Laboratory of Economics and Management
Sant'Anna School of Advanced Studies

Piazza Martiri della Libertà, 33 - 56127 PISA (Italy)
Tel. +39-050-883-341 Fax +39-050-883-344
Email: lem@sssup.it Web Page: http://www.sssup.it/~LEM/

LEM
Working Paper Series

Maximum Entropy Power Laws: an Application to the Tail of Wealth Distribution

Mishael Milaković *

New School for Social Research, New York.

2003/1 **January 2003**

ISSN (online) 2284-0400

Maximum Entropy Power Laws: An Application to the Tail of Wealth Distributions[1]

Mishael Milaković

Version: December 2002

Statistical equilibrium denotes the distribution of wealth that can be achieved in the largest number of ways while satisfying a first moment constraint on the rate of growth in wealth portfolios. Maximizing entropy subject to a logarithmic constraint yields a power law distribution whose characteristic exponent depends positively on the minimum wealth level, and inversely on the rate of growth and the average number of changes in the composition of wealth portfolios. Put differently, the distribution of wealth will be more unequal the faster the rate of growth in wealth and also the higher the number of turnovers.

[1] I would like to thank Giulio Bottazzi, Eugene Canjels, Carolina Castaldi, Giovanni Dosi, Peter Howitt, Salih Neftci, Ilfan Oh, Barkley Rosser, Willi Semmler, Sorin Solomon, and Lance Taylor for useful comments on earlier versions of this paper, which was presented at the 26th *Eastern Economic Association* conference and the 7th *Society for Computational Economics* meeting, both in 2001. I am most indebted to Duncan Foley for his many suggestions and the numerous conversations with him that have greatly enhanced my understanding of statistical equilibrium and led me to this formulation of wealth distribution theory. All errors are of course mine.

1. INTRODUCTION

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 [Anderson (1997); Drăgulescu and Yakovenko (2001); Levy and Solomon (1997); Steindl (1987); Wolff (1987)]. 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.[1] In contrast, models from probability theory [Champernowne (1953); Levy and Solomon (1996a); Mandelbrot (1960); Mantegna and Stanley (2000); Reed (2001)] provide insights into why a variable should be distributed according to a power law—but they often lack a clear relationship to economic theory.[2] The purpose of this paper is to develop an alternative probabilistic theory of wealth distribution that is supported by economic intuition.

Wealth consists of the current value of assets a household owns minus the current value of liabilities it services. The economic sources of wealth are income, inheritance, and the revaluation of assets or liabilities; 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.

The present paper takes 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 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.[3] To paraphrase the methodological view underlying the field of "complexity science"—as popularized for example in Bak (1996) or Waldrop (1992)—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.[4]

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' (1978) *maximum entropy program*, and on Foley's (1994) 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 aggregate growth constraint is the *statistical equilibrium* or *maximum entropy wealth distribution*.

An arithmetic mean constraint in the maximum entropy program leads to the (Gibbsian) exponential distribution.[5] 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 (1953) model of income distribution. More recently, Levy and Solomon (1996a) 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 (1996a) prove that the ergodic distribution of w will converge to a power law.[6]

The key to their proof lies in the logarithmic scale of wealth, so that 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.[7] 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 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 wealth level, the statistical equilibrium exponent depends inversely on the aggregate rate of growth in wealth portfolios, and it also depends inversely on the average number of turnovers that occur during the period.

2. 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 time t value of the *wealth portfolio* of household h, 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. In proper accounting however, the sources of wealth—like savings—have to equal the uses of wealth in value terms. 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 two well-defined macroscopic averages. The first is the average number of turnovers in the economy.

