Time Evolution of The Mutual Fund Size Distribution

Yonathan Schwarzkopf[1,2*] and J. Doyne Farmer[2,3*]

[1] *Department of Physics,*
California Institute of Technology,
1200 E California Blvd, mc 103-33
Pasadena, CA 91125
[2] *Santa Fe Institute, 1399 Hyde Park Road,*
Santa Fe, NM 87501
[3] *Luiss Guido Carli, Viale Pola 12 00198,*
ROMA Italy
(Dated: January 13, 2008)

Abstract

In this article we investigate the size, in terms of total assets managed, of equity mutual funds. We show that the large tail of the size distribution of equity funds is best described by a log-normal as opposed to a power law. We begin our analysis by comparing the power law tail hypothesis to the log-normal tail hypothesis and show that not only is the log-normal hypothesis more likely but that the power law tail may be rejected. We argue, using a stochastic growth model, that the mutual fund ecology is young and as such it is in a transient state and given enough time it will converge to a steady state in which the large tail of the distribution follows a power law. Previous analysis of the mutual fund ecology dangerously treated the process as stationary without identifying the time scales in which the process converges. Such a treatment predicts a stationary size distribution with a power law tail that was argued to follow Zipf's law. We construct a stochastic growth model for the ecology of mutual funds in which the evolution is governed by three processes; growth, modeled as a Gibrat process, creation and annihilation of funds. We provide an analytical solution to the model which allows us to identify the time scales of the evolution process and treat the process as time dependent . Our model can be made more realistic by modifying the growth process with the empirical observations that the variance and drift of the growth process are size dependent . We show, using simulations, that the modified model does an excellent job of describing the top tail of the size distribution. By describing the distribution using a random process we show, as was argued by Herbert Simon, that investor choice is not of fundamental importance for the description of all of the observable attributes of the economy.

_____

*Electronic address: `yoni@caltech.edu,jdf@santafe.edu`

I. INTRODUCTION

The investigation of distributions has been an area of active research in economic and social systems for many years. One approach for explaining these distributions is to examine the processes that govern their behavior and treat them as stochastic processes [19, 28, 32, 33]. By modeling these processes one can identify key properties responsible for creating the observed distribution and perhaps even find similarities across several seemingly disparate systems [33].

Some of the observed distributions in various types of systems seem to exhibit heavy tails. Among these distributions are the distribution of wealth [35], the distribution of firm sizes [2, 39] (aggregated across industries) and the distribution of number of inhabitants in cities [18] which are shown to follow Zipf's law. Lately there have been observations that several phenomena in financial markets obey heavy tailed distributions. Out of these phenomena two have been the center of recent research: the distribution of traded volume and the distribution of returns [20, 21]. These distributions were observed to have a power law tail with different tail exponents. The exponent of these two tails was extensively studied [13, 40] and there is an ongoing attempt to try to relate the two exponents through a simple mechanism [15, 17, 25, 35]. It is believed that the relation between these two exponents depends on the distribution of investors wealth which is believed too to obey a power law tailed distribution.

The upper tail of investor wealth distribution can be studied by investigating the characteristics of large players. Large players such as institutional investors were shown to play a large role in market activity [10] and as such they contribute to some of the behaviors we observe in financial markets. One can hypothesize that these Institutional investors are responsible for the tails of the distributions due to their size and stature in the financial world. As such, we are interested in studying the statistical properties of institutional investors. We chose to work with mutual funds due to the high quality data available. In this work we focus on the size distribution of mutual funds and show that the distribution of equity fund sizes is better described as a log-normal. The log-normality of the size distribution is in contrast with arguments that the distribution obeys Zipf's law [15, 16]. Since mutual funds are firms our work is a study of the size distribution of firms belonging to a single industry.

The distribution of firm sizes has been actively researched over the years yielding interesting observations and as a consequence models were proposed to explain them. It has been observed early on that the size distribution of firm sizes is highly skewed [32, 41]. This observation holds true for the size distribution of firms belonging to a single industry and for for the distribution of firm sizes across industries. The firm size distribution was found to be right skewed [1, 2, 4, 11, 32, 36, 37, 41] in the sense that the modal size is smaller than the median size and both are smaller than the mean firm size. We find that the size distribution of mutual funds exhibits similar right skewness. This is not surprising since mutual funds are firms. The observed skewness of the size distribution can be explained by modeling the growth process of the firms as a stochastic process and many such processes have been proposed [15, 16, 19, 28, 32, 33, 38, 41]. Even though these processes yield skewed distributions the upper tail of these distributions is log-normal in some and pareto in others.

The functional form of the upper tail of the firm size distribution is debated. One must discuss separately the size distribution of firms in a single industry from the distribution aggregated across industries. It is accepted that by aggregating across industries the resulting size distribution has a power law tail [2, 4, 11] but it is not clear whether it is the result of the aggregation process [4, 11]. For firms belonging to a single industry the upper tail was found to be a log-normal [1, 4, 11, 32, 36, 37, 41] while others found it to be a power law [2]. Previous work on the size distribution of mutual funds argued that the distribution follows a power law upper tail [15–17] while we, on the other hand, show in this work that the distribution is better described by a log-normal upper tail.

In order to explain the observed log-normal upper tail of the mutual fund size distribution we investigate the mechanism of growth [42]. Previous work on the mechanisms of growth of mutual funds include [3, 16]. The growth is decomposed into two mechanisms; return and money flux. Clearly, these two mechanisms, return and net money flux, are correlated since a fund yielding high returns will attract investors while a poor performance by a fund will deter investors. For work done on correlations between the flux of money into or out of a mutual fund see [7, 8, 22, 31, 34] and references there in.

Another interesting question is whether there is any size dependence in these changes. It was shown that the growth rate of firms is larger for small firms and is size independent for large firms [4, 5, 12, 23, 24, 26, 29]. Such a constant return to scale industry (for large firm sizes) is consistent with a Gibrat model for the growth process. We investigate the size dependence of the growth process on the mutual fund size and fins similar results for the size dependence. The return seems to be independent of size. Meaning, that on average the performance of a fund is independent of the fund size as was also investigated by[17]. We find this to be in agreement with an efficient market in which we can not achieve greater return by simply going to larger funds. This observation of efficiency in regard to fund size is complimentary to the believed efficiency with respect to performance [27]. In contrast to the return, the money flux decays with the fund size. Thus for small sizes the dominating mechanism is money flux whereas returns are the dominating growth mechanism for large funds sizes and the growth is then independent of the mutual fund size.

We show in this work that the variance of the growth process for mutual funds is size dependent. The variance in

growth rate decays with size with a rate that depends on the size. This is in accord with observations the the variance of the growth rate of firms in individual industries is size dependent [1, 4–6, 11, 14, 36, 37] and references there in. Not only do we observe a size dependence in the variance of the growth rate but we observe the same functional form. The variance in firm growth rates was shown to obey a power law decay and not surprisingly we show that so does the variance for the variance in the mutual fund growth rates.

With the empirical observation as guide lines, we offer a random growth model that captures the essence of the mechanism responsible for the observed distribution of fund sizes. We solve the model analytically and we show that the distribution evolves from a log-normal initial state towards a power law stationary state. Using a Green's function approach for the solution of the time dependent evolution of the fund size PDF we show that dynamics of the growth process yield a log-normal tail. This log-normal tail is an innate quality of the dynamics and does not depend on the way funds are created. The independence of the log-normal tail and the creation process is valid as long as funds are not created with a heavier tailed distribution then a log-normal. The distribution evolves towards a power law tailed distribution due to the annihilation process of funds. The role of a random annihilation process in creating a steady state power law tailed distribution was previously studied in [30].

The solution for the time dependent evolution of the distribution enables us to identify the important time scales in which the distribution converges toward a power law. Using the empirical rates we calculate the time scales and show that they are indeed large compared to the age of the industry. Thus, our analysis predicts that the distribution will be better described by a log-normal then a power-law, which is verified using the empirical data.

The paper is organized as follows. Section II examines the size distribution of mutual fund using the empirical data. The underlying processes responsible for the size distribution, as they appear in the data, are discussed in Section III. Using the empirical observation for the dynamical processes Section IV lays out our model and the resulting equations that govern the time evolution of the size distribution. The model is solved and discussed in Section V. Section VI presents simulation results of the proposed model and compares them to the empirical data. Section VII presents modification to the model as suggested by the data. The modified model is simulated and compared to the data.

II. THE OBSERVED DISTRIBUTION OF MUTUAL FUND SIZES

Recently the distribution of fund sizes was reported to have a power law tail which follows Zipf's law [15–17]. Zipf's law is defined such that the Cumulative Distribution Function (CDF) for the mutual fund size s obeys a power law

$$P(s > X) \sim X^{-\zeta_s},$$

with an exponent of $\zeta_s \approx 1$. where s denotes the mutual fund size.

In the following section we will investigate the size distribution using the CRSP database. The analysis will be carried out for equity funds, which are defined as funds with a portfolio consisting of at least 80% stocks. We test the hypothesis of a power law tail against the hypothesis that the distribution is actually a lognormal . In Figure 1 the cumulative distribution of sizes $P(s > X)$ is plotted. The cumulative distributions (CDF) for the years 1993,1998 and 2005 were calculated using the empirical data for equity funds. [43]. In [17] the CDF for fund sizes in the year 1999 is plotted on a double logarithmic plot. In Figure 2 we reconstruct the graph by plotting the CDF for funds with sizes $s > 10^6$ on a double logarithmic scale. The CDF is compared to an algebraic relation of exponent $\zeta_s = -1$ which is represented in the plot by a line of slope -1. It is not at all obvious that this is a power law. Thus, in the following subsection we will test the hypothesis of a power law tail in several ways. At first we will check the validity of a power law fit to the data using statistical tests.

A. Is the distribution of fund sizes a power law?

In this section we investigate the hypothesis of a power law upper tail for the fund size distribution. By quantitatively testing the validity of such a hypothesis, we argue that such a power law hypothesis is weak at best for some years and can be unquestionably rejected for the other years.

Following [9] and references there in, we test the validity of the power law hypothesis. The power law hypothesis is such that above a certain fund size denoted by s_{min} the probability density function $p(s)$ obeys a power law for sizes larger than s_{min}

$$p(s) = \frac{\zeta_s}{s_{min}} \left(\frac{s}{s_{min}} \right)^{-(\zeta_s+1)}, \tag{1}$$
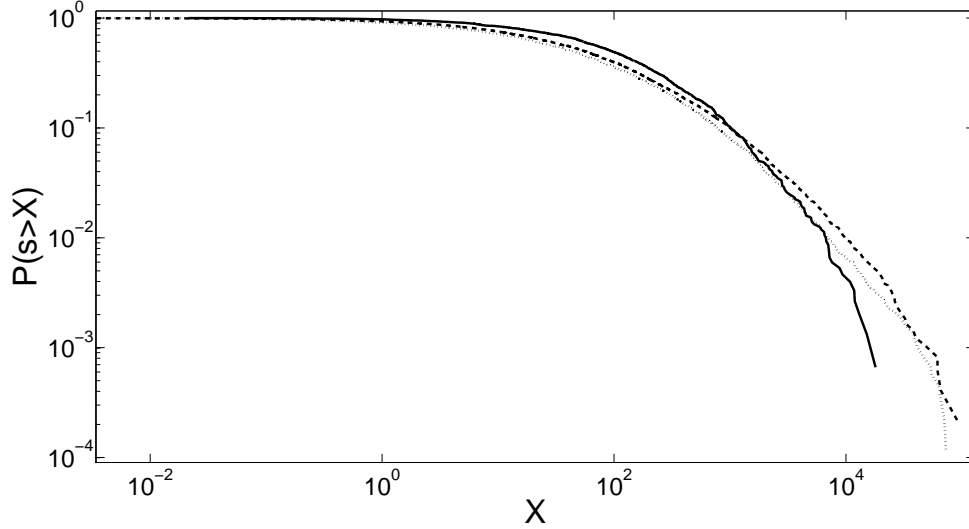
3

