policy Implications

The final step in the theoretical analysis of the distribution of wealth has to address the issue of how to connect the entropy-derived power law distribution to the empirically observed distribution. After all, the length of observation L remains arbitrary in the derivation of Proposition 1 and therefore also in (3.10). In order to interpret (3.10) as the actual distribution we have to make one more assumption.

Assumption 2 L and T are both large but their ratio is stable.

Then we can use Lemma 1 as a good approximation to the actual distribution because the arbitrariness in L will not matter. The following remark is a direct consequence of Assumption 2.

Remark 1. The choice of an initial period t_0 has no influence on the power law distribution of wealth.

Changes in the distribution of wealth are reflected through changes in the characteristic exponent; λ provides information about the degree of inequality in the economy, with a higher λ representing a more equal distribution of wealth [1, 25, 42]. Relevant policy prescriptions for lowering the degree of inequality in an economy thus have to address the issue of how to increase the absolute value of λ. From Proposition 1 we can single out the ratio of turnovers per observational period and the average growth of wealth as the economic determinants of the distribution of fortunes. In terms of economic theory, Proposition 1 leads to two trade-offs.

Remark 2. The faster average wealth grows in the upper tail, the more unequal is the maximum entropy distribution of wealth. The influence of turnovers on inequality depends on whether average growth is positive or negative.

Proof. We consider the partial derivatives of the statistical equilibrium characteristic exponent with respect to turnovers and the average rate of growth,

$$
\begin{aligned}
\frac{\partial \lambda^*}{\partial T} &= \frac{1}{T^2}\left(\frac{L}{T}\overline{\log R} - \log r_{\min}\right)^{-2}\overline{\log R}, \\
\frac{\partial \lambda^*}{\partial \overline{\log R}} &= -\left(\frac{L}{T}\overline{\log R} - \log r_{\min}\right)^{-2} L/T.
\end{aligned}
$$

Since turnovers are positive, the partial derivative with respect to average growth is always negative. If the growth rate is positive, more turnovers lead to a more equal distribution of wealth; for a negative growth rate, more turnovers imply a more unequal distribution. We say that the economic problem is not well defined if $T = 0$ or $\overline{\log R} = 0$. □

Regarding the trade-off between growth and wealth inequality, the last remark carries a somewhat similar flavor to Meade's [33] inherent conflict between income equality and productive efficiency in an economy. With respect to turnover activity, the distribution of fortunes will be more equal, *ceteris paribus*, the more agents "reshuffle" their portfolios. Uni-directional statements about wealth inequality in the upper tail are therefore not trivial because they depend on the magnitude of the effects of faster growth versus increased turnover activity.

A remaining and important question regarding the interpretation of the model is whether wealth levels (and their corresponding return factors) are measured in real or nominal terms. Does inflation matter for the degree of inequality in statistical equilibrium?

Remark 3. Inflation, understood as a change of scale, has no distributional consequences in statistical equilibrium.

Proof. Denote the inflation rate during the length of observation by $p \geq 0$. Since the characteristic exponent measures inequality, we have to establish how p affects λ^* in Proposition 1. Suppose the notation there refers to real magnitudes and we adjust return factors for inflation by multiplying them with $(1+p)$; then $\log\left[r_{\min}(1+p)\right] = \log r_{\min} + \log(1+p)$ and, because all portfolios face the same inflation rate, we can also write $\overline{\log R(1+p)} = \overline{\log R} + \log(1+p)$. The term $\log(1+p)$ cancels out in (3.11), leaving the characteristic exponent unchanged. □

We should clarify Remark 3 by re-iterating the crucial assumptions in our proof. First, we assumed that inflation will not affect the turnover rate. Second, we assumed that all economic uses are subject to the same inflation rate, thereby interpreting inflation as a change of scale that affects all portfolios *equally*. Viewed from a different perspective, the latter assumption ensures that the relative location of the minimum return factor adjusts so as to exactly offset the increase in the nominal growth rate. Of course the situation would be quite different if, for whatever reason, inflation changed the 'demarcation line' between the two distributional regimes disproportionately. But regardless of how we define inflation, the statistical equilibrium model has the desirable property that mere changes of scale will have no effect on the characteristic exponent.

V Summary of Theoretical Results

A power law is—in a powerful combinatorial sense—the most likely distribution in a system where the logarithmic mean is the only relevant constraint.

In contrast to the ergodic approach of Levy and Solomon [29], our statistical equilibrium model of wealth distribution determines the characteristic exponent not only from a lower bound but also from two other variables that are economically more relevant: the average rate of growth and the average number of changes in the composition of wealth portfolios. Statistical equilibrium predicts trade-offs between the two variables on one side and distributional equality on the other. The higher the rate of growth, the less equal the power law distribution of wealth. Turnovers, coupled with a positive growth rate of wealth, have an egalitarian influence on the tail distribution of wealth.

So far we only dealt with the theoretical properties of our model. To judge the empirical merit of the statistical equilibrium model, we would have to check whether the equality in (3.11) holds in practice. That will only be possible if we have data for all variables in Proposition 1: the characteristic exponent, the average rate of growth of wealth (within the power law tail, not over the entire population), the minimum return factor, and the average number of turnovers. At least in principle the first three should be observable, whereas privacy issues render observation of turnovers extremely unlikely. If we cannot test the quantitative accuracy of our model directly because we do not observe turnovers, and assuming that we do have in fact information about the other three variables, we should ask which turnover activity our model implies. We take up this question, along with other issues regarding the calibration of the model, in Chapter 5. Meanwhile we focus our attention on the characteristic exponent of the power law distribution in the next chapter, elaborating how to estimate the characteristic exponent from Lorenz data, but also estimating the characteristic exponent directly from a subset of the wealthiest individuals in the United States.

Chapter 4

Empirical Investigations Into the Power Law Tail[1]

The characteristic exponent of the power law distribution should be the 'most robustly observable' parameter among the three needed to calibrate the statistical equilibrium model in the upper tail. We first extract the exponent from Lorenz data in the next section, also briefly sketching the historic magnitudes of the exponent for a small sample of western countries. In the second section, we estimate the exponent directly from seven recent *Forbes 400* lists of the wealthiest US citizens, and finally we use the same data set to conduct a simple mobility analysis.

---

[1] **The sections dealing with estimation of the characteristic exponent from the *Forbes 400* data, as well as the mobility analysis, have been co-authored with Carolina Castaldi.**

I Characteristic Exponent and Lorenz Data

Quite often wealth data are reported in Lorenz form, i.e. $x\%$ of the population owns $y\%$ of total wealth. We can infer the characteristic exponent if we know the wealth share S of a top percentile P with the following lemma.

Lemma 2. *If wealth is distributed according to a power law, then one point (P, S) on the Lorenz curve enables us to determine the characteristic exponent of the distribution as*

$$\alpha = 1 + \left(1 - \frac{\log S}{\log P}\right)^{-1}. \tag{4.1}$$

Proof. Set minimum wealth to unity and consider the probability density function $n(x)$ of wealth x, $n(x) = cx^{-\alpha}$, with c as an appropriate normalizing constant and w as the (unknown) wealth level corresponding to the pair (P, S); then P and S are defined by the ratios

$$P \equiv \frac{\int_w^\infty n(x)\,\mathrm{d}x}{\int_1^\infty n(x)\,\mathrm{d}x} = w^{-\alpha+1}, \text{ and } S \equiv \frac{\int_w^\infty xn(x)\,\mathrm{d}x}{\int_1^\infty xn(x)\,\mathrm{d}x} = w^{-\alpha+2},$$

provided Assumption 1 holds. Since empirically observed wealth is finite, this assumption is not restrictive. Given P and S, we solve the system of two equations in two unknowns for α to obtain the above statement. □

To avoid confusion, we recall that the characteristic exponent derived here refers to the density and not the (inverse) distribution function, which

P	.0001	.001	.002	.005	.01	.03	.05
S^c	.167	.232	.261	.310	.359	.466	.534
$\alpha - 1$	1.241	1.267	1.275	1.283	1.286	1.278	1.265

Table 4.1: Critical values for the propagation of errors.

is usually cited in the literature. It is readily verified that the exponent of the distribution function will be $\alpha - 1 = (1 - \log S/\log P)^{-1}$.