A *turnover* describes a change in the household's position between period t and t'. Let $T_h(t, t')$ designate the number of elements where $a_h^k(t) \neq a_h^k(t')$ for all $k \in \mathbf{K}$, that is to say $T_h(t, t')$ gives the number of changes in the composition of household h's portfolio between period t and t'. Hence, we define the *average number of turnovers* in the economy between period t and t' as

$$T(t,t') \equiv n^{-1} \sum_h T_h(t,t').$$

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 a *return factor* $r(t) \equiv w(t)/w(t_0)$. If the economy starts with an egalitarian distribution of wealth at t_0, where $w_h(t_0) = w_0$ for all $h \in \{1, \ldots, n\}$, then

$$r_h(t) \propto w_h(t).$$

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.

The results of our model depend on the following assumption, capturing the idea of proportionality in returns to wealth.

Assumption 1 We express proportionality in return factors with a logarithmic scale, $\log w(t)$.

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.[8] Let n_i denote the number of households with wealth w_i. The logarithmic mean $\overline{\log\left[w\left(t\right)/w_0\right]}$ over the $t-t_0$ periods of observation is the sum of returns per average turnover $T(t_0,t)$ weighted by the frequency $p_i\left(w\right) \equiv n_i/n$ of households enjoying wealth level w_i

$$\frac{\overline{\log\left[w(t)/w_0\right]}}{t-t_0} = \sum_i p_i(w) \frac{\log\left[w_i(t)/w_0\right]}{T(t_0,t)}. \tag{1}$$

Let us disentangle what might look at first like a somewhat contorted formulation of a logarithmic mean constraint by pointing out the necessary dimensionality. A logarithmic mean by itself has no time dimension. We could use 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. Thus the constraint reads the sum of weighted log-returns per turnover has to add up to the growth rate per unit time of observation. As we will argue in Section 5, the use of two different time scales allows us to resolve the conceptual issue of a 'zero period' and the absolute lapse of time.

It is important to notice that we are no longer summing over households but over the number of *theoretically possible wealth levels* $w_i(t)$ 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

$$\sum_i p_i(w) = 1. \tag{2}$$

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 (1) 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 (1) and the natural constraint (2). Which probability distribution should we choose in the absence of any further information?

3. MAXIMUM ENTROPY WEALTH DISTRIBUTION

A feasible wealth distribution obeys (1) and (2). 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, $M[\{n_c\}] \equiv n!/n_1! \cdots n_c! \cdots n_C!$. 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.

For notational convenience, let $L \equiv t - t_0$ denote the length of observation and let $T \equiv T(t_0, t)$ denote the number of turnovers during the period. The *maximum entropy program* maximizes entropy $H[\{p_i(w)\}]$ subject to the natural constraint (2) and a finite number of moment constraints. In our case, we are dealing with a single logarithmic constraint (4) that measures the average return $r_i(t)$ over the length of observation as the average return per turnover

$$\max_{\{p_i\}} H[\{p_i(w)\}] = -\sum_i p_i(w) \log p_i(w) \tag{3}$$

subject to

$$\begin{aligned} \sum_i p_i(w) \frac{\log r_i(t)}{T} &= \frac{\overline{\log r(t)}}{L}, \\ \sum_i p_i(w) &= 1. \end{aligned} \tag{4}$$

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 [Jaynes (1978); Kapur and Kesavan (1992)].[9] If an entropy-maximizing wealth distribution exists, it is unique because the objective function (3) is strictly concave and the constraints define a convex set [Jaynes (1978); Foley (1994)]. The maximum entropy program yields the proof of a similar theorem by Levy and Solomon (1996a).

THEOREM 1 (Power laws are logarithmic Boltzmann laws). *For all return factors* $r_i(t) > 0$ *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*

$$p_i^*(w) = \frac{r_i(t)^{-\lambda^*}}{Z(\lambda^*)} \tag{5}$$

where

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

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

Proof. If for any constant $c > 0$, $r_i(t) \neq c$ for all 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

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

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

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

From (2) we then obtain $\exp(\lambda_0^*) = Z(\lambda^*)$, resulting in the power law $p_i^*(w) = r_i(t)^{-\lambda^*}/Z(\lambda^*)$. ∎

Theorem 1 says that the most disorderly mixing of return factors leads to a power law distribution of wealth. To paraphrase Foley's (1994) 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 wealth levels* w_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 of feasible wealth levels. ∎

4. CHARACTERISTIC EXPONENT OF THE WEALTH DISTRIBUTION IN STATISTICAL EQUILIBRIUM

Using the statistical equilibrium distribution in the growth constraint (4), we obtain a parametric solution for λ

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

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 Section 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 A, where we also derive the general conditions for the invariance of the entropy measure. Moreover, we argue why introducing the new measure does not alter our results qualitatively.[10] 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 2 The power law distribution has finite mean, i.e. $\lambda > 2$.