FIG. 1: The CDF for the mutual fund size s (in millions) is plotted on a double logarithmic plot. The cumulative distribution for funds existing at the end of the years 1993, 1998 and 2005 are given by the full, dashed and dotted lines respectively.
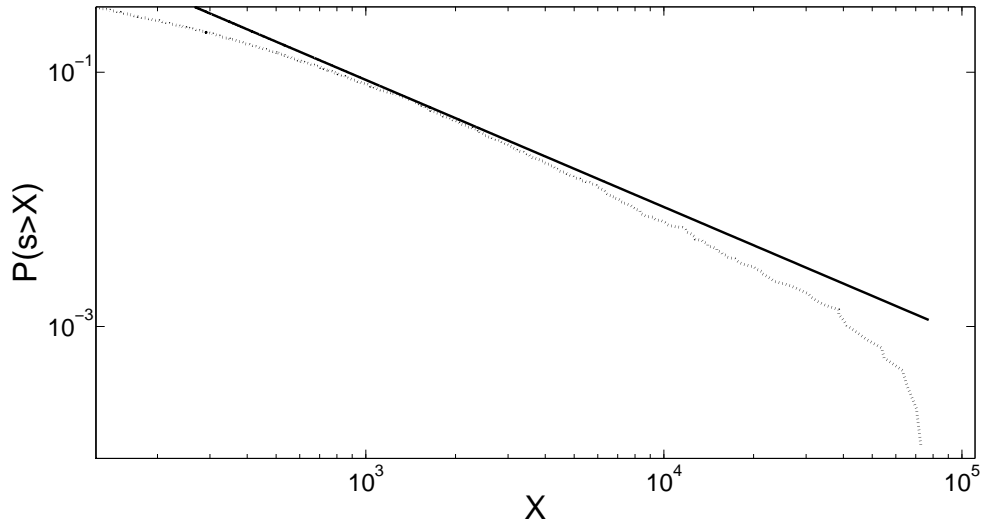

FIG. 2: The CDF for the mutual fund size s (in millions) existing in the end of 2005 is plotted in a double logarithmic scale (dotted line) is compared to an algebraic relation with exponent -1 (full line)

where the distribution is normalized in the interval $[s_{min}, \infty)$. Thus, such a power law fitting has two parameters s_{min} and the exponent ζ_s. The MLE for the scale parameter ζ_s depends on s_{min}. This crossover size s_{min} is not chosen arbitrarily (by visualization) but is chosen such that it minimizes the Kolmogorov-Smirnov (KS) statistic D [9]. The KS statistic is the distance between the CDFs of the empirical data $P_e(s)$ and the fitted model $P_f(s)$

$$D = \max_{s \geq s_{min}} |P_e(s) - P_f(s)|.$$

By following the above procedure we fit the size distribution of funds existing at the end of the years 1991 to 2005. The estimated parameters are summarized in Table I. The value for ζ_s and s_{min} is given for each of the each of the years. The mean of the yearly values is calculated $\bar{\zeta}_s = 1.09 \pm 0.04$ and can be regarded as agreeing with Zipf's law for which the exponent is ≈ -1. The standard error is calculated by dividing the standard error by the square root
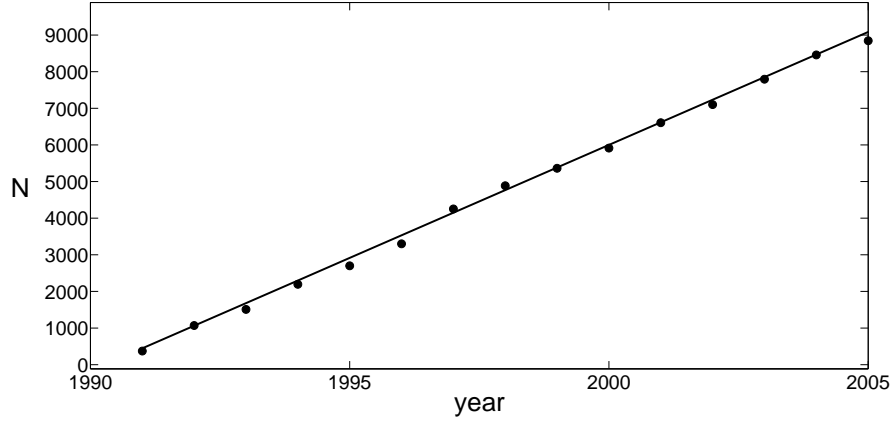
4

FIG. 3: The number of equity funds as a function of time for the yeas 1991 to 2005($\bullet$). The data is compared to a linear dependence.
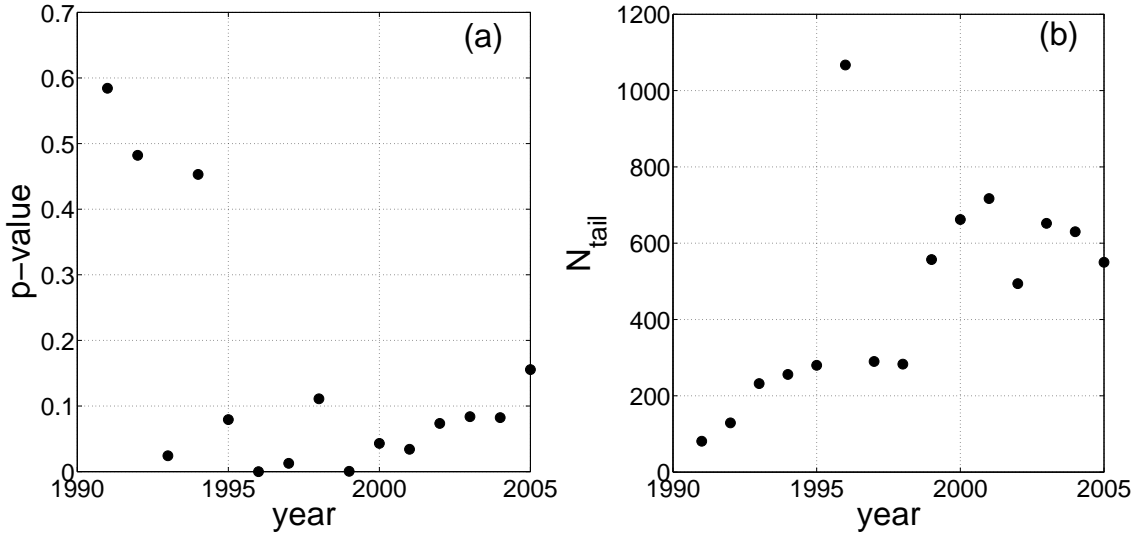

FIG. 4: (a) The p-value for a power law tail hypothesis calculated for the years 1991 to 2005. (b) The number of equity funds $N_{tail}(t)$ in the upper tail s.t. $s \geq s_{min}$ as a function of time for the years 1991 to 2005($\bullet$).

of the number of observations (number of years).

Our ability to fit the data with a power law does not mean that the power law fit is a good one. In order to check the validity of the power law hypothesis we will try and address the plausibility that the sample we observe is actually drawn from the hypothesized power law model. We do so by calculating the p-value for this model. The p-value is the probability that a data set of the same size randomly drawn from the hypothesized distribution will have a goodness of fit not better than the empirical data. As the goodness of fit we use the KS statistic D described above. The p-value is calculated numerically using a monte-carlo method [9]. The method is such that we generate a large number of synthetic data sets. Each synthetic data set is drawn from the empirical distribution for $s \leq s_{min}$ and for $s > s_{min}$ we draw from a power law distribution with an exponent that is the best fit to the empirical data given s_{min}. For each data-set we calculate the KS statistic to it's best fit. The p-value is the fraction of the data sets for which the KS statistic to it's own best fit is larger then the KS statistic for the empirical data and it's best fit. The resulting p-value were calculated numerically using 10000 randomly chosen data sets.

In Figure 4(a) the p-value for each of the years 1991 to 2005 is given. The results are summarized in Table I where the p-value for the power law fit is calculated for the data at the end of each year as well as the mean across the years. The standard error for the mean was calculated by dividing the standard deviation by the square root of the number

variable	1991	1992	1993	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	mean	std	std-err
N	372	1069	1509	2194	2699	3300	4253	4885	5363	5914	6607	7102	7794	8457	8845	-	-	-
ζ_s	1.33	1.36	1.19	1.15	1.11	0.78	1.08	1.10	0.95	0.97	1.01	1.07	1.07	1.10	1.14	1.09	0.14	0.04
s_{min}	955	800	695	708	877	182	1494	1945	1147	903	728	836	868	1085	1383	974	408	105
N_{tail}	81	129	232	256	280	1067	290	283	557	662	717	494	652	630	550	-	-	-
p-value	0.58	0.48	0.02	0.45	0.07	0	0.01	0.11	$5\,10^{-4}$	0.04	0.03	0.07	0.08	0.08	0.15	0.15	0.19	0.05
$\mu(10^{-3})$	26.1	41.8	81.2	46.2	72.7	67.9	57.7	38.8	39.5	19.6	-30	-10	50.2	17.8	30.5	38.5	26.1	6.7
σ	0.078	0.200	0.260	0.293	0.229	0.282	.289	0.302	0.283	0.278	0.255	0.242	0.236	0.240	0.246	0.247	0.055	0.014
$\mathcal{R}$	-0.50	-1.35	-1.49	-1.71	-3.29	-18.42	-2.25	-1.29	-6.57	-4.96	-2.63	-2.95	-2.00	-1.05	-0.99	-3.43	4.45	1.15

TABLE I: The values for each of the parameter (rows) are given for each year (columns). The mean and standard deviation are evaluated for the yearly values. The standard error is calculated by dividing the standard deviation by the square root of the number of observations, which corresponds to the number of years.
N - the number of equity funds existing at the end of each year.
ζ_s - the power law tail exponent (1).
s_{min} - the lower tail cutoff above which we fit a power law (1).
N_{tail} - the number of equity funds belonging to the upper tail s.t. $s \geq s_{min}$.
μ - the drift term for the geometric random walk (9).
σ - the standard deviation of the mean zero Wiener process (9).
$\mathcal{R}$ - the log likelihood ratio of a power law fit relative to a log-normal fit (3) .

of observations. One can notice that the p-value decreases with time s.t. the power law hypothesis can not be rejected at the beginning but as time progresses it can be rejected. The number of equity funds increases approximately linear with time as can be seen in Figure 3 and so do the number of equity funds in the upper tail N_{tail} given in Figure 4(b). The number of funds in the upper tail is defined as the number of funds in each of the years with a size $s \geq s_{min}$. Figure 4 suggests that as the number of funds in the upper tail increases one can reject the power law hypothesis due to better statistics. However, the hypothesis can not be rejected for the year 2005. The values of N_{tail} and N for each year are summarized in Table I.

We can conclude that the power law tail hypothesis is questionable however it cannot be unequivocally rejected. In the following section we show that regardless of the validity of the power law hypothesis, the data is more likely to be explained by a log-normal hypothesis.