We could calculate error bounds for α from (4.1) if we knew the variance σ^2 in the measurement error of S by simply calculating α for $S \pm \sigma/2$, holding P constant. Alternatively, we can use the *law of propagation of errors* to approximate the effect of a mismeasurement in S. The law of propagation of errors states that, for $\sigma \ll S$, the extent of the error bounds will be

$$\left|\frac{\partial \alpha}{\partial S}\right| \sigma = \left(1 - \frac{\log S}{\log P}\right)^{-2} \frac{\sigma}{|\log P| S} \quad \text{for } P, S \in (0,1). \tag{4.2}$$

As the name suggests, the law of propagation of errors allows us to determine how much of the measurement error in S is 'passed through' to α since $|\partial\alpha/\partial S| < 1$ means a less than proportionate increase in the error bounds of α compared to σ; the opposite is true for $|\partial\alpha/\partial S| > 1$. The following table shows the numerically computed solutions for the 'critical values' S^c where $|\partial\alpha/\partial S| = 1$. We chose P based on the data in Table 4.2. For a given P, equation (4.1) shows α as a strictly convex function of $S > P$. Therefore, measurement errors will be magnified in α if $0 < S < S^c$, while the reverse is true if $S^c < S < 1$.

Country	Year	Top households P	Wealth share S	$\alpha - 1$
Sweden	1920	0.0001	0.090	1.354
		0.0010	0.240	1.260
		0.0100	0.500	1.177
	1975	0.0010	0.060	1.687
		0.0020	0.080	1.685
		0.0050	0.125	1.646
		0.0100	0.170	1.625
		0.0200	0.240	1.574
	1983	0.0010	0.080	1.576
		0.0020	0.100	1.589
		0.0050	0.145	1.573
		0.0100	0.195	1.550
		0.0200	0.260	1.525
	1975	0.01	0.160	1.661
Belgium	1969	0.01	0.280	1.382
Canada	1970	0.01	0.196	1.548
Denmark	1973	0.01	0.250	1.431
Germany	1973	0.01	0.280	1.382
United States	1972	0.01	0.250	1.431
	1983	0.01	0.34	1.306
	1989	0.01	0.39	1.257
United Kingdom	1923	0.01	0.61	1.120
		0.05	0.82	1.071
	1929	0.01	0.56	1.144
		0.05	0.79	1.085
	1975	0.01	0.24	1.449
		0.05	0.44	1.378
	1980	0.01	0.23	1.469
		0.05	0.43	1.392
France	1977	0.01	0.19	1.564
		0.05	0.45	1.363
	1986	0.01	0.26	1.413
		0.05	0.43	1.392

Table 4.2: Lorenz data are taken from [46] and the papers collected in [45]; different data sources are indicated by a horizontal line.

Table 4.2 presents Lorenz data for different countries at different points in time, taken from Wolff [45, 46], together with the characteristic exponents of the distribution function that we calculated using Lemma 2. The results are encouraging: within the upper tail—typically the top one to three percent of households—the functional form of a power law seems consistent with the data. (Particularly since the deviations that occur do so where the propagation of errors will be pronounced.) Moreover, variation across time is much more pronounced than variation across countries, all this in spite of the fact that international wealth data are usually not measured in the same fashion [45] and in spite of the rather coarse nature of wealth percentiles. Wealth inequality in the upper tail was significantly higher at the beginning of the twentieth century compared to the 'Golden Age,' i.e. the decades between World War II and the collapse of Bretton-Woods.

II The *Forbes 400* List

Since 1982 *Forbes* has annually published the *Forbes 400* list of the wealthiest US individuals in its print edition, and also posted the list free of charge on its web site for the years 1996-2002.[2]

To be considered for the list, individuals must be US citizens and should own more than $550 million. In order to estimate net worth, a dozen-strong staff collects and updates publicly available information regarding stakes in

[2] http://www.forbes.com/2002/09/13/rich400land.html

public stock, real estate, and unincorporated businesses. Yet the data need to be gauged with the appropriate degree of skepticism. We often hear—*Forbes* being no exception—that, say, Bill Gates' fortune is made up almost exclusively from his Microsoft shares. It is by no means clear why he should not have diversified his position, nor do we get to know on which evidence such a claim rests. That said, the *Forbes* staff calculates wealth from public stock holdings by multiplying share prices with the number of shares known to be in the agents' possession; wealth from real estate is calculated on the basis of publicly known square feet-ownership multiplied by local market rates; wealth from unincorporated businesses is estimated by assuming that the enterprise operates under the same margins as a publicly traded company in the same sector; finally, projected tax liabilities are subtracted to obtain an estimate of net worth.

On the other hand, more reliable wealth data from estate tax records and household surveys are particularly inaccurate in the tails of the distribution [7], so the *Forbes 400* represents valuable information on the upper tail in spite of its imperfect measurement. Since Pareto's initial discovery of a power law distribution there has been ongoing controversy about the functional form of wealth distributions [6, 35], but also a consensus that the upper tail of wealth distributions indeed obeys a power law [1, 7, 10, 30, 35, 42]. To check whether the *Forbes 400* list confirms the claim of a power law distributed upper tail, we plot the inverse cumulative distribution function for the original lists and for a modified set where we grouped the wealth of Sam