As before, we define return factors as $r(t) \equiv w(t)/w(t_0)$ and denote them, for notational simplicity, without the time index simply as r. The maximum entropy program then takes the form

$$\max_{f(r)} H[f(r)] \equiv -\int_W f(r) \log f(r)\, \mathrm{d}r \tag{6}$$

subject to

$$\int_W f(r) \log r\, \mathrm{d}r = \frac{T}{L}\overline{\log r} \tag{7}$$

$$\int_W f(r)\, \mathrm{d}r = 1. \tag{8}$$

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

$$f^*(r) = \frac{\lambda^* - 1}{r_{\min}^{-\lambda^*+1}} r^{-\lambda^*}. \tag{9}$$

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

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

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 (8),

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

Substitution completes the proof. ∎

Lemma 1 enables us derive the central proposition of the paper, 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*

$$\lambda^* = \left(\frac{T}{L}\overline{\log r} - \log r_{\min}\right)^{-1} + 1. \tag{10}$$

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 (7) yields (10). ∎

Proposition 1 identifies the three determinants of the characteristic exponent: the average rate of growth 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 (1996a) on the minimum wealth level as the determinant of the characteristic exponent. Drăgulescu and Yakovenko (2001) show that the empirically observed wealth distribution changes its functional form at a particular wealth level.[11] 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.

We are now in a position to interpret the economic implications of

Proposition 1, and discuss how to operationalize and calibrate the model from data.

5. INTERPRETATION, CALIBRATION, AND SOME CASUAL EMPIRICISM

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 (9). In order to interpret (9) as the actual distribution we have to make one more assumption.

Assumption 3 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 3.

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 [Anderson (1997); Kirman (1987); Steindl (1987)]. 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 (logarithmic) 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 the economy grows, and the higher the average number of turnovers in the economy, the more unequal the maximum entropy distribution of wealth.

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

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

For $\overline{\log r}\,, T > 0$ the partial derivatives are negative. 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 (1964) inherent conflict between income equality and productive efficiency in an economy. With respect to turnover activity, the distribution of fortunes will be more unequal, *ceteris paribus*, the less institutional frictions exist in the economy's financial markets.

Financial liberalization over the last three decades has considerably increased investment opportunities. But the largely increased flow of capital can also be considered as an increase in turnover activity, and casual observation would have us believe that inequality has increased since the collapse of Bretton-Woods. But keep in mind that our notion of inequality applies to the power law tail, and not to the relationship between the very wealthy and the rest. To judge whether inequality within the power law tail has indeed increased with financial liberalization, we need (rarely available) information about the characteristic exponent. Most wealth data are reported in Lorenz form but if we know the wealth share S of a top percentile P, we can infer the characteristic exponent 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*

$$\lambda = 1 + \left(1 - \frac{\log S}{\log P}\right)^{-1}. \tag{11}$$

Proof. Without loss of generality, we set minimum wealth to unity and consider the probability density function $n(x)$ of wealth x, $n(x) = cx^{-\lambda}$, with c as the appropriate normalizing constant; then P and S are defined by the ratios

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

again provided Assumption 2 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 λ 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 is usually cited in the literature. It is readily verified that the exponent of the distribution function will be $(1 - \log S / \log P)^{-1}$.