B. Is the distribution of fund sizes log-normal?

Whether we choose to reject the power law tail hypothesis or not we can test the data for another hypothesis. We propose that the data is better described by a log normal distribution. The log normal distribution is defined such that the density function $p_{LN}(s)$ obeys

$$p(s) = \frac{1}{s\sigma\sqrt{2\pi}} \exp\left(\frac{(\log(s) - \mu_s)^2}{2\sigma_s^2}\right)$$

and the CDF is given by

$$P(s' > s) = \frac{1}{2} - \frac{1}{2}\mathrm{erf}\left(\frac{\log(s) - \mu_s}{\sqrt{2}\sigma_s}\right).$$

A qualitative method to compare a given sample to a distribution is by a probability plot in which the quantiles of the empirical distribution are compared to the suggested distribution. Figure 5(a) is a log-normal probability plot for the size distribution of funds existing at the end of the years 1991 to 2005 while Figure 5(b) is a log-normal probability plot for the size distribution of funds existing at the end of 1998. The empirical quantiles are compared to the theoretical log-normal values. For both years, the quantiles fall on the dashed line which represents the log-normal theoretical values. Thus, for most of the tail the distribution seems to follow a log-normal distribution. However, the largest values in the tail of the distribution are above the dashed line for both the years 1991 and 2005. This means that for very large sizes the empirical distribution decays faster then a log-normal. Since a power law distribution decays slower then the log-normal this hints that a log-normal tail hypothesis is better suited.
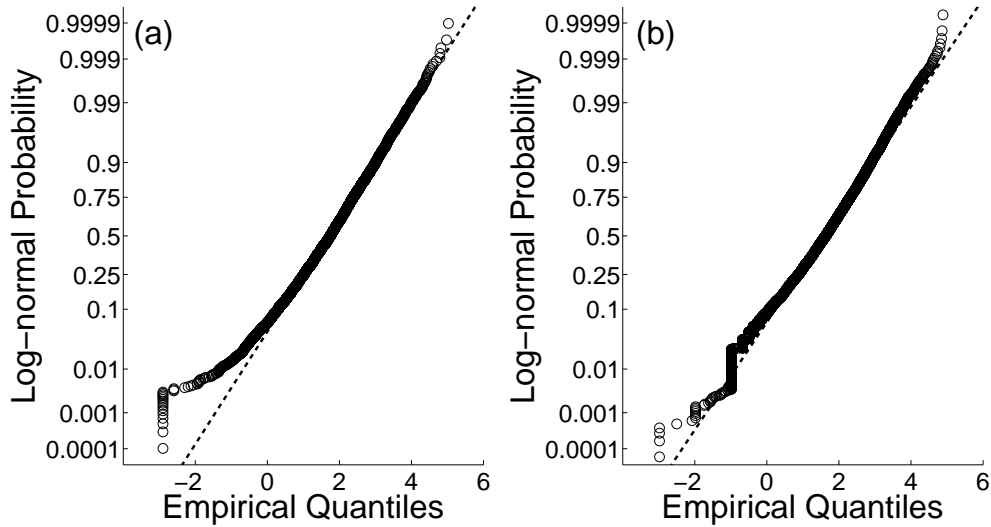
6

FIG. 5: A log-normal probability plot for the size distribution (in millions of dollars) of equity funds. The size quantiles are given in a base ten logarithm. The empirical quantiles (◦) are compared to the theoretical values for a log-normal distribution (dashed line). This is similar to a QQ-plot with the empirical quantiles on the x-axis and the theoretical quantiles for a log-normal on the y-axis. (a) The empirical quantiles are calculated from funds existing at the end of the year 1998. (b) The empirical quantiles are calculated from funds existing at the end of the year 2005.
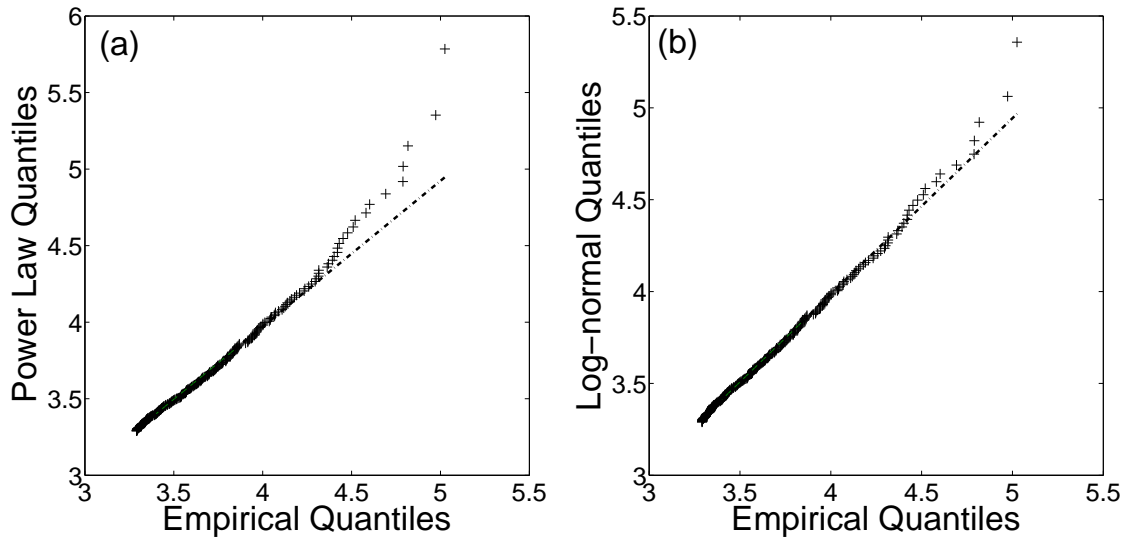

FIG. 6: A Quantile-Quantile (QQ) plot for the size distribution (in millions of dollars) of equity funds. The size quantiles are given in a base ten logarithm. The empirical quantiles are calculated from the size distribution of funds existing at the end of the year 1998. (a) A QQ-plot with the empirical quantiles as the x-axis and the quantiles for the best fit power law as the y-axis. The power law fit for the data was done using the maximum likelihood described in Section II A and the fit parameters are $s_{min} = 1945$ and $\alpha = 1.107$. The empirical data was truncated from below such that only funds with size $s \geq s_{min}$ were included in the calculation of the quantiles. (b) A QQ-plot with the empirical quantiles as the x-axis and the quantiles for the best fit log-normal as the y-axis. The log-normal fit for the data was used the maximum likelihood estimation given s_{min} (2) yielding $\mu = 2.34$ and $\sigma = 2.5$. The value for s_{min} is taken from the power law fit evaluation.
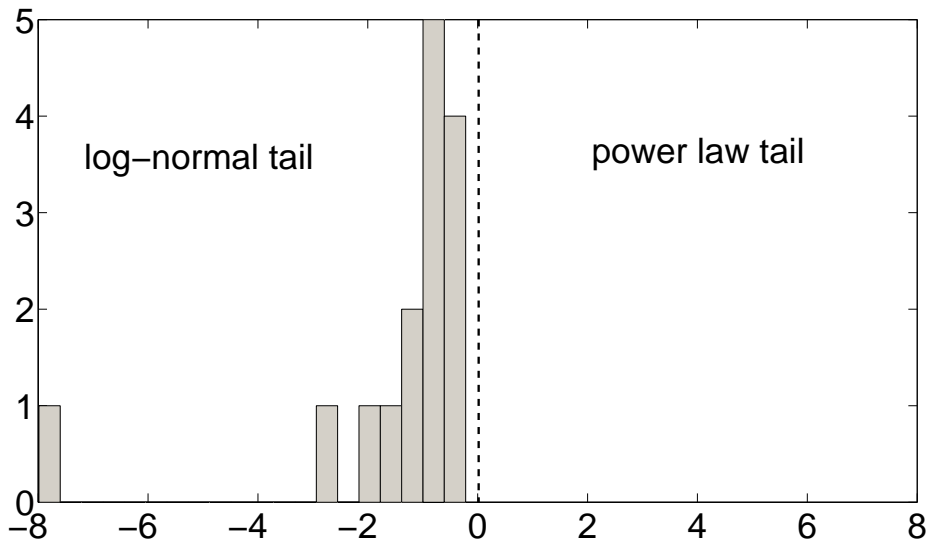
FIG. 7: A histogram of the base 10 logarithm of the log likelihood ratios $\mathcal{R}$ computed using (3) for each of the years 1991 to 2005. A negative log likelihood ratio implies that it is more likely that the empirical distribution is log-normal then a power law. Since the log likelihood ratio for any of the years is negative we can conclude that a log-normal hypothesis better suits the size distribution of funds existing at the end of the years 1991 to 2005.

A visual comparison between the two hypothesis can be made by looking at the Quantile Quantile (QQ) plots for the empirical data compared to each of the two hypotheses. In a QQ-plot we scatter plot the quantiles of one distribution as the x-axis and the other's as the y-axis. If the two distributions are the same then we expect the points to fall on a straight line. Figure 6 is a QQ-plot for the size distribution (in millions) of equity funds compared to the two hypotheses. The empirical quantiles are calculated from the size distribution of funds existing at the end of the year 1998. In Figure 6(a) the empirical quantiles are the x-axis while the quantiles from the best fit power law are the y-axis. The power law fit for the data was done used the maximum likelihood described in Section II A and the fit parameters are $s_{min} = 1945$ and $\alpha = 1.107$. The empirical data was truncated from below such that only funds with size $s \geq s_{min}$ where included in the calculation of the quantiles. (b) A QQ-plot with the empirical quantiles as the x-axis and the quantiles for the best fit log-normal as the y-axis. The log-normal fit for the data was done used the maximum likelihood estimation given s_{min} (2) yielding $\mu = 2.34$ and $\sigma = 2.5$. The value for s_{min} is taken from the power law fit evaluation. In both we can conclude that the tail of the empirical distribution decays faster then either a power law or a log-normal. The log-normal is clearly a better fit.

A more quantitative method to address the question of which hypothesis better describes the data is to compare the likelihood of the observation in both hypotheses. We define the likelihood for the tail of the distribution to be

$$L = \prod_{s_j \geq s_{min}} p(s_j).$$

We define the power law likelihood as $L_{PL} = \prod_{s_j \geq s_{min}} p_{PL}(s_j)$ with the probability density of the power law tail given by (1). The lognormal likelihood is defined as $L_{LN} = \prod_{s_j \geq s_{min}} p_{LN}(s_j)$ with the probability density of the lognormal tail given by

$$
\begin{aligned}
p_{LN}(s) &= \frac{p(s)}{1 - P(s_{min})} \\
&= \frac{\sqrt{2}}{s\sqrt{\pi}\sigma} \left[\operatorname{erfc}\left(\frac{\ln s_{min} - \mu}{\sqrt{2}\sigma} \right) \right]^{-1} \exp\left[-\frac{(\ln s - \mu)^2}{2\sigma^2} \right].
\end{aligned}
\tag{2}
$$

The more probable it is that the empirical sample is drawn from a given distribution, the larger is the likelihood value for that set of observations. The ratio of these likelihoods implies from which probability distribution is it more likely that the observed data follows. However it does not address the issue of the validity of such a fit to the data.

We define the log likelihood ratio as

$$\mathcal{R} = \ln\left(\frac{L_{PL}}{L_{LN}}\right). \tag{3}$$

The sign of the log likelihood is determined by the value of the ratio of likelihoods for the two distributions. If $\mathcal{R}$ is positive it implies that the likelihood for a power law tail is larger then the likelihood for the sample to be drawn from a log-normal distribution. Thus, a positive value for $\mathcal{R}$ implies that the power law hypothesis is more probable then a log-normal. A negative value for $\mathcal{R}$ implies that the log-normal tail is more probable . For each of the years 1991 to 2005 we computed the maximum likelihood estimators for both the power law fit and the log-normal fit to the tail, as explained above and in Section II A. Using the fit parameters, the log likelihood ratio was computed (3) for each of the years 1991 to 2005 and the results are summarized in Figure 7. The histogram clearly shows that for each of the years (all points) the resulting $\mathcal{R}$ is negative, implying that the log-normal hypothesis is more probable. We conclude that the log-normal tail hypothesis is better suited to describe the empirical data.

III.   EMPIRICAL JUSTIFICATION FOR THE MODEL

So far we have shown that the upper tail of the size distribution is better described as a log-normal than a power law as was previously . In the following sections we attempt to justify the observation through a stochastic growth model. We wish to construct a model that is as simple as possible while capturing the essence of the dynamics of the mutual funds as observed from the empirical data. Our empirical analysis is carried out using the CRSP Survivor-Bias-Free US Mutual Fund Database. This data base enables us not only to study the sizes over time but also to investigate the mechanism of growth. We investigate the time behavior of equity funds over the years 1991 to 2005. We define an equity fund as a fund with a portfolio consisting of at least 80% stocks. Even though the data base has data on mutual funds dating back to 1961 the data on equity mutual funds as defined above is available for funds mainly from the year 1991. Thus, we chose to treat the data as if no equity funds were available prior to 1991.

We define a model consisting of three possible events; creation of funds, annihilation of funds and changes in the size of funds. In the following section we investigate these three processes while trying to characterize it on the simplest manner. Some of the characteristics of these processes will be neglected in order to keep the model as simple as possible. Nevertheless, we show later on that this simple model captures the essence of the process.

We begin by analyzing the growth process in which is the process in which mutual funds change their size. The growth is decomposed into two mechanisms; return and money flux. The first process is that of the mutual fund performance in which the assets under management change due to the return. The second growth process is the flux of money from investors. This flux can be either negative corresponding to investors leaving the fund or positive corresponding to money invested in the fund. However, we can only measure the net flux of money.