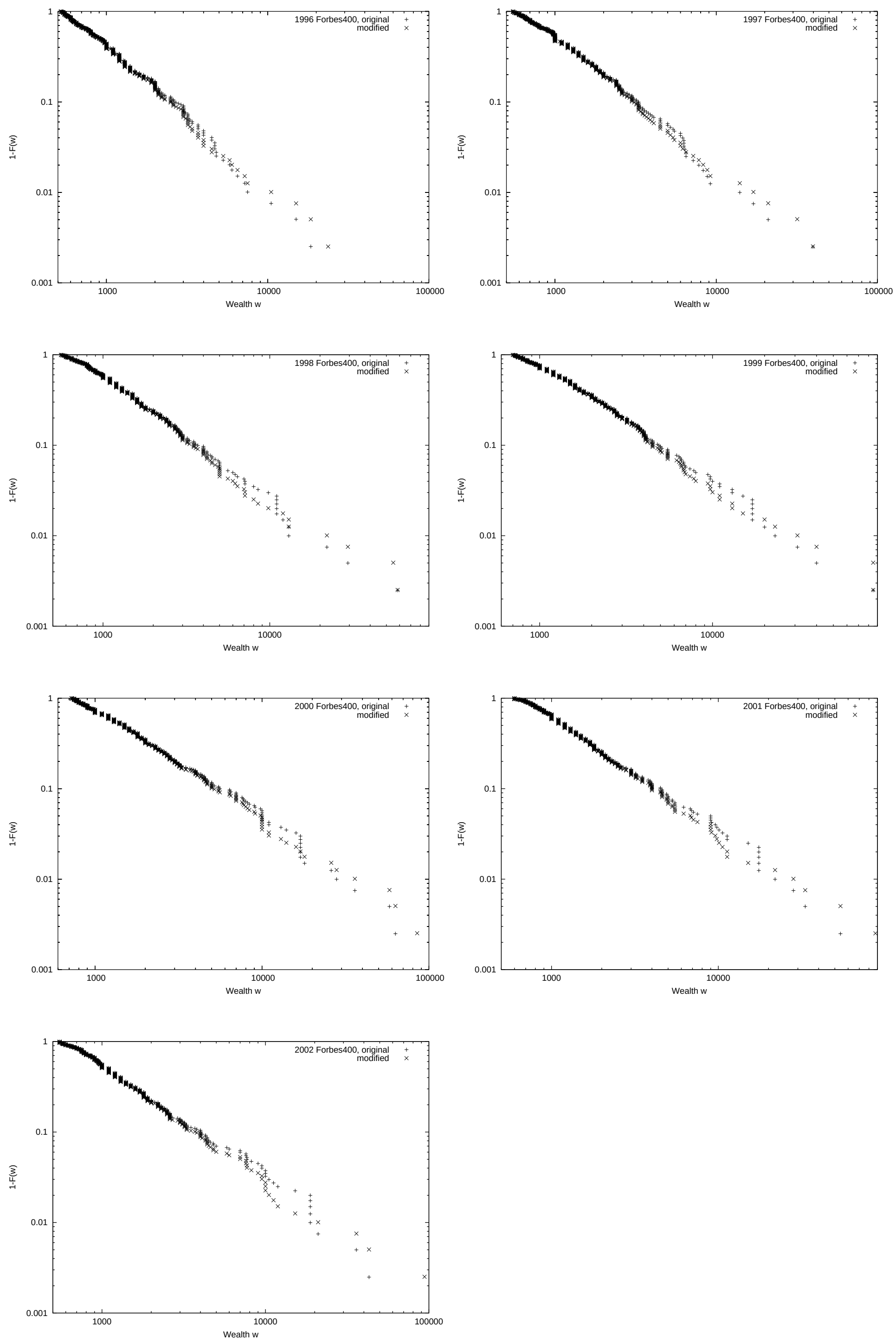


Figure 4.1: Inverse cumulative distribution of personal wealth from the *Forbes 400* list, plotted on double-logarithmic scale for each of the years 1996–2002.

Walton's five heirs to his retail empire into one observation because they always have identical net worth—an indication that they share equal stakes in a *single* portfolio, which is the relevant information for our theoretical framework. We also removed a few minor inconsistencies, e.g. if a fortune shows up in one year as a family item, then as multiple individual items, and in later years again as a single family item.

If wealth is distributed as a power law, the inverse cumulative distribution will be a straight line on double-logarithmic scale. Casual inspection of Figure 4.1 certainly confirms the conventional wisdom, and it is much more likely that observed deviations are due to measurement error rather than the other way around, i.e. that measurement error creates a strong but spurious impression of a power law.

III Estimation of Power Law Exponent

We will estimate the characteristic exponent for each of the years 1996–2002 from the modified *Forbes 400* lists, where wealth levels are arranged in descending order. (A form that is also referred to as *reverse order statistics.*) The statistical equilibrium model predicts the probability density $f(w) = (\alpha - 1)c^{\alpha-1}w^{-\alpha}$ for wealth w, where we denote the minimum wealth level by $c = \min\{w_i : i = 1, \ldots, n\}$ to explicitly distinguish it from the minimum return factor $r_{\min}$.[3] The log-likelihood function for the reverse order

[3]Notice that we have consistently labeled the characteristic exponent as α and not as λ in this chapter because λ is the exponent of the distribution of return factors. Chapter 5

statistics of size n is

$$\log L = n\left[\log(\alpha - 1) + (\alpha - 1)\log c\right] - \alpha \sum_{i=1}^{n} \log w_i. \tag{4.3}$$

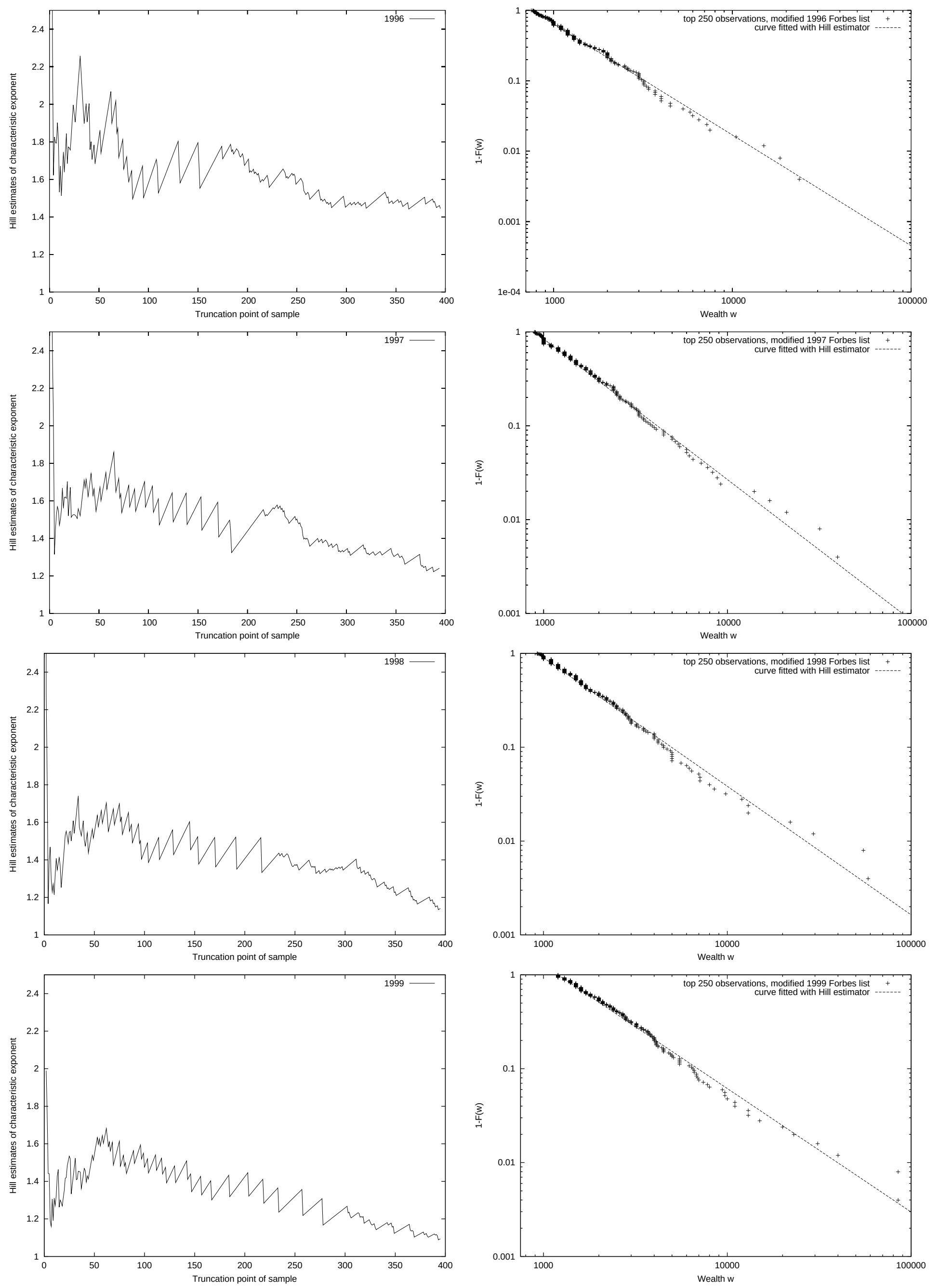
Defining $\phi = \alpha - 1$, we obtain the maximum likelihood estimator for the characteristic exponent ϕ of the cumulative distribution as

$$\hat{\phi} = \left(n^{-1} \sum_{i=1}^{n} \log w_i - \log c \right)^{-1}, \tag{4.4}$$

which is the well-known Hill estimator [18]. The Hill estimator depends crucially on the "cut-off" parameter n and is particularly sensitive to the term $\log c$ because c depends directly on the chosen sample size n. In order to decide on the 'optimal' cut-off point in our samples we employ a tool that is known as the Hill plot, graphing the characteristic exponent $\hat{\phi}$ as a function of the used sample size n. In the case of very accurate data and, of course, provided the distribution is a power law, the Hill plot typically shows a few initial oscillations that quickly disappear and settle into a horizontal line at the 'true' value of ϕ [9].