We could calculate error bounds for λ from (11) if we knew the variance σ^2 in the measurement error of S by simply calculating λ 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

$$\left|\frac{\partial \lambda}{\partial S}\right| \sigma = \left(1 - \frac{\log S}{\log P}\right)^{-2} \frac{\sigma}{|\log P| S} \quad \text{for } P, S \in (0,1). \tag{12}$$

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 λ since $|\partial\lambda/\partial S| < 1$ means a less than proportionate increase in the error bounds of λ compared to σ; the opposite is true for $|\partial\lambda/\partial S| > 1$. Table 2 reports the numerically computed solutions for the 'critical values' S^c where $|\partial\lambda/\partial S| = 1$. We chose P based on the data in Table 1.

For a given P, equation (11) shows λ as a strictly convex function of $S > P$. Therefore, measurement errors will be magnified in λ if $0 < S < S^c$, while the reverse is true if $S^c < S < 1$.

Table 1 presents Lorenz data for different countries at different points in time, taken from Wolff (1987, 1996), together with the characteristic exponents of the distribution function (cdf) that we calculated with 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, see for example Wolff (1987), and in spite of the rather coarse nature of wealth percentiles.

Wealth inequality was significantly higher at the beginning of the twentieth century compared to the decades between World War II and the collapse of Bretton-Woods. This would suggest (a) that our assumption of a large but stable ratio of turnovers per observational period is essential to a meaningful interpretation of the theoretical model, and (b) that inequality within the power law indeed increases with financial liberalization, provided we agree that turnover activity was higher during the 1920's and after Bretton-Woods than during the 'Golden Age.'

An important question for the calibration of the model is whether wealth levels (or the 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 (10), 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.

So far we argued in a rough-and-ready manner that the qualitative features of the model are supported empirically or, at any rate, do not suffer from obvious empirical contradictions. To judge the quantitative abilities of the statistical equilibrium model, we would have to check whether the equality in (10) holds empirically. 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.

Closer inspection reveals that such an endeavor is everything but trivial. Our conceptual device of an initial period with egalitarian distribution of wealth has been very helpful in arguing why return factors can be considered as wealth levels, and the assumption of a stable turnover rate allowed us to interpret the phenomenological characteristic exponent as our theoretical λ. Like the dimensionless exponent, the growth rate is a dimensionless ratio (per unit of observation) and so is the minimum return factor. Empirically, however, we do not observe the minimum return factor but the minimum wealth level in currency terms. The simplest, least satisfactory, and most *ad hoc* way around this would be to assume a particular minimum return factor, say, unity, in which case $r_{\min}$ vanishes altogether from equation (10). A much sounder solution is available if we know either the arithmetic mean return factor $\langle r \rangle$ or any (100k)th quantile r_k of the power law. Given the characteristic exponent, we can use the definitions of $\langle r \rangle$ or r_k to determine that $r_{\min}$ equals $\langle r \rangle(1 - 1/(\lambda - 1))$ or $r_k(1 - k)^{1/\lambda - 1}$. Unfortunately, we have no reason to believe that $\langle r \rangle$ or r_k are any easier to observe than $r_{\min}$. Yet another possibility—if we can trace the identity of a subset of wealthy agents in the power law tail—would be to proxy $r_{\min}$ with the smallest observed return factor among the subset.[12]

6. CONCLUSION

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 (1996a), 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 and the more turnovers occur, the more unequal the power law distribution of wealth.

The present model opens several venues for further research. On a theoretical level, we would certainly like to extend our model into a 'unified' theory that applies to the entire wealth distribution, and not just to the power law tail. The most intriguing question remains what determines the location of the minimum wealth level that separates the power law regime from the left part of the distribution, for example the exponential regime that Drăgulescu and Yakovenko (2001) observe in the UK. It would also be desirable to embed wealth and income taxation in the entropy model.

Empirically, the next step should be to calculate implied turnover rates from sources that provide named wealth data for the power law tail of the wealth distribution, or at least for a subset of it.

Dept. of Economics, New School University, 65 Fifth Ave., New York, NY 10003, U.S.A.; milakm01@newschool.edu.