The data set has monthly values for the Total Assets Managed (TASM) by the fund and the Net Asset Value (NAV). Using these values we can define the size s of a fund (in millions of US dollars) at time t as $s(t) \equiv TASM(t)$ and the fractional change in the fund size s as

$$\Delta_s(t) = \frac{s(t+1) - s(t)}{s(t)}. \tag{4}$$

The return at time t is defined by measuring the fractional change in the Net Asset Value (NAV) of the fund at time t

$$\Delta_r(t) = \frac{NAV(t+1) - NAV(t)}{NAV(t)}. \tag{5}$$

The fractional change corresponding to net flux of money Δ_f is defined as

$$\Delta_f(t) = \frac{s(t+1) - [1 + \Delta_r(t)]s(t)}{s(t)}. \tag{6}$$

The fractional change in size due to return Δ_r and the fractional change due to money flux Δ_f are related through the relation

$$\Delta_s(t) = \Delta_f(t) + \Delta_r(t). \tag{7}$$

Clearly, these two mechanisms, return and net money flux, are correlated since a fund yielding high returns will attract investors while a poor performance by a fund will deter investors. For work done on correlations between the flux of money into or out of a mutual fund see [7, 8, 22, 31, 34] and references there in.
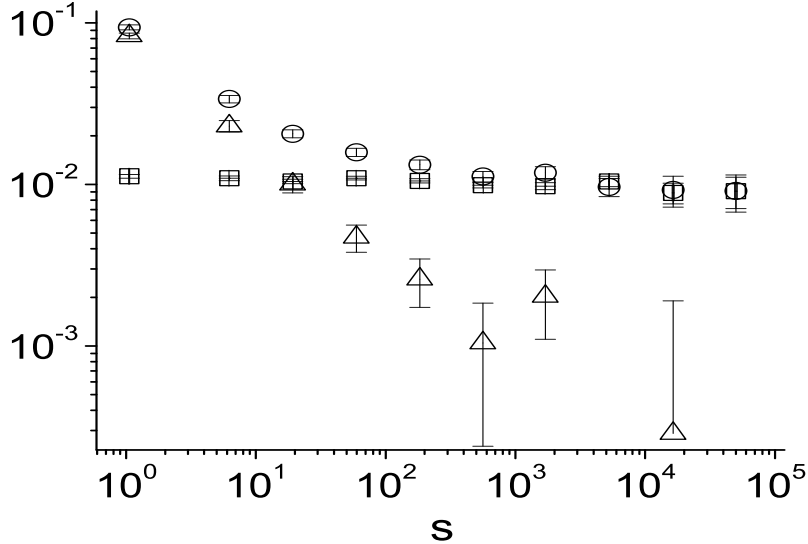
9

FIG. 8: The fractional changes in a funds size as a function of size (in millions). The total fractional size change Δ_s (4), the return Δ_r (5) and the fractional flux of money Δ_f (6) are represented by ($\circ$), ($\square$) and ($\triangle$) respectively. The calculation is for fund size changes in the year 2005. The data was binned into 8 logarithmically spaced bins. The error bars were calculated as the standard error of the mean (in both x and y) in each bin.

Another interesting question is whether there is any size dependence in these changes. To help answer this question, the empirical fractional size changes; Δ_s, Δ_r and Δ_f were calculated for the year 2005 and are plotted as a function of the fund size s in Figure 8. The return Δ_r, as can be seen in Figure 8, seems to be independent of size. Meaning, that on average the performance of a fund is independent of the fund size and equity funds can be viewed as a constant return to scale industry[17]. We find this to be in agreement with an efficient market in which we can not achieve greater return by simply going to larger funds. This observation of efficiency in regard to fund size is complimentary to the believed efficiency with respect to performance [27]. In contrast to the return, the money flux Δ_f decays with the fund size. Thus for small sizes the dominating mechanism is money flux whereas returns are the dominating growth mechanism for large funds sizes. This will be further investigated in a following paper. For the purpose of this work we concentrate on the total change Δ_s.

The size movements are modeled as an iid random process such that a size change Δ_s at time t is uncorrelated with previous size changes. At first this seems like an improper assumption since a size change consists of return and money flux (7) components that were shown to be correlated. Nevertheless, although Δ_s is correlated with past performance since the market is assumed to be efficient, we can regard performance (return) to be a random process and as such so will be the size movements. We can conclude then that an iid approximation for the size changes is a valid approximation.

As a first approximation, we model the distribution of size changes as a log-normal. In such an approximation the logarithmic changes are modeled as a normal random variable $N(\mu, \sigma)$ with a mean μ and standard deviation σ which are independent of the fund size in agreement with a Gibrat type process. This is a simplification that allows us to solve the model analytically and to gain insight into the time evolution of the distribution. We will later show that there is a size dependence for both the drift and standard deviation terms and we will offer modified models to take these size considerations into account.

The parameters μ and σ were measured for each of the years 1991 to 2005 by taking the mean and standard deviation of the log size changes. The results for each of the years, their mean and their standard error are summarized in Table I. The standard error was calculated by dividing the standard deviation by the square root of the number of years. The rates we use here after in our numeric analysis are the means of these yearly rates as given in Table I.

next we examine the creation of new funds. We investigate both the number of funds created each year $N_c(t)$ and the sizes in which they are created. Using a linear regression we find no statistically significant dependence between
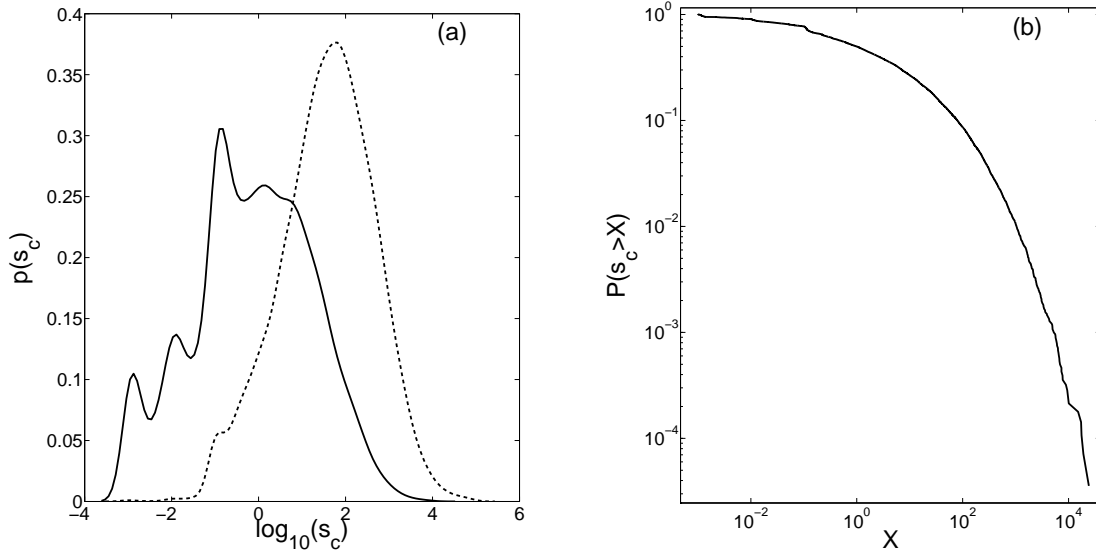
FIG. 9: (a) - The creation size s_c Probability Density Function (PDF) for funds created in the years 1991 to 2005 (full line). The creation size PDF (full line) is compared to the size PDF of equity funds existing in the end of 2005 (dashed line). The PDF estimate for the size of created funds s_c was calculated using a gaussian kernel smoothing technique. The kernel smoothing window is optimal for a normal density (MATLAB built in function). The PDF was calculated using the sizes s_c of all the funds created in the years 1991 to 2005. (b) - The Cumulative Probability Density $P(s_c > X)$ for the sizes of funds created in the years 1991-2005. The CDF was calculated using order ranking of the empirical data.

$N_c(t)$ and N$(t - 1)$. Thus, we approximate the creation of funds as a Poisson process with a constant rate ν. The number of funds as a function of time is plotted in Figure 3 and suggests a linear time dependence for the number of funds $N(t)$.

In Figure 9(a) we plot the PDF for the size s_c of new funds. The density function is estimated using a gaussian kernel smoothing technic. The density was estimated for the aggregated data for all created funds. In Figure 9(b) the CDF $P(s_c > X)$ of new fund sizes is plotted. With the above plot as motivation we approximate the probability of a new fund to be created with a size ω, $f(\omega)$, as a normal in log size with a mean $\mu_s = 0$ and standard deviation $\sigma_s = 3$. In order to gauge qualitatively how the empirical initial fund size distribution differs from a log-normal we use a probability plot Figure 10. The empirical initial fund size distribution differs mostly in the upper and lower tails as can be seen in Figure 10 where the data points diverge from the straight line. For the lower tail (small initial fund sizes) the data points fall beneath the line which means that the empirical distribution is heavier tailed. In contrast, for the upper tail (large initial fund sizes) the data points fall above the line which means that the empirical distribution is lighter tailed compared with a log-normal. In the mid of the Data point falling on the dashed line correspond to quantiles which are in agreement with a log-normal. Thus, we can safely approximate the process as log-normal since the upper tail of the creation size distribution decays faster then a log-normal and as such can not be regarded as a potential source for heavy tails in the size distribution. Moreover, in Figure 10 the PDF for the size of new funds is compared to the size PDF of funds existing at the end of 2005. It is clear that the creation process is not responsible for the observed size distribution.

The third process is the annihilation of funds. We argue that the annihilation process is such that the total number of annihilated funds is proportional to the number of funds, i.e. there is a size dependent probability $\lambda(\omega)$ for any given fund to be annihilated. In Figure 11 we plot the number of annihilated funds $N_a(t)$ as a function of the total number of equity funds existing at the previous year $N(t - 1)$. Using a linear regression we conclude that $N_a(t)$ depends linearly on $N(t)$. As a result, we define the annihilation process as follows; each existing fund is annihilated with a rate $\lambda(\omega)$ that might depend on the fund size ω. If we define the number density as $n(\omega, t)$, the rate Λ of annihilated funds is given by

$$\Lambda = \int_{-\infty}^{\infty} \lambda(\omega) n(\omega, t) \mathrm{d}\omega.$$

If we make the simplifying assumption that the rate is independent of size, the rate λ is just the slope of the linear
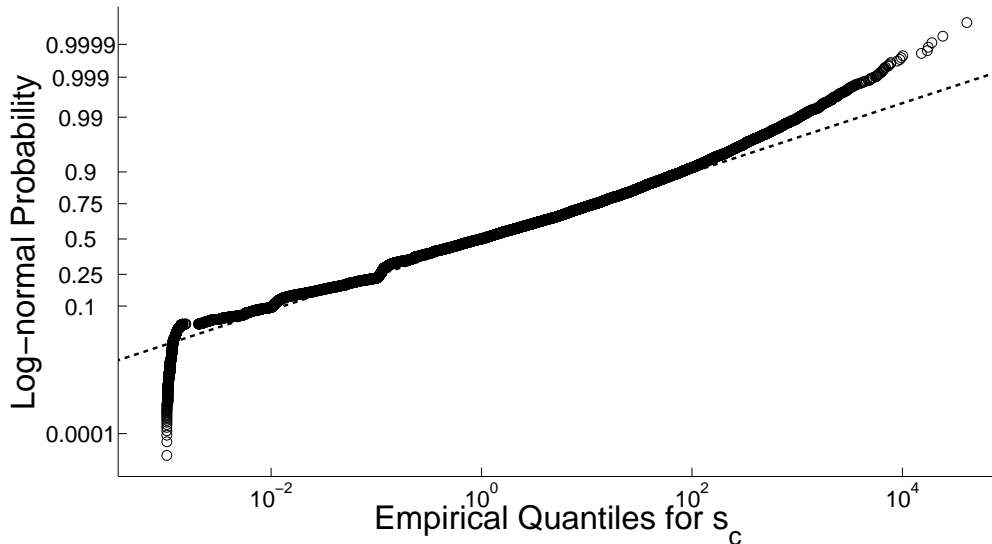
11

FIG. 10: A log-normal probability plot for the size distribution (in millions) of new equity funds. The empirical quantiles (∘) are compared to the theoretical values for a log-normal distribution (dashed line). This is similar to a QQ-plot with the empirical quantiles on the x-axis and the theoretical quantiles for a log-normal on the y-axis.

regression in Figure 11 yielding an annihilation rate per annum of $\lambda = 0.092 \pm 0.015$ at a 95% confidence level. Since we assume the annihilation to be a Poisson process the monthly rate is just the yearly rate divided by the number of month per year. Using simulations, we verified that the size dependence of the annihilation rate does not change the essence of the dynamics and this simplification is justified.