Unfortunately, as we see in the left panel of Figure 4.2, the Hill plots for the *Forbes 400* lists do not show any tendency to settle down at a specific value of $\hat{\phi}$ as the sample size increases but instead oscillate over the entire order statistics in a given year. There are two possibilities why this should be the case. Either the distribution is not a power law in the first place, or

takes up the issue in more detail.


1996
1997
1998
1999
Hill estimates of characteristic exponent
Truncation point of sample
top 250 observations, modified 1996 Forbes list
top 250 observations, modified 1997 Forbes list
top 250 observations, modified 1998 Forbes list
top 250 observations, modified 1999 Forbes list
curve fitted with Hill estimator
1-F(w)
Wealth w

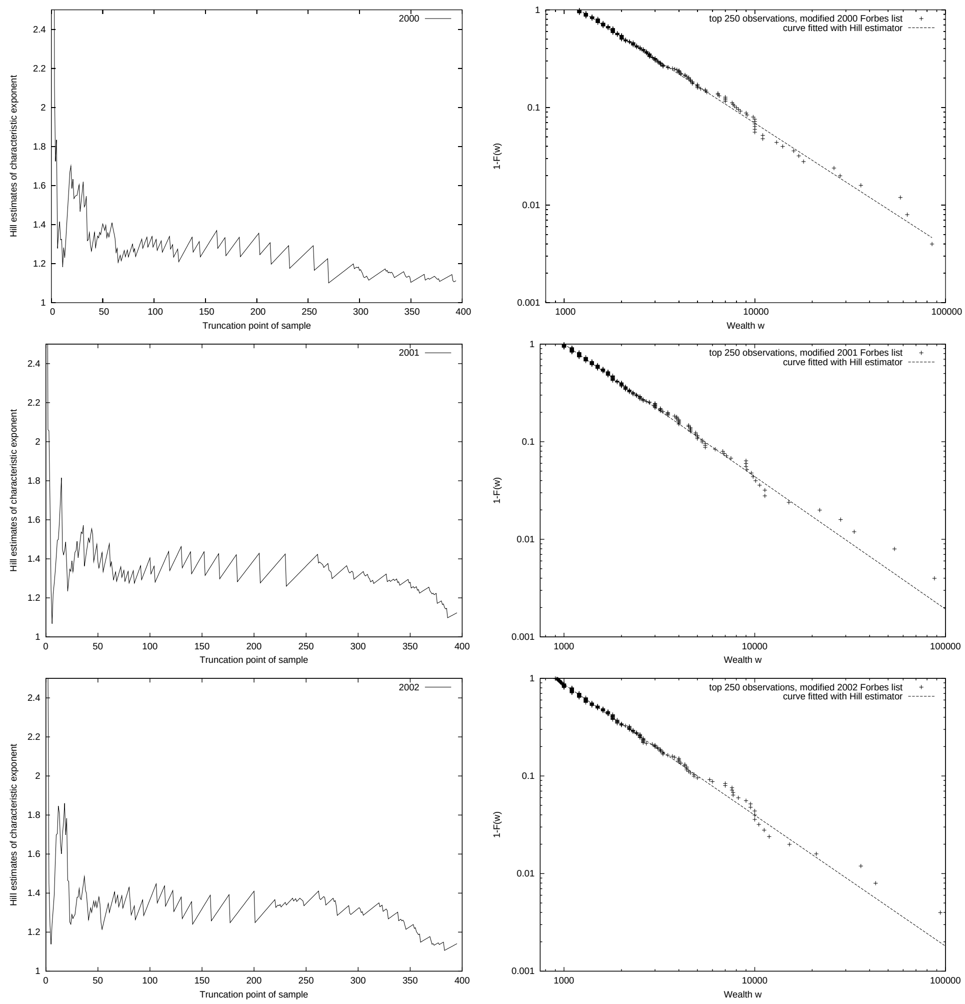


Figure 4.2: (Continued from the previous page.) The left panel shows the Hill plots of the modified *Forbes 400* list for the years 1996–2002. The right panel plots the inverse cumulative distribution of the top 250 observations in the corresponding year for the modified lists on double-logarithmic scale, as well as the curve fitted from the Hill estimator of the characteristic exponent $\hat{\phi}$ at $n = 250$.

Year	1996	1997	1998	1999	2000	2001	2002
$\hat{\phi}$	1.573	1.497	1.369	1.315	1.262	1.358	1.341
	(.100)	(.095)	(.087)	(.083)	(.080)	(.086)	(.085)

Table 4.3: Characteristic exponent and standard error, top 250 observations of the *Forbes 400*. Standard errors reported in parentheses.

measurement error prevents the discovery of the 'true' exponent. Since our theoretical model strongly suggests a power law distribution, and because we know about the imperfect compilation of the data, we side with the latter explanation.

Nevertheless, the Hill plots reveal two interesting features. First, there is a pronounced decrease at the right end of the curves. Second, the curves show a stable region where the graphs oscillate around a stable level before falling off to the right. Therefore, the appropriate sample size seems to be the one that separates the stable region from the decreasing tail in the Hill plots.[4] Instead of elaborating—ultimately arbitrary—criteria for choosing the 'best' sample size for each individual year from the Hill plots, we prefer to select a uniform cut-off point for all seven years. Based on graphical inspection, we chose $n = 250$ to estimate the characteristic exponent and show the resulting fit in the right panel of Figure 4.2.

We report the values of the estimated characteristic exponent $\hat{\phi}$ in Ta-

[4] One could speculate that measurement error is more pronounced for the lower part of the lists because there is less discriminating information about half-billionaires than about the top ranked agents, or possibly less effort is devoted to the compilation of 'lower' ranks in the list.

ble 4.3. It has been shown [8, 17] that $\sqrt{n}(\hat{\phi} - \phi) \xrightarrow{d} N(0, \phi^2)$ such that we can estimate the standard error of $\hat{\phi}$ as $\hat{\phi}/\sqrt{n}$. We also performed a Monte Carlo simulation to check whether the standard error of the estimator conforms to the asymptotic expression at sample size 250.[5] The results were virtually identical and Table 4.3 reports the estimated standard error.

IV Wealth Mobility

Inequality measures the dispersion in the distribution of wealth at a given point in time. Mobility, on the other hand, measures how agents move within the distribution over time. The question of interest is whether inequality and mobility are clearly correlated, i.e. does higher inequality go along with higher (or lower) mobility, or are the two measures possibly even uncorrelated?

The primary purpose of this section is to obtain a rough-and-ready impression of mobility and to relate it to inequality, not to provide a rigorous measurement of mobility. Hence we neglect issues of positional versus non-positional measures of mobility [11] as well as issues regarding the desirable axiomatic minimum requirements of such measures [40].

The nature of the *Forbes 400* data enables us to provide particularly simple, yet informative, measures of mobility among the wealthiest US citizens. We are interested both in assessing how volatile or persistent membership

[5] We are happy to provide the code upon request.

Number of years	1	2	3	4	5	6	7
Frequency (total=685)	149	83	68	95	41	44	205

Table 4.4: Frequency table for the number of years that individuals stay in the *Forbes 400* list.

in this privileged club is, as well as in measuring the relative mobility within the list. Regarding membership over the seven years, we observe that 205 individuals constitute the core group that is present in all years. A total of 685 people (or families) appear at least once, while 149 of them appear only once. We calculated the frequency distribution for the length of stay in the list, shown in Table 4.4. The average length of stay between 1996 and 2002 is four years.

In addition, we traced year-by-year entries—i.e. the number of people that appear in a given year but were not there in the previous one—as well as the year-by-year exits.[6] We define *rank mobility* as the average absolute change in rank of agents between consecutive years of the list. Notice that in calculating rank mobility we only consider agents that appear in the list from one year to the next, thereby avoiding the assignment of fictitious rank for people who are entries or exits during that period. The results, reported in Table 4.5, indicate a considerable churn rate in the Richest Club with an average of 62 year-on-year entries and exits,[7] and an average annual rank

[6]Note that the difference between number of entries and number of exits reported in Table 4.5 is exactly the difference in the sample size between the two consecutive years.