REFERENCES

ANDERSON, P. W. (1997): "Some Thoughts about Distribution in Economics," in *The Economy as an Evolving Complex System II*, ed. by B. Arthur, S. N. Durlauf, and D. A. Lane. Reading, MA: Addison-Wesley.

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

BROCK, W. A. (1999): "Scaling in Economics: A Reader's Guide," *Industrial and Corporate Change*, 8, 409–446.

CHAMPERNOWNE, D. G. (1953): "A Model of Income Distribution," *Economic Journal*, 63, 318–351.

DRĂGULESCU, A., AND V. M. YAKOVENKO (2001): "Exponential and Power Law Probability Distributions of Wealth and Income in the UK and US," *Physica* **A**, 299, 213–221.

FOLBRE, N. (1995): *The New Field Guide to the U.S. Economy.* New York: The New Press.

FOLEY, D. K. (1994): "A Statistical Equilibrium Theory of Markets," *Journal of Economic Theory*, 62, 321–345.

GABAIX, X. (1999): "Zipf's Law for Cities: An Explanation," *Quarterly Journal of Economics*, 114, 739–767.

JAYNES, E. T. (1978): "Where Do We Stand on Maximum Entropy? reprinted in *E. T. Jaynes: Papers on Probability, Statistics and Statistical Physics*, ed. by R. D. Rosenkrantz. Dordrecht: Reidel.

KAPUR, J. N., AND H. K. KESAVAN (1992): *Entropy Optimization Principles with Applications.* San Diego, CA: Academic Press.

KENNICKELL, A. B., M. STARR-MCCLUER, AND B. J. SURETTE (2000): "Recent Changes in US Family Finances: Results from the 1998 Survey of Consumer Finances," *Federal Reserve Bulletin*, 86, 1–29.

KIRMAN, A. P. (1987): "Pareto as an Economist," in *The New Palgrave Dictionary of Economics*, ed. by J. Eatwell, M. Milgate, and P. Newman. London: MacMillan.

KOTLIKOFF, L. J., AND L. H. SUMMERS (1981): "The Role of Intergenerational Transfers in Aggregate Capital Accumulation," *Journal of Political Economy*, 89, 706–732.

LEE, Y., L. A. NUNES AMARAL, D. CANNING, M. MEYER, H. E. STANLEY (1998): "Universal Features in the Growth Dynamics of Complex Organizations," *Physical Review Letters*, 81, 3275–3278.

LEVY, M., AND S. SOLOMON (1996a): "Power Laws are Logarithmic Boltzmann Laws," *International Journal of Modern Physics* **C**, 7, 595–601.

——— (1996b): "Dynamical Explanation for the Emergence of Power Law in a Stock Market Model," *International Journal of Modern Physics* **C**, 7, 65–72.

——— (1997): "New Evidence for the Power Law Distribution of Wealth," *Physica* **A**, 242, 90–94.

MANDELBROT, B. (1960): "The Pareto-Levy Law and the Distribution of Income," *International Economic Review*, 1, 79–106.

MANTEGNA, R. N., AND H. E. STANLEY (2000): *An Introduction to Econophysics.* New York: Cambridge University Press.

MEADE, J. E. (1964): *Efficiency, Equality and the Ownership of Property.* London: George Allen & Unwin.

OULTON, N. (1976): "Inheritance and the Distribution of Wealth," *Oxford Economic Papers*, 28, 86–101.

PERSKY, J. (1992): "Pareto's Law," *Journal of Economic Perspectives*, 6, 181–192.

QUADRINI, V., AND J. RIOS-RULL (1997): "Understanding the US distribution of Wealth," *Federal Reserve Bank Minneapolis Quarterly Review*, 21, 22–36.

REED, W. J. (2001): "The Pareto, Zipf and Other Power Laws," *Economics Letters*, 74, 15–19.

STEINDL, J. (1987): "Pareto Distribution," in *The New Palgrave Dictionary of Economics*, ed. by J. Eatwell, M. Milgate, and P. Newman. London: MacMillan.