IV. A MODEL FOR THE MUTUAL FUND GROWTH DYNAMICS

We begin with a simple model for the growth dynamics of mutual funds. A simple model for stochastic growth of firms was proposed by Simon et al [32] and more recently similar model describing the growth of equity funds was proposed by Gabaix et al [16]. There are key differences in the approach we offer. First, they model the tail of the distribution taken to be funds with size in the top 15% whereas we model the entire spectrum of fund sizes. We define an equity fund as a fund with a portfolio containing at least 80% stock for a larger sample that will ensure better statistics. We found that the distribution is the same even for a more strict demand of 95%. However, such a strict constraint leads to smaller samples and weaker statistical statements. The model proposed by Gabaix et al has a fund creation rate which is linear with the number of funds whereas the data suggests that the rate has no linear dependence on the total number of funds. Moreover, the model is solved for the steady state distribution whereas we solve for the time dependence and show that the asymptotic power law behavior is reached only after a long time.

In our model, the number of mutual funds increases with time. The growth is both in the number of funds and in the total size (money) of the ecology. We denote by $N(t)$ the number of funds at time t and by $S(t)$ the total size at time t. The number density of funds of size s at time t is denoted by $n(s,t)$. We begin by assuming that the size of a fund follows a geometric brownian motion

$$\mathrm{d}s(t) = s(t)\left(\mu \mathrm{d}t + \sigma \mathrm{d}W_t\right), \tag{8}$$

where W_t is $N(0,1)$, a mean zero and unit variance normal random variable. For simplicity we will work with the natural logarithm of the mutual fund size which is denoted by $\omega \equiv \log(s)$ which satisfies the following stochastic evolution

$$\mathrm{d}\omega(t) = \mu \mathrm{d}t + \sigma \mathrm{d}W_t. \tag{9}$$

The Fokker-Planck equation (also known as the forward Kolmogorov equation) for the number density can be written as

$$\frac{\partial}{\partial t} n(\omega, t) = \left[-\mu \frac{\partial}{\partial \omega} + \frac{\sigma^2}{2} \frac{\partial^2}{\partial \omega^2} \right] n(\omega, t). \tag{10}$$
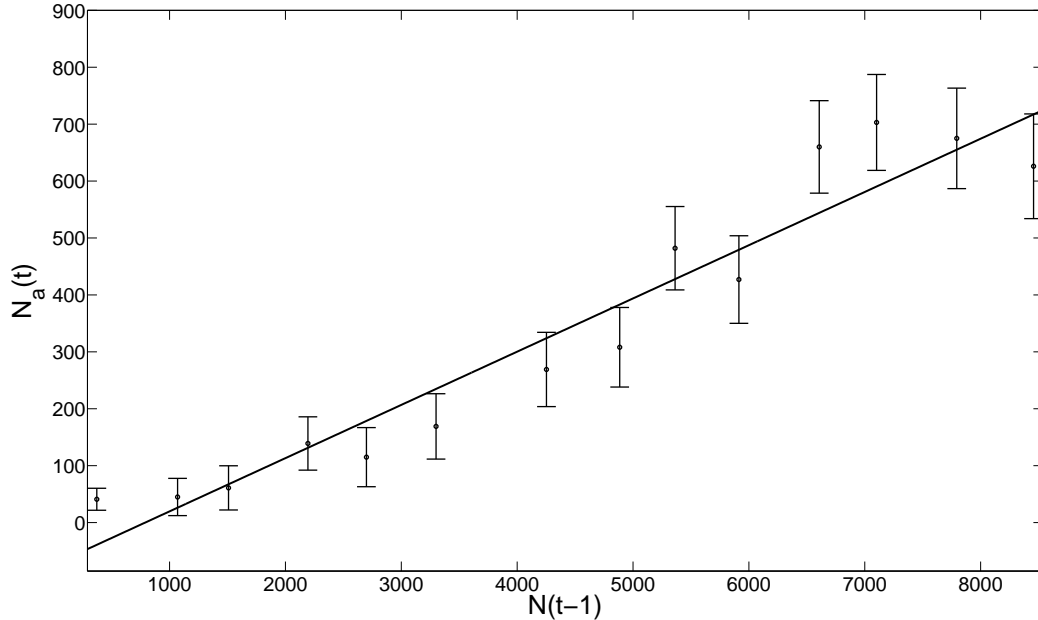
12

FIG. 11: The number of equity funds annihilated $N_a(t)$ in the year t as a function of the total number of funds existing at the previous year $N(t-1)$ ($\circ$). The plot is compared to a linear regression of the data points (full line). The error bars were calculated for each point under a Poisson process assumption to be the square root of the average number of annihilated funds at that year. The average number of annihilated funds in each year was taken to be $N_a(\bar{t}-1) = \sqrt{\lambda N(t-1)}$ where λ was calculated as the slope of the linear regression to the points.

However, not only do the fund sizes change but funds can be created and annihilated.

We denote by $\lambda(\omega)$ the rate in which a fund of size ω is annihilated. This can be restated so that any fund of size ω has a probability of $\lambda(\omega)\mathrm{d}t$ to be annihilated from time t to $t+\mathrm{d}t$. The rate of creation of new funds is defined as ν . The creation process is such that at time t a new fund is created with a probability $\nu\mathrm{d}t$. The size of these new funds is distributed with a distribution $f(\omega, t)$. Thus, at time t, when a new fund is created, it has a size between ω' and $\omega' + \mathrm{d}\omega'$ with a probability $f(\omega', t)\mathrm{d}\omega'$.

The time evolution of the number density can be written as

$$\frac{\partial}{\partial t}n(\omega, t) = -\lambda(\omega)n(\omega.t) + \nu f(\omega, t) + \left[-\mu\frac{\partial}{\partial \omega} + \frac{\sigma^2}{2}\frac{\partial^2}{\partial \omega^2}\right]n(\omega, t). \tag{11}$$

Thus, given the creation and annihilation rates λ and ν, the size distribution of new funds $f(\omega, t)$ and the initial condition $n(\omega, 0)$ we can solve for the number density using (11). The total number of funds is defined as

$$N(t) = \int_{-\infty}^{\infty} n(\omega', t)\mathrm{d}\omega' \tag{12}$$

and obeys the following time evolution

$$\frac{\mathrm{d}N(t)}{\mathrm{d}t} = -\int_{-\infty}^{\infty} \lambda(\omega)n(\omega, t)\mathrm{d}\omega + \nu. \tag{13}$$

Once the number density is known it is straight forward to calculate the probability density through

$$p(\omega, t) = \frac{n(\omega, t)}{N(t)}. \tag{14}$$

Since the total number of funds $N(t)$ is independent of size ω, the probability density and the number density exhibit the same size dependence up to a time dependent constant. Thus, we can solve for the number density instead of the probability density in order to examine the size dependence of the distribution.

13

V. ANALYTICAL SOLUTION FOR THE NUMBER DENSITY $n(\omega, t)$

We define the dimensionless size $\tilde{\omega} = (\mu/D)\omega$ where $D = \sigma^2/2$, the dimensionless time $\tau = (\mu^2/D)t$ and the rate $\gamma = 1/4 + (D/\mu^2)\lambda$ for which the Fokker-Plank equation is written in the simple form

$$\left[\frac{\partial}{\partial \tau} + \gamma - \frac{\partial^2}{\partial \tilde{\omega}^2}\right] \eta(\tilde{\omega}, \tau) = \frac{D}{\mu^2} \nu e^{-\tilde{\omega}/2} f(\tilde{\omega}, \tau). \tag{15}$$

Using a Laplace transform to solve for time dependence and a Fourier transform to solve for the size dependencethe time dependent number density $n(\omega, t)$ can now be calculated for a given source $f(\omega, t)$.

A.   An impulse response (Green's function)

The Fokker-Planck equation in (15) is given in a linear form

$$\mathcal{L}\eta(\tilde{\omega}, \tau) = \mathcal{S}(\tilde{\omega}, \tau), \tag{16}$$

where $\mathcal{L}$ is a linear operator and $\mathcal{S}$ is a source function. We define the Green's function $G(\tilde{\omega}, \tau)$ to be the solution for a point source in both size and time

$$\mathcal{L}G(\tilde{\omega} - \tilde{\omega}_0, \tau - \tau_0) = \delta(\tilde{\omega} - \tilde{\omega}_0)\delta(\tau - \tau_0), \tag{17}$$

where δ is the Dirac delta function . Using the Green's function, the number density for any general source can be written as

$$\eta(\tilde{\omega}, \tau) = \int \int \mathcal{S}(\tilde{\omega}_0, \tau_0)G(\tilde{\omega} - \tilde{\omega}_0, \tau - \tau_0)\mathrm{d}\tilde{\omega}_0 \mathrm{d}\tau_0. \tag{18}$$

We solve for the Green function using a source of the form

$$\frac{D}{\mu^2} \nu e^{-\tilde{\omega}/2} f(\tilde{\omega}, \tau) = \delta(\tilde{\omega} - \tilde{\omega}_0)\delta(\tau - \tau_0). \tag{19}$$

This is a source for funds of size w_0 generating an impulse at $t = t_0$. We will assume that prior to the impulse at $t = 0^-$ there were no funds which means that the initial conditions are $\eta(\tilde{\omega}, 0^-) = 0$. The Green's function can be solved analytically as described in Appendix A. The solution is given by

$$G(\tilde{\omega} - \tilde{\omega}_0, \tau - \tau_0) = \frac{1}{\sqrt{\pi 4(\tau - \tau_0)}} \exp\left[-\frac{(\tilde{\omega} - \tilde{\omega}_0)^2}{4(\tau - \tau_0)} - \gamma(\tau - \tau_0)\right] \theta(\tau - \tau_0). \tag{20}$$

B.   A continuous source of constant size funds

After obtaining the green's function we will look at a continuous source of funds of size ω_s starting at a time t_s which can be written as

$$f(\omega, t) = \delta(\omega - \omega_s)\theta(t - t_s). \tag{21}$$

As described in Appendix A the large times limit of the upper tail of the number density, calculated using an asymptotic expansion in large $\tilde{\omega}$, is given by

$$n(\tilde{\omega}, \tau) = \exp\left(\frac{\tilde{\omega} - \tilde{\omega}_s}{2} - \sqrt{\gamma}|\tilde{\omega} - \tilde{\omega}_s|\right)\frac{\nu D}{2\mu^2\sqrt{\gamma}}\left[\mathrm{erf}\left\{2\sqrt{\gamma}\left(\frac{(\tilde{\omega} - \tilde{\omega}_s)^2}{4\gamma}\right)^{1/4}\right\} + \mathrm{erf}\left\{2\sqrt{\gamma}\left(\sqrt{\tau - \tau_s} - \left(\frac{(\tilde{\omega} - \tilde{\omega}_s)^2}{4\gamma}\right)^{1/4}\right)\right\}\right].$$
$$\tag{22}$$

Since $\gamma > 1/4$ the density is final for both $\tilde{\omega} \to \infty$ and $\tilde{\omega} \to -\infty$. The steady state number of funds is given by

$$\lim_{t \to \infty} N(t) = \frac{\nu}{\lambda}. \tag{23}$$

14

This number is such that the total number of funds created is equal to the number of funds annihilated For an annual creation rate of $\nu = 700 \pm 300$ and an annihilation rate of $\lambda = 0.092 \pm 0.015$ the steady state number of funds is expected to be $N = 7600 \pm 3500$ which can be larger then the 8845 equity funds existing at the end of 2005. The value for ν was calculated as the constant parameter in a linear regression between the number of created funds $N_c(t)$ in year t and the total number of funds existing in the previous year $N(t-1)$. The value for λ was calculated as the proportion parameter in a linear regression for the number of annihilated funds $N_a(t)$ in year t and the total number of funds in the previous year $N(t-1)$ as can be seen in Figure 11.