[7]Note that the difference between number of entries and number of exits is exactly the difference in the sample size between the two consecutive years.

Year	Entries	Exits	Rank mobility
1996-97	69	71	32.50
1997-98	55	55	31.10
1998-99	70	71	35.81
1999-00	77	68	33.35
2000-01	55	55	40.96
2001-02	47	47	32.38

Table 4.5: Entry and exit, and average absolute change in rank between consecutive years.

mobility of 34. It is quite interesting that rank mobility and entry-and-exit do not show a clear positive correlation—the period 1999-2001 is a powerful illustration that mobility and entry-and-exit can be negatively correlated!

Unfortunately, we have no way of knowing whether the proportion of entries and exits or the magnitude of rank mobility that we observed for the very small subset of the richest four hundred are also representative of the entire power law tail. Be that as it may, the fact remains that within the *Forbes 400* subset we cannot observe a clear pattern between inequality and mobility, at least not in the short run, as a comparison of the exponents in Table 4.3 with the mobility measures in Table 4.5 illustrates.

A final remark concerns the average growth of wealth in the subset. Neglecting entry-and-exit, we simply calculated the growth rate of 'total' wealth from the top four hundred observations in each of the seven years. The annual growth rates of wealth between 1996 and 2002 were 28%, 17%, 37%, 6%, −21%, and −8%. Obviously, the pattern in the average growth of wealth cannot account for the inequality pattern in Table 4.3. The reason, of course, is that the average growth of wealth is only one variable that influences inequality through the characteristic exponent. The remaining task, taken up in the next chapter, is to operationalize the model with the ultimate goal of obtaining estimates for unobservable turnover activity.

Chapter 5

Calibration Issues

In the last chapter we estimated the characteristic exponent of the distribution of wealth, using two different methods and various sources. But why or how does the empirically observed exponent α relate to λ, the exponent of the distribution of return factors developed in Chapter 3? We have not yet discussed how to determine the minimum return factor, either. As the following sections demonstrate, the calibration of the theoretical model is everything but a trivial exercise.

I How Does the Theoretical Exponent Relate to the Empirically Observed One?

Our theoretical model defined wealth levels to be proportional to return factors because we assumed all portfolios to start out with the same wealth level. Hence the random variable determining the observed wealth level is the return factor in (3.1). Given equation (3.10), the probability density of return factors in statistical equilibrium, and given the definition (3.1), we can use an elementary theorem of probability theory to find the density of wealth levels. According to the theorem, see for example [39, p. 229, Theorem 7.1], the density function for wealth levels will be given by[1]

$$f_w(z) = f_r\left(g^{-1}(z)\right)\left|\frac{\mathrm{d}g^{-1}(z)}{\mathrm{d}z}\right|, \tag{5.1}$$

where $g^{-1}(z)$ denotes the inverse functional relationship between wealth levels and return factors, i.e. $g^{-1}(z) = (z/w_0)^{1/T}$. To simplify things, we assume that the mythical initial wealth level equals unity.

Furthermore, the turnover parable implies that T is very large, as we explicity pointed out in Assumption 2 on page 31. Thus we can argue that $r_{\min}$ has to equal unity, since otherwise the definition in (3.1) implies either a minimum wealth equal to zero (if $r_{\min} < 0$) or equal to infinity (if $r_{\min} > 0$), both of which have to be ruled out. That leaves us with a much simpler expression for the distribution of wealth levels. Now that $g^{-1}(z) = z^{1/T}$ and $f(r) = (\lambda - 1)r^{-\lambda}$, we have

$$f_w(z) = \xi z^{-(\xi+1)}, \quad \text{where} \quad \xi \equiv \frac{\lambda - 1}{T}. \tag{5.2}$$

[1] To be more precise, the application of the theorem presupposes that return factors are a continuous random variable and wealth levels are a continuous differentiable and monotonic function—unproblematic in the case of our support.

The characteristic exponent of the (inverse) cumulative distribution of wealth will therefore be ξ, corresponding to our 'observed' or estimated ϕ from the previous chapter. Recalling Proposition 1, with $r_{\min} = 1$, turnovers per observational period are equal to $T/L = \overline{\log R}(\lambda - 1)$. But from the definition of ξ in (5.2) we also know that $\lambda - 1 = (\xi + 1)T$, which finally leaves us with the expression

$$\frac{1}{L} = (\xi + 1)\,\overline{\log R}. \tag{5.3}$$

While the right-hand side contains only observables now, it is also true that turnovers have canceled out—the baby has been thrown out with the bathwater! Starting from an alternative formulation of the basic theory, we attempt a remedy in the next section.

II An Alternative Statistical Equilibrium Model of the Power Law Tail

The pre-analytical vision remains the same as in Chapter 3. All portfolio's start out with the same wealth level, for convenience set to unity. At the end of the turnover period, wealth will be distributed according to the maximum entropy principle under a logarithmic constraint. But instead of conceptualizing wealth as a return factor, we use continuous compounding and define the *rate* of return per turnover $r = \log w$. The one period constraint on wealth levels remains logarithmic, so that the mean constraint on rates of return is arithmetic.

Turning the crank, we obtain an exponential distribution of rates of return that corresponds to a Pareto distribution of wealth levels, as we already have demonstrated in (2.7) and (2.8) on page 12. Suppose the observed minimum wealth level is less than the initial wealth level, $m < w_0$. That means that the corresponding minimum return factor is negative, $\log m \equiv r_m < 0$. Maximizing (2.1) on the support $R = [r_m, \infty)$ subject to (2.3) yields the distribution of return factors

$$f(r) = \mu \mathrm{e}^{-\mu(r-r_m)}, \tag{5.4}$$

where μ is the Lagrange multiplier associated with the arithmetic mean constraint (2.3), and obeys

$$\mu = 1/(\bar{r} - r_m)\,. \tag{5.5}$$

What happens after T successive turnovers have taken place, i.e. what is the distribution of $\tilde{r} = \sum_{t=1}^{T} r_t$? Assuming that each turnover follows the same one-period statistical equilibrium distribution, the sum of T exponentially distributed random variables with parameter μ will be a random variable that is Gamma distributed with parameters μ and T [39, p. 266, Proposition 3.1]. In other words, T convolutions of an exponential distribution with parameter μ result in the Gamma distribution

$$f(\tilde{r}) = \frac{\mu^T}{\Gamma(T)} \tilde{r}^{T-1} \mathrm{e}^{-\mu \tilde{r}} \,. \tag{5.6}$$

But what does a Gamma distribution of returns imply about the tail behavior of the distribution of wealth levels $\tilde{w} \equiv \mathrm{e}^{\tilde{r}}$ after T turnovers? Applying the previously used theorem for the density of a function of a random variable, we have

$$f(\tilde{w}) = \tilde{w}^{-(\mu+1)} \frac{\mu^T}{\Gamma(T)} (\log \tilde{w})^{T-1} \,. \tag{5.7}$$

In a first approximation, we argue that the asymptotic slope of the density of period T wealth will be, on logarithmic scale,

$$\begin{aligned} \lim_{\tilde{w} \to \infty} \frac{\partial \log f(\tilde{w})}{\partial \log \tilde{w}} &= \lim_{\tilde{w} \to \infty} \frac{T-1}{\log \tilde{w}} - (\mu + 1) \\ &= -(\mu + 1) \,. \end{aligned} \tag{5.8}$$

In a (probably too rough) second approximation, we argue by analogy to the 'pure' power law that the absolute slope of the cumulative distribution will be equal to the empirically observed one, $\mu \approx \hat{\phi}$. Finally, employing the same logic as in Chapter 3, we know that $T\,\bar{r} = L\,\bar{R}$, where R is now the *rate* of return per unit time, and L still designates the elapsed time since the process started. Therefore $T/L = \bar{R}/\bar{r}$, but as before we cannot observe

Year	$\bar{R}$	R_m	T/L
1996-97	28%	-59%	1.31
1997-98	17%	-47%	0.88
1998-99	37%	-47%	1.11
1999-00	7%	-61%	0.85
2000-01	-21%	-90%	0.94
2001-02	-8%	-68%	0.81

Table 5.1: Implied average turnover activity calculated from equation (5.10).