STIGLITZ, J. E. (1969): "Distribution of Income and Wealth Among Individuals," *Econometrica*, 37, 382–397.

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

WOLFF, E. N. (Ed.) (1987): *International Comparisons of the Distribution of Household Wealth.* New York: Oxford University Press.

——— (1996): *Top heavy: The Increasing Inequality of Wealth in America and What Can Be Done about It*, New York: The New Press.

——— (2000), "Why Has Median Wealth Grown So Slowly in the 1990s?" working paper, NYU, November 2000.

NOTES

[1]Quadrini and Rios-Rull (1997) provide a survey of the literature.

[2]See Brock (1999) and Gabaix (1999) for critical assessments.

[3]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.

[4]Wolff (1996, 2000) 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.

[5]Kapur and Kesavan (1992) present numerous applications of the maximum entropy program under different constraints taken from the natural and social sciences. Foley (1994) 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).

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

[7]If the characteristic exponent is less or equal to unity, the power law distribution has infinite mean. Correspondingly, the *density* function loses its first moment if the characteristic exponent is smaller or equal to two.

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

[9]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.

[10]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, euro instead of dollars does not affect the scale of return factors.

[11]According to Drăgulescu and Yakovenko (2001), the wealth distribution changes from an exponential shape to a power law for the top five percent of households in the UK.

[12]In a forthcoming paper, co-authored with Carolina Castaldi, we do exactly that by calibrating the statistical equilibrium model from seven consecutive years of the *Forbes 400* list, a named subset of the four hundred wealthiest US individuals.

TABLE 1

Characteristic exponent λ calculated from Lorenz curve pairs (P, S) with Lemma 2. Data are taken from Wolff (1996) and the papers collected in Wolff (1987); different data sources are indicated by horizontal lines.

Country	Year	Top households P	Wealth share S	λ (cdf)
US	1983	0.01	0.340	1.306
US	1989	0.01	0.390	1.257
Sweden	1920	0.0001	0.090	1.354
Sweden	1920	0.001	0.240	1.260
Sweden	1920	0.01	0.500	1.177
Sweden	1975	0.001	0.060	1.687
Sweden	1975	0.002	0.080	1.685
Sweden	1975	0.005	0.125	1.646
Sweden	1975	0.01	0.170	1.625
Sweden	1975	0.02	0.240	1.574
Sweden	1983	0.001	0.080	1.576
Sweden	1983	0.002	0.100	1.589
Sweden	1983	0.005	0.145	1.573
Sweden	1983	0.01	0.195	1.55
Sweden	1983	0.02	0.260	1.525
UK	1923	0.01	0.610	1.120
UK	1923	0.05	0.820	1.071
UK	1929	0.01	0.560	1.144
UK	1929	0.05	0.790	1.085
UK	1975	0.01	0.240	1.449
UK	1975	0.05	0.440	1.378
UK	1980	0.01	0.230	1.469
UK	1980	0.05	0.430	1.392
France	1977	0.01	0.190	1.564
France	1977	0.05	0.450	1.363
France	1986	0.01	0.260	1.413
France	1986	0.05	0.430	1.392
Germany	1973	0.01	0.280	1.382
Belgium	1969	0.01	0.280	1.382
Denmark	1973	0.01	0.250	1.431
Sweden	1975	0.01	0.160	1.661
Canada	1970	0.01	0.196	1.548
US	1972	0.01	0.250	1.431

TABLE 2
Critical values for the propagation of errors.

P	.0001	.001	.002	.005	.01	.03	.05
S^c	.167	.232	.261	.310	.359	.466	.534
λ (cdf)	1.241	1.267	1.275	1.283	1.286	1.278	1.265

APPENDIX A

CONTINUOUS MEASURE OF ENTROPY

The following heuristic motivation for the use of $H(f)$ as a continuous entropy measure can be found in Kapur and Kesavan (1992). 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 (1978) 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.