We can now determine a more quantitative notion of large sizes or large times. Given the above asymptotic solution the size is considered large if

$$4\gamma(\tilde{\omega} - \tilde{\omega}_s)^2 \gg 1, \tag{24}$$

which can be rewritten as

$$\omega - \omega_s \gg \frac{D}{4\mu\gamma} = \frac{D}{\mu + 4D\lambda/\mu}. \tag{25}$$

Using the values for the monthly rates given in Table I we get that $D/(4\mu\gamma) \approx 0.5$ which means that we should define large size to be

$$s \gg s_s, \tag{26}$$

where s_s is the size of funds created by the source in millions of dollars approximated as $s_s = 1$ (in millions). By demanding that at large times the solution be time independent, we arrive at the following definition for large times

$$t - t_s \gg \frac{|\omega - \omega_s|}{2\mu\sqrt{\gamma}} = \frac{|\omega - \omega_s|}{\sqrt{\mu^2 + 4D\lambda}}. \tag{27}$$

Using the measured monthly rates in Table I and the value for $\lambda = 0.092$ given in Section III we get that $1/(\sqrt{\mu^2 + 4D\lambda}) \approx 420$ and the large times condition is written as

$$t - t_s \gg 420\,|\omega - \omega_s|. \tag{28}$$

Since we use t is units of months, this means that even for a fund of size $\omega = 1$ a large time is considered such that it is much larger than approximately 40 years. For large times the number density is independent of time and is given, as a function of the dimensional parameters, by

$$n(\omega) = \frac{\nu}{2\sqrt{\gamma}} \exp\left(\frac{\mu(\omega - \omega_s)}{2D} - \frac{\mu\sqrt{\gamma}|\omega - \omega_s|}{D}\right). \tag{29}$$

Since the steady sate number of funds is constant we can calculate the probability density by dividing by $N = \nu/\lambda$

$$p(\omega) = \frac{\lambda}{2\sqrt{\gamma}} \exp\left(\frac{\mu(\omega - \omega_s)}{2D} - \frac{\mu\sqrt{\gamma}|\omega - \omega_s|}{D}\right). \tag{30}$$

The tail distribution is given by

$$p(w) \sim n(w) \sim \exp\left(\frac{\mu}{D}(\frac{1}{2} - \sqrt{\gamma})(\omega - \omega_s)\right). \tag{31}$$

For which the CDF has a power law tail with an exponent ζ_s such that

$$P(s > X) \sim X^{-\zeta_s}, \tag{32}$$

with

$$\zeta_s = \frac{-\mu + \sqrt{\mu^2 + 4D\lambda}}{2D}. \tag{33}$$

Using the parameter values in Table I the exponent has the value

$$\zeta_s = 0.18 \pm 0.04. \tag{34}$$

The asymptotic value for the exponent is smaller then the measured exponents from the empirical data assuming a power law tail which are given in Table I. Since the distribution is currently log-normal we expect that if we were to measure the exponent for a power law tail fit it will be larger since the tail has to widen with time. This is indeed the case.
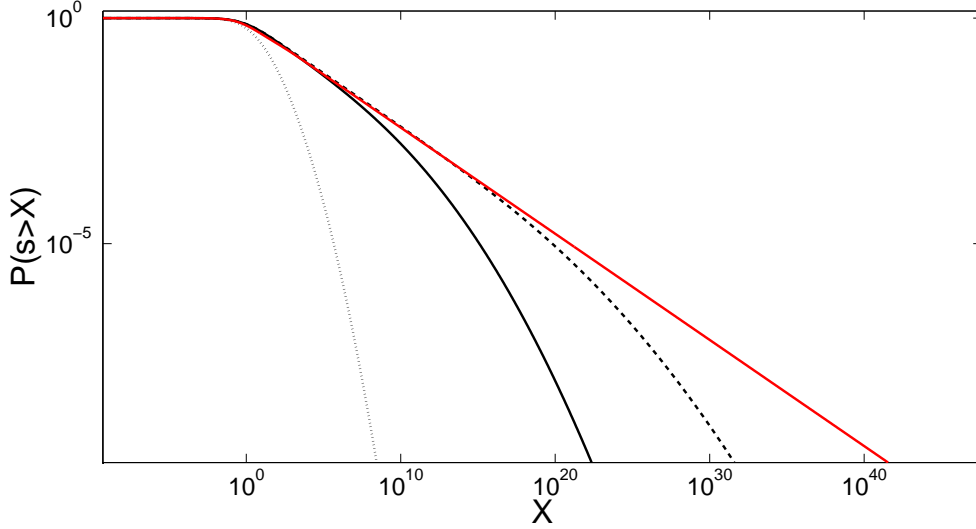
FIG. 12: The CDF from numerical calculation for time horizons of 5, 50 and 100 years given by (from left to right) the dotted, full and dashed lines respectively. The distributions are compared to the $t \to \infty$ distribution represented by right full line.

C. A normal source of funds

The empirical observations suggest that the funds are created with a lognormal distribution in the fund sizes s or a normal distribution in the log sizes ω. The above analysis can be easily applied to a source of the type

$$f(\tilde{\omega}, \tau) = \frac{1}{\sqrt{\pi \sigma_s^2}} \exp\left(-\frac{(\tilde{\omega} - \tilde{\omega}_s)^2}{\sigma_s^2}\right) \theta(\tau - \tau_s). \tag{35}$$

By solving for $\eta(\tilde{\omega}, \tau)$, as described in Appendix A, we get that the solution for the number density is given by

$$n(\tilde{\omega}, \tau) = \int_{\tau_s}^{\tau} \frac{\nu D}{\mu^2 \sqrt{\pi \left(\sigma_s^2 + 4(\tau - \tau')\right)}} \exp\left[\frac{(\tau - \tau')(\sigma_s^2(1 - 4\gamma) + 8\tilde{\omega}) - 4(4\gamma(\tau - \tau')^2 + (\tilde{\omega} - \tilde{\omega}_s)^2 + 2(\tau - \tau')\tilde{\omega}_s)}{4(\sigma_s^2 + 4(\tau - \tau'))}\right] d\tau'. \tag{36}$$

There is no closed form (that we could get) for (36) and we leave the solution in the integral form. The number density can be calculated numerically using (36) and indeed the resulting distribution is normal for short time scales and becomes exponential in ω (power law in s) for long times. In Figure 12 the CDF for the fund size s was calculated numerically using (36) for time horizons of 5, 50 and 100 years. As expected, the size distribution seems to converge towards a power law. The convergence is such that, for any finite time horizon, the tail of the distribution is lognormal while the a larger portion of the CDF becomes a power law as the time horizon increases. Only at infinitely large times does the distribution become a complete power law. This is due to the fact that the Green function is log-normal and for the tail to become a power law funds must diffuse in size space towards larger and larger sizes.

VI. SIMULATION OF THE MODEL

In this section we describe simulation results of the above growth model and compare the resulting probability density to the one given above.

A. The Simulation Model

The details of the actual simulation of the model are given in Appendix B. The rates used in the simulation were calculated from the empirical data. The annihilation rate was calculated using a linear regression as shown in Figure 11. The annihilation rate was taken to be $\lambda = 0.092/12$ per month. The creation rate was taken to be
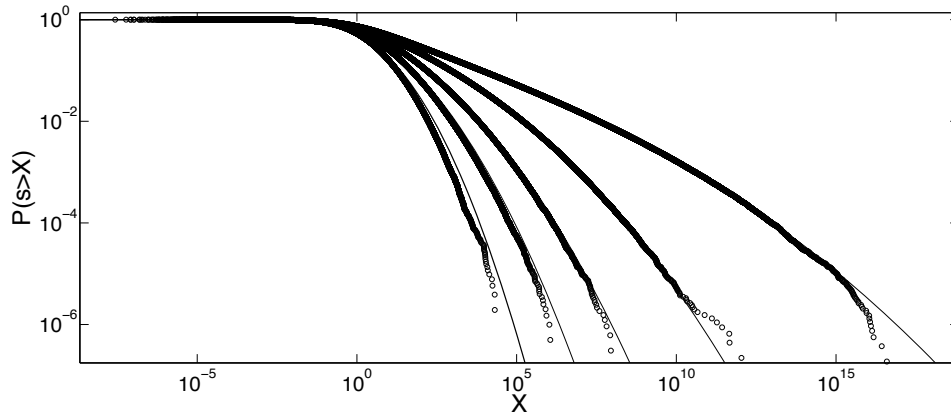
FIG. 13: The cumulative size density $P(s > x)$ from simulations ($\circ$) is compared to the numerical calculation of the cumulative size density (full line). The comparison is given for the time horizons:1,5,10,20 and 50 years from left to right respectively.

$\nu = 500/12$ per month. The value of ν affects only the number of funds in the simulation and does not affect the resulting distribution.

The parameters for the size changes were calculated from the CRSP monthly data for equity funds in the years 1991 to 2005 and are given in Section III. The drift μ was calculated as the mean of the logarithmic size changes and σ was calculated as the standard deviation of these logarithmic size changes. The results were $\mu = 0.038$ and $\sigma = 0.247$.

The new fund source was taken to be a mean zero normal (log normal in s) with a standard deviation of 3, $f(\omega) = N(0, 3)$. The initial state for the simulation is no funds. At time $t' = 0$, the source is introduced and funds are created in a poisson process. The simulation continues up to time $t' = t$ which is the defined time horizon. At this time we record the size distribution. By repeating the process many times better statistics are achieved.

B. Simulation Results

The simulation was carried out with the above parameters for several times t where the observation of funds sizes is made. Each simulation was repeated 1000 times for each time horizon and the cumulative probability density $P(s > X)$ was calculated. In Figure 13 the CDF obtained from simulation is compared to the numerical calculation using (36) calculated for the same time horizon. The comparison is given for varying time horizons of 1,5,10,20 and 50 years. It is apparent from the plot that as time progresses the distribution converges to a power law as expected. Nevertheless, the time scales are quite large and are of the order of 100 years.

A comparison between the distribution obtained by simulating the model and the empirical distribution is given in Figure 14(a) and Figure 14(b). Since the number of equity funds is negligible at the beginning of 1990 we can consider the dynamics to have started then resulting in a time horizon for the year 1998 of 8 years. Thus, the empirical distribution in Figure 14(a) and Figure 14(b) is compared to a simulation with an evolution time horizon of 8 years. Even though the results from this simplified model do not completely coincide with the empirical data it captures important aspects of the dynamics and more importantly it emphasizes the transient state in which the mutual fund ecology is currently residing.

VII. A MORE REALISTIC MODEL

The simplified model given is by no means an exact model for the mutual fund growth process. However it does convey the transient phase for which the distribution of fund sizes is better described by a lognormal. Using the simplistic model we were able to calculate the time scales in which the distribution converges to the steady state distribution. However, the growth process is surprisingly more complex and exhibits two very interesting phenomena.