$\bar{r}$. We know, however, from (5.5) that $\bar{r} = r_m + 1/\hat{\phi}$ and thus

$$T/L \approx \bar{R}/(r_m + 1/\hat{\phi})\,. \tag{5.9}$$

In contrast to the previous section, we are now able to derive an expression for turnover activity that depends on the same observables, but where turnovers per observational period do not cancel out. The final step is to express the unobservable minimum return per turnover r_m in terms of the observable minimum return per calendar time R_m. But the convolutions imply that $L\,R_m = T\,r_m$, and we are finally left with

$$T/L \approx (\bar{R} - R_m)\hat{\phi}. \tag{5.10}$$

Conceptually, turnovers per observational period are a dimensionless positive constant implied by the theoretical model. T/L is always positive since $\hat{\phi}$ is always positive (in fact greater than unity), and because it also must be true that $\bar{R} > R_m$. Table 5.1 reports the turnover activity

Year	Entries	Exits	Rank mobility	$\hat{\phi}$	T/L
1996-97	69	71	32.50	1.497	1.31
1997-98	55	55	31.10	1.369	0.88
1998-99	70	71	35.81	1.315	1.11
1999-00	77	68	33.35	1.262	0.85
2000-01	55	55	40.96	1.358	0.94
2001-02	47	47	32.38	1.341	0.81

Table 5.2: Mobility, inequality, and turnover activity in the *Forbes 400*.

calculated from the seven consecutive *Forbes 400* lists studied in Chapter 4.

We conclude the chapter with a reproduction of the data from Tables 4.3, 4.5, and 5.1, to better visualize the movements in inequality, mobility, and turnover activity. A comparison of the series describing wealth inequality, turnover activity, and wealth mobility shows that mobility and turnover activity are strongly positively correlated, with increases (decreases) in turnover activity always going along with higher (lower) mobility. Conversely, turnover activity and wealth inequality are also positively correlated, with only one exception during the period 1998-99, which is also true of the relationship between inequality and mobility. The increase in inequality during that period is caused by the enormous growth rate of wealth that cannot be compensated for by the increase in turnover activity.

Chapter 6

The Economics of Gamma Distributed Wealth

We conclude our investigations into the distribution of wealth by examining the necessary economic ingredients for a statistical equilibrium theory of the left part of the distribution. A recent article on the UK distribution of wealth claims to have observed an exponential (inverse) cumulative distribution for ninety-five percent of the population with positive wealth, together with a power law tail that covers the top five percent of wealth holders [10].

The authors admit, however, that they are aware of evidence that points towards a Gamma distribution but—out of convenience—only estimate the parameter μ (see page 12) for the exponential law instead of fitting the additional parameter of the Gamma distribution. As we have shown in Chapter 2, the inverse cumulative distributions of the exponential and Gamma laws look extremely similar so that a relatively good fit of an exponential law is not surprising, even if the 'true' distribution is Gamma. Since the Gamma and exponential laws have very similar cumulative distributions, we would expect a better fit to data from the Gamma law, simply because of the additional degree of freedom. Our ambition in this final chapter is not to develop a statistical test that discriminates between the two distributions based on a data sample, but rather to argue which of the two distributions is more likely from the viewpoint of statistical equilibrium theory.

As usual, our analysis starts from the stylized fact concerning the composition of wealth portfolios that make up the part of the distribution we want to investigate. In the left part of the wealth distribution, the net position in owner-occupied housing, deposits, and life insurance and pension plans are the main assets of agents. We assume, in contrast to our previous analysis where we only cared about the uses of wealth and not its sources, that these assets are financed from earned income that will mostly flow from wages and salaries. Other possible sources are government transfer payments, rents, and profits arising from unincorporated businesses and financial assets. Regardless of the source of income, and this is the crucial point, we assume that additions to wealth will *not* be re-invested in the way we envisaged for the very wealthy in Chapter 3.

Instead, the existing level of wealth will be augmented by additions out of current income such that wealth remains roughly proportional to income. It turns out the crucial theoretical argument why the statistical equilibrium distribution should be a Gamma instead of an exponential law boils down to the definition of what 'roughly' should mean in a 'more exact' context.

Suppose, for the time being, that for all individuals $j \in \{1, \ldots, n\}$, $n < \infty$, who accumulate wealth in such fashion $y_j^e(t)$ designates the disposable income from source $e \in \mathbf{E} = \{1, \ldots, E\}$ at time period t. Moreover, if $w_j(t)$ denotes the wealth of agent j in period t then the change in wealth between periods will depend on how much of the agent's income has been 'saved' from the different sources

$$\Delta w_j(t) = \sum_{e \in \mathbf{E}} s_j^e(t) y_j^e(t) \qquad \forall j \in \{1, \ldots, n\}, \tag{6.1}$$

where $s_j^e(t)$ represents the proportion of income from source e that agent j utilizes to augment wealth at time t. The stock of wealth $w_j(\tau)$ that agent j has accumulated up to period τ, at which we observe the personal wealth distribution, will depend on her accumulation behavior s_j^e, her fortunes in the (labor) market y_j^e, and of course on the number of periods $T_j \equiv T_j^e + T_j^r$ in which she has an income either earned during T_j^e periods of working life or flowing during the T_j^r periods after retirement (e.g. when the agent made provisions through pension and life insurance schemes)

$$w_j(\tau) \equiv \sum_{t=1}^{T_j} \Delta w_j(t) \qquad \forall j \in \{1, \ldots, n\}. \tag{6.2}$$

Thus $s_j^e(t)$ should not be understood in the classical sense of a 'saving propensity' since we want to allow for a negative $s_j^e(t)$, for example to represent the decrease of wealth that occurs when an agent retires and spends her previously accumulated pension income for consumption.[1]

As before, the enormous amount of information prevents us from explaining the empirically observed distribution of personal wealth by tracing the destiny of all agents. So, by what should hopefully be standard fare at this point, we start from the macroscopic constraint on the average wealth accumulated by the population at time τ

$$\overline{w_\tau} = n^{-1} \sum_j \sum_t^{T_j} \sum_{e \in \mathbf{E}} s_j^e(t) y_j^e(t) \,. \tag{6.3}$$

We denote the set of theoretically possible wealth levels by $W = [0, m)$, where m is the wealth level that separates the empirically observed exponential and power law regimes.[2] Let $i \in \{1, \ldots, z\}$ run over the set of discrete wealth levels $w_i \in W$ and let n_i be the fraction of agents with wealth w_i. Then it must also be true that

$$\overline{w_\tau} = \sum_i^z w_i \frac{n_i}{n} \equiv \sum_i^z w_i p_i \,, \tag{6.4}$$

and to ensure that all agents are assigned to a wealth level again, the natural constraint $\sum_i p_i = 1$ must hold as well. The wealth distribution that allows

[1] Consumption should be understood here in the broad sense of an economic use not included within the five categories we introduced in Chapter 1 that count as personal wealth.

[2] The exponential law cannot account for individuals with negative wealth. Thus, for the sake of completeness, we should also have a constraint that prevents negative wealth in (6.2). Similar to the conventional life-cycle model, it would boil down to the postulate that life-time earnings should be greater than or equal to life-time 'consumption.'

for the largest number of individual destinies and behaviors consistent with the observed average stock of wealth in W will be given by the solution to the maximum entropy program that maximizes informational entropy subject to (6.4) and the natural constraint. Not surprisingly, as in the continuous case (2.7), the resulting distribution will be of an exponential type also known as the canonical Gibbsian distribution [13]

$$p_i = \frac{\mathrm{e}^{-\mu w_i}}{Z(\mu)}, \tag{6.5}$$

with the partition function $Z(\mu) \equiv \sum_i \mathrm{e}^{-\mu w_i}$. It is quite straightforward to show that on the continuous support $W = [0, m)$ the relationship between μ and $\overline{w_T}$ is given by $\overline{w_T} = 1/\mu - m/(\mathrm{e}^{m\mu} - 1)$, which for a large enough m approximates the familiar result $\mu = 1/\overline{w_T}$.