The first phenomena is that the variation in growth rate σ seems to decay with size. A similar phenomena was observed for the processes of firm growth [1, 4, 36, 37]. It was found that the standard deviation of the growth process obeys the same power law with the same exponent. This is not too surprising if one takes into account that
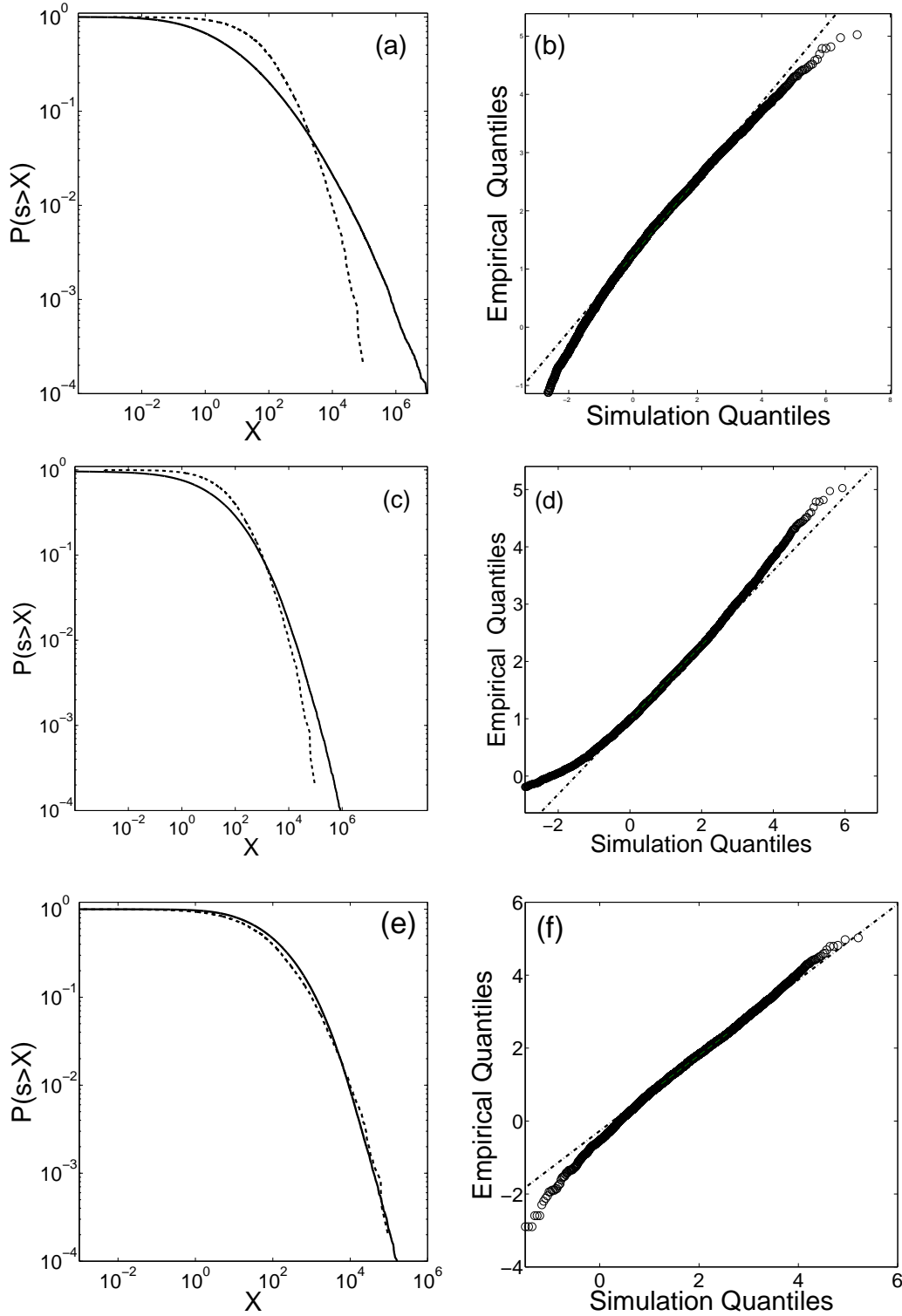
17

FIG. 14: In the following plots we compare the results from simulating three models with the empirical distribution. The comparison is between the empirical distribution at the end of 1998 given and the CDF obtained from simulation with the corresponding time horizon of 8 years. For each model the results are given by a pair of plots (a row). The left column of plots are comparisons between the cumulative size density $P(s > x)$ from simulations (full line) to the empirical cumulative size density (dashed line). The right column of plots are Quantile-Quantile (QQ) plots of the empirical distribution quantiles in the Y-axis and the simulation quantiles as the X-axis. The quantiles were calculated for the logarithm of the of the sizes (in millions). The first row, Figures (a) and (b) are simulation results of the simple model described in Section VI. The second row, Figures (c) and (d) are simulation results of the model with a size dependent diffusion constant (39). The third row, Figures (e) and (f) are simulation results of the model with a size dependent diffusion constant (39) and drift term (40).
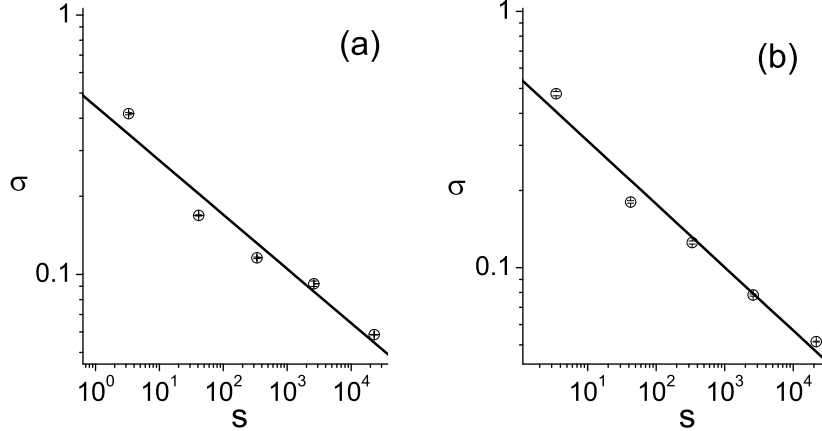
18

FIG. 15: The standard deviation in the logarithmic size change of an equity fund σ as a function of the fund's size s. The variance was calculated for the aggregated data; (a) for the years 1991 to 2006 and (b) for the year 1998. The data is compared to a linear regression for the logarithms.

funds are firms belonging to the same industry. In Section VII A we modify the simplistic model to incorporate the size dependence of the diffusion constant $D = \sigma^2/2$. The model is then simulated and compared to the empirical distribution.

The second interesting observation is that the average growth rate μ is larger for smaller funds then it is for larger ones as was described in Figure 8 in Section I. As was noted in the beginning the dominating form of growth for small funds is by influx of money whereas for larger ones it is return on the assets. Using a linear regression between the return and the fund size we find, surprisingly, that there is no simple correlation between return and mutual fund size. As a result, the growth rate becomes constant for large funds for which the return on assets is the dominating source for growth. In Section VII B we incorporate the decaying growth rate into our model. The model is then simulated and compared to the empirical distribution. The resulting distributions (from simulations) for different years are compared to the corresponding empirical distributions. The agreement is overwhelmingly good. Thus, it seems that this modified model is complex enough to capture the essence of the process and using it we can predict the future shape of the equity mutual fund size distribution.

A.   model 2.1

A striking result is that the variance in the logarithmic size change is size dependent with a power law dependence of the form

$$\sigma(s) \sim s^{-\beta}, \tag{37}$$

with an exponent of $\beta \approx 0.2$. This can be seen in Figure 15(a) and Figure 15(b) where the standard deviation was binned into five logarithmically spaced bins with respect to the fund size. The fund size was calculated as the average value of funds sizes in each bin and the standard error was calculated to be the standard deviation divided by the square root of the bin occupancy. The value of the standard deviation for each bin and the error were calculated in the same manner.

The figures suggest that a power law hypothesis relation between σ and s is a good approximation and regardless of the exact functional form of the relation, it is apparent that $\sigma(s)$ is a decreasing function of s. The consequence is that the diffusion rate in log size space for the mutual funds is decreasing with their size and as a result the tails of the distribution should remain thinner. This thinning of the tails results in an even slower conversion to the power law tail existing in the steady state as suggested by our simplistic model.

Not only does the convergence to the power law tail be slower but the exponent to which the distribution converges to is also affected by the decay of σ. For large ω we take the standard deviation to be small such that $\lambda D/\mu^2 \ll 1$ for
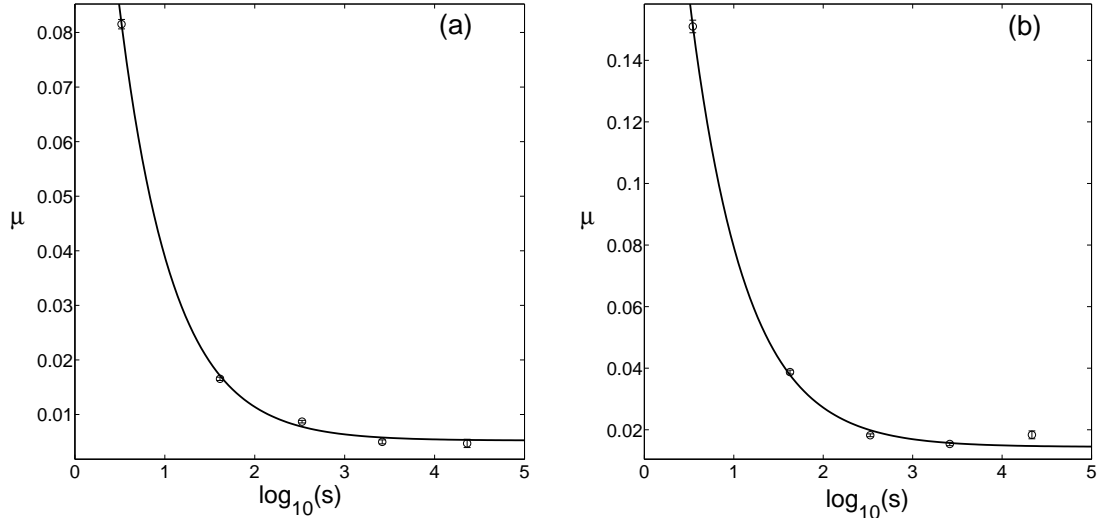
FIG. 16: (a) The mean logarithmic size change of an equity fund μ as a function of the fund's size s (in millions). The mean was calculated for the aggregated data for the years 1991 to 2005. The data is compared to a power law relation given by (40). (b) The mean logarithmic size change of an equity fund μ as a function of the fund's size s for the aggregated data for the years 1991 to 2006.

which we approximate the tail exponent which is given by (33) as

$$\zeta_s = \frac{\lambda}{\mu}, \tag{38}$$

which for the measured monthly annihilation and drift rates yields $\zeta_s \approx 0.2 \pm 0.05$. The value for the annihilation rate is $\lambda = 0.092 \pm 0.015$ given in Section III and for the drift we use $\mu = 0.038 \pm 0.007$ given in Table I.

We modified the simplistic model to have a size dependent standard deviation of the form

$$\sigma'(s) = \sigma_0 s^{-0.2}, \tag{39}$$

where $\sigma_0 = 10^{-0.25} \approx 0.56$. The model was simulated using the same values for the remaining parameters as in the previous simple model .

In Figure 14(c) and Figure 14(d) we compare the modified simulation with the empirical distribution. The comparison is for a simulation with a time horizon of 8 years which corresponds to an observation on the empirical size distribution of equity funds at the end of 1998 since the number of equity funds in the data set is negligible prior to 1991. It is clear that the distribution is closer to the observed distribution than the model with a constants standard deviation. This can be seen using the QQ-plot given in Figure 14(d) where the quantiles obtained from the modified simulation were compared to the quantiles obtained from the empirical data for the logarithm (base 10) of the size of funds (in millions) existing at the end of 1998.

B. Model 2.2

Another aspect of the growth process that we have neglected so far is the dependence of the drift term μ on the size of the fund as can be seen in Figure 8. We now wish to further enhance our model to take this size dependence into account. As a first approximation we fit a power law relation between the drift μ and the size s (exponential in the log size ω) of the form

$$\mu(s) = \mu_0 s^{-\alpha} + \mu_\infty. \tag{40}$$

This relation implies that the growth rate decays with the size of the fund. Smaller funds grow (in percent wise) faster then larger funds. For very larger sizes the growth rate becomes approximately constant with a rate μ_∞. Similar to

variable	1991-1996	1991- 1998	1991- 2002	1991- 2005
σ_0	0.21 ± 0.07	0.26 ± 0.06	0.32 ± 0.08	0.35 ± 0.07
β	0.29 ± 0.03	0.25 ± 0.03	0.21 ± 0.03	0.21 ± 0.03
μ_0	0.38 ± 0.14	0.33 ± 0.11	0.22 ± 0.04	0.19 ± 0.05
α	0.69 ± 0.26	0.7 ± 0.24	0.76 ± 0.14	0.74 ± 0.21
μ_∞	0.016 ± 0.01	0.014 ± 0.008	0.0015 ± 0.0025	0.005 ± 0.004

TABLE II: σ_0 and α - The fitted parameter values for the size dependence of the variance of the growth process (37). μ_0, α and μ_∞ - The fitted parameter values for the size dependence of the average growth rate (16). The values are given with 95% confidence for the data averaged over different time horizons . The time horizons start in the year 1991 and end in the years 1996, 1998, 2002 and 2005. The results for each parameter of the fit (line) is given in a different column from left to right corresponding to the different time horizons.

what we have done previously , we compute the mean of the log size change in each of the 5 bins and check for a size dependence in the drift term μ. The results are plotted in Figure 16(a) for the aggregated size changes of funds from the years 1991 to 2005 and in Figure 16(b) for the log size changes in the years 1991 to 1998. As can be seen in Figure 16 the functional dependence of the average growth rate μ seems to be constant over the years. However the fitted, parameters change as is expected since the overall growth of the market differs from year to year. Fitting (40) for the data, aggregated over different years, we get slightly different values for the fit parameters given in Table II.