If wealth in the left part of the distribution was determined by accumulated earnings alone, we would favor the exponential distribution from a theoretical point of view. But we know that the uses of wealth chosen by the majority of agents are owner-occupied housing, deposits, and life insurance and pension plans, all of which earn a rate of return.[3] In principle, we should be able to observe the average rate of return that the aforementioned uses earn over a certain period, analogously to the model of Chapter 3, but without the turnover concept—after all we are arguing that the majority of agents does not perpetually re-invest returns and therefore does not change

[3]Returns are possibly negative, particularly regarding the value of owner-occupied real estate.

the composition of their portfolios very much. Dropping the turnover concept, we might as well consider the 'pure' logarithmic mean without time dimension as a valid approximation to the return on wealth.

As we have demonstrated in Chapter 2, the statistical equilibrium distribution under simultaneous arithmetic and logarithmic means will be the Gamma distribution. Hence, a theoretically founded case can be made in favor of the Gamma law if we agree that (at least part of) accumulated wealth from lifetime earnings yields a return.

Obviously, our argument in favor of the Gamma law has not been developed in a formally satisfactory way, yet the intuition should be quite clear. What remains to be done is to clarify the relationship between the parameters of the Gamma distribution and the observed arithmetic and logarithmic means of wealth. Such an endeavor is complicated by the fact that we cannot obtain closed-form solutions that relate moments to parameters [22], making calibration of a model along the lines proposed here at least as difficult as in the case of the model for the power law tail.

Chapter 7

Concluding Remarks

The distribution of wealth is the result of a highly complex set of interactions among economic agents and yet it displays robust functional regularities across space and time. The upper tail of the wealth distribution, covering about five percent of the population, obeys a power law, while the remaining majority of agents with positive wealth levels can be described by a Gamma distribution.

Statistical equilibrium views these phenomena as reflecting two distinct processes in the accumulation of wealth, the power law tail being caused by the perpetual re-investment of returns from the different economic uses of wealthy portfolios, and the Gamma distribution following from a life-cycle type of savings behavior and the presence of retirement provisions. It is noteworthy from a formal point of view that the power law distribution implies the complete absence of any aggregate constraints other than a logarithmic mean. Put differently, the fact that we observe a power law distribution tells us by itself about the inherently dynamic character of the underlying process (since the logarithmic mean reflects a growth constraint) that is not in any way constrained by the stock of wealth at a particular point in time. Contrary to the jargon and intuition of thermodynamics or statistical physics, we detect the absence of a conservation principle in the personal wealth of the mighty rich.

The analogy to the first law of thermodynamics, the conservation of energy, might apply in large part to the wealth of the vast majority of economic agents but if it was to apply to the very wealthy as well, it would simply be impossible to observe the power law tail. With respect to the existing theories of power law distributed phenomena [2, 3, 4, 15, 29, 31, 32, 37], the most significant contribution of the statistical equilibrium approach is probably the explicit acknowledgment of the absence of any aggregate constraint other than the logarithmic one.

The strength of statistical equilibrium theory is its formal underpinning that allows robust statements about the aggregate constraints that shape the microscopic outcomes of a system. At the same time its strength is also its major drawback because we cannot identify individual destinies nor the microeconomic forces—for instance inheritance, investment behaviors, or personal skills—that economists often regard as the determinants of individual destinies in the distribution of wealth. In spite of this limitation, and probably much more important from the viewpoint of economic policy, statistical equilibrium tells us something about the functional form of the wealth distribution. As long as we can only influence the parameters of a distribution and not its functional form—which is the basic result of distributional theory in statistical equilibrium and does not depend on the aforementioned individual characteristics of agents—the task of policymakers will be to provide institutional environments that are able to change these parameters. Think, for example, of the equality within the power law tail. If we believe that a more equal distribution among the very wealthy is desirable—quite possibly making a society 'more democratic' because varied interests are better able to compete with each other through similarly powerful pressure groups—a policymakers' task would be to increase turnover activity among the wealthiest portfolios by providing adequate institutional means.

The material in Chapters 5 and 6—i.e. the calibration of the power law theory and the extension of statistical equilibrium theory to the Gamma distribution of wealth—should be understood as a road map for the things that lie ahead and need to be resolved in more detail. More generally, the "unification" of distributional theory across the different regimes and "empirical testing" of statistical equilibrium theory are the pressing items on the research agenda. But there are also other questions that arise from the current work. A particular detail concerns the wealth level that separates the power law from the Gamma distribution. How rich does one have to be in order to enjoy the multiplicative regime that is merely constrained by the growth rate of wealth? A trained economist will almost naturally ask whether this wealth level represents a threshold that fundamentally changes the economic behavior of agents in such a way that they start to re-invest all their income. Here we can only speculate that at the threshold agents do not need to worry anymore about fulfilling their material needs, as agents below the threshold do, and thus can afford to re-invest returns perpetually.

Our theory started from the stylized fact that different households put their wealth to different economic uses, and the statistical equilibrium models neatly tied up the composition of wealth portfolios to the functional form of observed wealth distributions. Reflecting more broadly on what we have learned so far about the distribution of wealth, however, the fundamental question remains why the majority of agents does not want to—or is not able to—diversify its wealth portfolios into assets that continuously earn a rate of return. Different preferences will hardly serve as a satisfactory explanation to this question.

Appendix A

Definitions of Wealth

Marketable wealth, or net worth, is composed of (1) the gross value of owner-occupied housing; (2) other real estate owned by the household; (3) cash and demand deposits; (4) time and savings deposits, certificates of deposit (CDs), and money market accounts; (5) bonds (government, corporate, foreign) and other financial securities; (6) the cash surrender value of life insurance plans; (7) the cash surrender value of pension plans; (8) corporate stock, including mutual fund holdings; (9) net equity in unincorporated businesses; and (10) equity in trust funds. Subtracting the current value of mortgage debt, consumer debt, and other debt yields a household's marketable wealth.

When items (6) and (7) are included, the measure is sometimes also referred to as *augmented* wealth, while the definition of *financial* wealth subtracts the net equity position in owner-occupied housing, i.e. the difference between the property value and outstanding mortgage debt.

Appendix B

Continuous Entropy Measure

The following heuristic motivation for the use of $H(f)$ as a continuous entropy measure can be found in [23]. Let the points x_i form an equally spaced partition of $A = [a, b]$ where $x_0 = a$ and $x_n = b$ such that $\Delta x_i = x_i - x_{i-1} = \frac{b-a}{n} \equiv h$. The discrete probability p_i can be approximated by $f(x_i)\Delta x_i$ in the sense that

$$
\begin{aligned}
-\sum_{i=1}^{n} p_i \ln p_i &\approx -\sum_i f(x_i)\Delta x_i \ln\left(f(x_i)\Delta x_i\right) \\
&= -\sum_i f(x_i)\Delta x_i \ln f(x_i) - \sum_i f(x_i)\Delta x_i \ln \Delta x_i \\
&= -\sum_i f(x_i) \ln f(x_i)\Delta x_i - \ln h \sum_i f(x_i)\Delta x_i \\
&\approx -\int_a^b f(x) \ln f(x)\, \mathrm{d}x - \ln h \int_a^b f(x)\, \mathrm{d}x \\
&= -\int_a^b f(x) \ln f(x)\, \mathrm{d}x - \ln h.
\end{aligned}
$$

The term $-\ln h$ causes some difficulty since $-\ln h \to \infty$ as $h \to 0$. However, if we consider the difference between the entropy of $f(x)$ and the entropy of another density function $g(x)$ corresponding to the probability distribution q_i for $i = 1, \ldots, n$ then the term cancels out. In this sense $H(f)$ represents a measure not of absolute but of relative uncertainty (relative to any other distribution). Of course, this is not a rigorous but merely a heuristic justification for the use of $H(f)$ as a measure of entropy. Instead of h, Jaynes [21] considers the limiting density of discrete points in h and arrives at

$$H^m(f) = -\int f(x) \ln \frac{f(x)}{m(x)} \, \mathrm{d}x,$$

where $m(x)$ is proportional to the limiting density of points in h. As Jaynes points out, under a change of variables the functions $f(x)$ and $m(x)$ transform in the same way so that $H^m(f)$ will be invariant. The probability density function under a constraint on the logarithmic mean obeys

$$f(x) = x^{-\lambda} \frac{m(x)}{Z(\lambda)}.$$

This implies that the functional form of a power law will be preserved if the measure $m(x)$ obeys a power law itself. Since the measure should be finite over its support, it must be of the generic form $m(x) = x^{-(1+\epsilon)}$ for all $\epsilon > 0$. Intuitively, such a measure would provide a proportionally spaced rather than an equally spaced partition of points on the support.

Bibliography

[1] Anderson, P. W. Some thoughts about distribution in economics. In *The Economy as an Evolving Complex System II*, B. Arthur et al., Eds. Addison-Wesley, Reading, MA, 1997.

[2] Bak, P. *How Nature Works: The Science of Self-organized Criticality.* Springer-Verlag (Copernicus), New York, 1996.

[3] Brock, W. A. Scaling in economics: A reader's guide. *Industrial and Corporate Change 8* (1999), 409–446.

[4] Champernowne, D. G. A model of income distribution. *Economic Journal 63* (1953), 318–351.

[5] Chiang, A. C. *Elements of Dynamic Optimization.* McGraw-Hill, New York, 1992.

[6] Dagum, C., and Zenga, M., Eds. *Income and Wealth Distribution, Inequality and Poverty.* Springer-Verlag, Berlin, 1990.

[7] DAVIES, J. B., AND SHORROCKS, A. F. The distribution of wealth. In *Handbook of Income Distribution*, A. B. Atkinson and F. Bourguignon, Eds. Elsevier Science B. V., Amsterdam, 2000.

[8] DE HAAN, L., AND RESNICK, S. On asymptotic normality of the Hill estimator. *Stochastic Models 14* (1998), 849–867.

[9] DREES, H., DE HAAN, L., AND RESNICK, S. How to make a Hill plot. *Annals of Statistics 28* (2000), 254–274.

[10] DRĂGULESCU, A., AND YAKOVENKO, V. M. Exponential and power law probability distributions of wealth and income in the UK and US. *Physica **A** 299* (2001), 213–221.

[11] FIELDS, G. S., AND OK, E. A. The meaning and the measurement of income mobility. *Journal of Economic Theory 71* (1996), 349–377.

[12] FOLBRE, N. *The New Field Guide to the U.S. Economy.* The New Press, New York, 1995.

[13] FOLEY, D. K. A statistical equilibrium theory of markets. *Journal of Economic Theory 62* (1994), 321–345.

[14] FOLEY, D. K. Statistical equilibrium in a simple labor market. *Metroeconomica 47* (1996), 125–147.

[15] GABAIX, X. Zipf's law for cities: An explanation. *Quarterly Journal of Economics 114* (1999), 739–767.

[16] Greenwood, D. T., and Wolff, E. N. Changes in wealth in the US, 1962-1983: Savings, capital gains, inheritance, and lifetime transfers. *Journal of Population Economics 5* (1992), 261–288.

[17] Hall, P. On some simple estimates of an exponent of regular variation. *Journal of the Royal Statistical Society **B** 44* (1982), 37–42.

[18] Hill, B. M. A simple general approach to inference about the tail of a distribution. *Annals of Statistics 3* (1975), 1163–1174.

[19] Jaynes, E. T. *Prior Probabilities.* In Rosenkrantz [38], 1989, ch. 7 (first published in 1968).

[20] Jaynes, E. T. *What Is the Question?* In Rosenkrantz [38], 1989, ch. 13 (first published in 1981).

[21] Jaynes, E. T. *Where Do We Stand on Maximum Entropy?* In Rosenkrantz [38], 1989, ch. 10 (first published in 1978).

[22] Kapur, J. N. *Maximum-Entropy Models in Science and Engineering.* John Wiley & Sons, New York, 1989.

[23] Kapur, J. N., and Kesavan, H. K. *Entropy Optimization Principles with Applications.* Academic Press, San Diego, 1992.

[24] Kennickell, A. B., et al. Recent changes in US family finances: Results from the 1998 Survey of Consumer Finances. *Federal Reserve Bulletin 86* (2000), 1–29.

[25] KIRMAN, A. P. Pareto as an economist. In *The New Palgrave Dictionary of Economics*, J. Eatwell et al., Eds. MacMillan, London, 1987.

[26] KOTLIKOFF, L. J., AND SUMMERS, L. H. The role of intergenerational transfers in aggregate capital accumulation. *Journal of Political Economy 89* (1981), 706–732.

[27] LEE, Y., ET AL. Universal features in the growth dynamics of complex organizations. *Physical Review Letters 81* (1998), 3275–3278.

[28] LEVY, M., AND SOLOMON, S. Dynamical explanation for the emergence of power law in a stock market model. *International Journal of Modern Physics **C** 7* (1996), 65–72.

[29] LEVY, M., AND SOLOMON, S. Power laws are logarithmic Boltzmann laws. *International Journal of Modern Physics **C** 7* (1996), 595–601.

[30] LEVY, M., AND SOLOMON, S. New evidence for the power law distribution of wealth. *Physica **A** 242* (1997), 90–94.

[31] MANDELBROT, B. The Pareto-Levy law and the distribution of income. *International Economic Review 1* (1960), 79–106.

[32] MANTEGNA, R. N., AND STANLEY, H. E. *An Introduction to Econophysics.* Cambridge University Press, New York, 2000.

[33] MEADE, J. E. *Efficiency, Equality and the Ownership of Property.* George Allen & Unwin, London, 1964.

[34] Oulton, N. Inheritance and the distribution of wealth. *Oxford Economic Papers 28* (1976), 86–101.

[35] Persky, J. J. Pareto's law. *Journal of Economic Perspectives 6* (1992), 181–192.

[36] Quadrini, V., and Rios-Rull, J.-V. Understanding the US distribution of wealth. *Federal Reserve Bank Minneapolis Quarterly Review 21* (1997), 22–36.

[37] Reed, W. J. The Pareto, Zipf and other power laws. *Economics Letters 74* (2001), 15–19.

[38] Rosenkrantz, R. D., Ed. *E. T. Jaynes: Papers on Probability, Statistics and Statistical Physics.* Kluwer Academic Publishers, Dordrecht, Netherlands, 1989.

[39] Ross, S. M. *A First Course in Probability.* Prentice Hall, Upper Saddle River, NJ, 1998.

[40] Shorrocks, A. F. The measurement of mobility. *Econometrica 46* (1978), 1013–1024.

[41] Shorrocks, A. F. The portfolio composition of asset holdings in the UK. *Economic Journal 92* (1982), 268–284.

[42] Steindl, J. Pareto distribution. In *The New Palgrave Dictionary of Economics*, J. Eatwell et al., Eds. MacMillan, London, 1987.

[43] STIGLITZ, J. E. Distribution of income and wealth among individuals. *Econometrica 37* (1969), 382–397.

[44] WALDROP, M. M. *Complexity: The Emerging Science at the Edge of Order and Chaos.* Simon & Schuster (Touchstone), New York, 1992.

[45] WOLFF, E. N., Ed. *International Comparisons of the Distribution of Household Wealth.* Oxford University Press, New York, 1987.

[46] WOLFF, E. N. *Top Heavy: The Increasing Inequality of Wealth in America and What Can Be Done about It.* The New Press, New York, 1996.

[47] WOLFF, E. N. Why has median wealth grown so slowly in the 1990s? Working Paper, New York University, November 2000.