Where $\alpha = 1 \pm 0.5$, $\mu_0 = 0.35 \pm 0.3$ and $\mu_\infty = 0.01 \pm 0.005$ with 95% confidence. Using the above value for μ_∞ in (38) for the steady state power law exponent yields $\zeta_s = 0.8 \pm 0.4$ which is in agreement with the observations.

We simulated the model with the added size dependence for the drift term and compared to the empirical distribution in Figure 14(e). The QQ-plot for the comparison of the simulation results and the empirical distribution is given in Figure 14(f) where the quantiles obtained from the modified simulation were compared to the quantiles obtained from the empirical data for the logarithm (base 10) of the size of funds (in millions) existing at the end of 1998.

The empirical evidence seems to suggest that not only does the diffusion constant and the drift term depend on size but so does the annihilation rate. The annihilation rate dependence is such that it decays with size for large sizes. Through out our models and simulations we kept the annihilation rate a constant independent of size. It is in fact the annihilation which is responsible for the formation of a power law tail. The effect of the annihilation is that most of the annihilated funds will be with smaller sizes, due to the relatively larger number, and as a result the distribution of funds changes to compensate for the loss of small funds and the upper tail gets wider. From simulation results we concluded that the fact that the annihilation rate is size dependent does not affect the size distribution as much as the size dependence of the diffusion rate and drift term.

To verify that the model captures the evolution of the size distribution the parameters of model 2.0 were calculated from the aggregated data for several periods. Using the power law relations for the decaying variance in growth rates (37) and the average growth rate (16) the value of the parameters were fit for the periods 1991 to 1996, 1998, 2002 and 2005. The value of the parameters is given in Table II. Using these estimated parameters model 2.2 was simulated for the corresponding time horizons of 6, 8 ,12 and 15 years. The comparison of the simulation results and the empirical distributions is given in Figure 17. It is apparent from the figure that the model accomplishes to capture the essence of the distribution for the different years. It is apparent from the QQ-plot that the functional forms of the two distributions are in agreement.

To conclude, the modified model takes into account the violation of Gibrat's law through the size dependence of the growth process. The empirical data suggests that there is an algebraic dependence between both the average growth rate μ and the variance of the growth process σ. These departures from Gibrat's law are such that the time scale calculated above for the convergence into a steady state distribution will be upper bounds. Thus, the time needed for the system to evolve into a state described by a power law for the tail of the distribution is even longer then the estimated times described previously. These departures from the simple model are shown to explain the difference in the observed distribution from the expected distribution from the simple model.

VIII.   CONCLUSIONS

In this work we investigated the distribution of equity fund sizes and offered a stochastic model for the growth process. We showed that the upper tail of the distribution is better described as a lognormal then a power law as our model predicts. We identified three key processes; size change, creation and annihilation of funds, as responsible
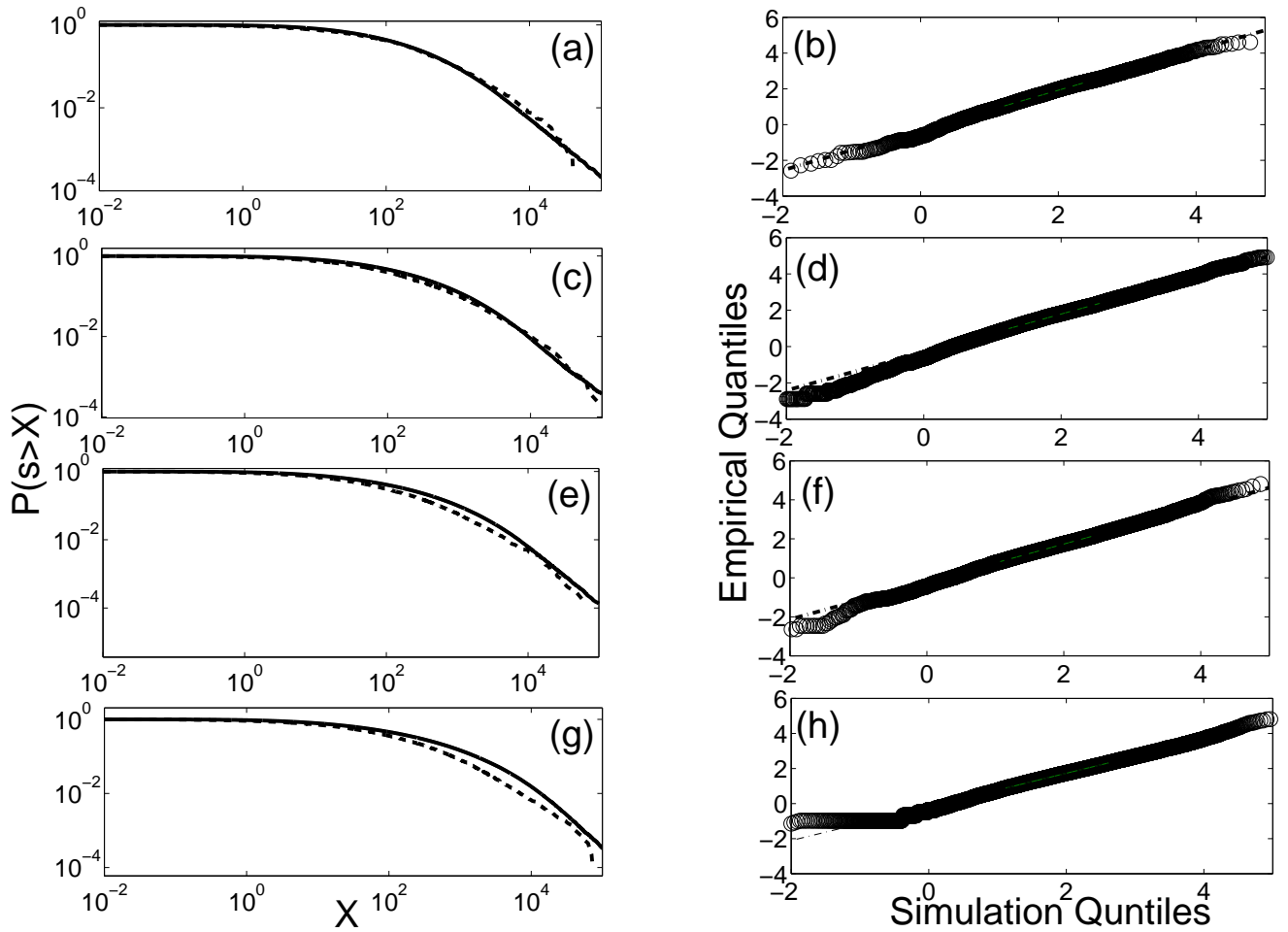
FIG. 17: Model 2.2 was simulated for different time horizons and is compared to the empirical distribution. The left Column is a comparison of the CDF obtained by simulating Model 2.2 (full line) and is compared to the empirical CDF (dashed line). The right Column is a QQ-plot comparing the two distributions. (a), (b) - The distribution at the end of 1996 is compared to a simulation of a time horizon of 6 years.
(c), (d) - The distribution at the end of 1998 is compared to a simulation of a time horizon of 8 years.
(e), (f) - The distribution at the end of 2002 is compared to a simulation of a time horizon of 12 years.
(g), (h) - The distribution at the end of 2005 is compared to a simulation of a time horizon of 15 years.
The parameters for the simulation are given in Table II

for the distribution we observe. In contrast with past work in the subject we did not assume that the distribution is stationary. We allowed the distribution to evolve with time and we were able to identify the time scales on which our model evolves. A system that can be described by our model will evolve such that the tails are initially log-normal and become power-law after a long time. We showed that the growth process governing the mutual funds is more complicated and the terms are size dependent. Nevertheless, this does not affect the functional form of the distribution but only increases the time it will take the distribution to reach a power law state. Thus, we can conclude that mutual funds obey a log-normal size distribution and given another millennia for the market to evolve we will find them in a distribution better described by a power law.

Appendix A: Analytical Solution To Model 1.0

We define the dimensionless size $\tilde{\omega} = (\mu/D)\omega$ where $D = \sigma^2/2$ and the dimensionless time $\tau = (\mu^2/D)t$ for which

$$\left[\frac{\partial}{\partial\tau} + \frac{D}{\mu^2}\lambda(\omega) + \frac{\partial}{\partial\tilde{\omega}} - \frac{\partial^2}{\partial\tilde{\omega}^2}\right]n(\tilde{\omega},\tau) = \frac{D}{\mu^2}\nu f(\tilde{\omega},\tau) \tag{A1}$$

By defining the function

$$\eta(\tilde{\omega},\tau) = e^{-\tilde{\omega}/2}n(\tilde{\omega},\tau) \tag{A2}$$

The Fokker-Plank equation (A1) is rewritten as

$$\left[\frac{\partial}{\partial\tau} + \frac{D}{\mu^2}\lambda(\omega) + \frac{1}{4} - \frac{\partial^2}{\partial\tilde{\omega}^2}\right]\eta(\tilde{\omega},\tau) = \frac{D}{\mu^2}\nu e^{-\tilde{\omega}/2}f(\tilde{\omega},\tau) \tag{A3}$$

We simplify the problem by approximating the annihilation rate as independent of size and by defining the rate $\gamma = 1/4 + (D/\mu^2)\lambda$ the Fokker-Plank equation is written in the simple form

$$\left[\frac{\partial}{\partial\tau} + \gamma - \frac{\partial^2}{\partial\tilde{\omega}^2}\right]\eta(\tilde{\omega},\tau) = \frac{D}{\mu^2}\nu e^{-\tilde{\omega}/2}f(\tilde{\omega},\tau) \tag{A4}$$

We define the Laplace transformed function $\phi(\tilde{\omega},u)$ as

$$\phi(\tilde{\omega},u) = \int_0^\infty \eta(\tilde{\omega},\tau)e^{-u\tau}d\tau \tag{A5}$$

and the Fourier transformed functions

$$\psi(k,u) = \int_{-\infty}^\infty \phi(\tilde{\omega},\tau)\frac{e^{-ik\tilde{\omega}}}{\sqrt{2\pi}}d\tilde{\omega} \tag{A6}$$

$$\psi_0(k) = \int_{-\infty}^\infty \eta(\tilde{\omega},0^-)\frac{e^{-ik\tilde{\omega}}}{\sqrt{2\pi}}d\tilde{\omega} \tag{A7}$$

$$\tag{A8}$$

For which (A4) is written as

$$\left[u + \gamma + k^2\right]\psi(k,u) = \frac{D}{\mu^2}\nu\tilde{f}(k,u) + \psi_0(k) \tag{A9}$$

with the source distribution transformed as follows

$$\tilde{f}(k,u) = \int_0^\infty \int_{-\infty}^\infty e^{-\tilde{\omega}/2}f(\tilde{\omega},\tau)\frac{e^{-ik\tilde{\omega}-u\tau}}{\sqrt{2\pi}}d\tilde{\omega}d\tau \tag{A10}$$

We define a generalized source as

$$\mathcal{F}(k,u) = \frac{D}{\mu^2}\nu\tilde{f}(k,u) + \psi_0(k) \tag{A11}$$

For which the solution for $\psi(k,u)$ is then

$$\psi(k,u) = \frac{\mathcal{F}(k,u)}{u + \gamma + k^2} \tag{A12}$$

The time dependent number density $n(\omega,t)$ can now be calculated for a given source $f(\omega,t)$.

1. An impulse response (Green's function)

The Fokker-Planck equation in (A4) is given in a linear form

$$\mathcal{L}\eta(\tilde{\omega}, \tau) = \mathcal{S}(\tilde{\omega}, \tau), \tag{A13}$$

where $\mathcal{L}$ is a linear operator and $\mathcal{S}$ is a source function. We define the Green's function $G(\tilde{\omega}, \tau)$ to be the solution for a point source in both size and time

$$\mathcal{L}G(\tilde{\omega} - \tilde{\omega}_0, \tau - \tau_0) = \delta(\tilde{\omega} - \tilde{\omega}_0)\delta(\tau - \tau_0), \tag{A14}$$

where δ is the Dirac delta function . Using the Green's function, the number density for any general source can be written as

$$\eta(\tilde{\omega}, \tau) = \int \int \mathcal{S}(\tilde{\omega}_0, \tau_0)G(\tilde{\omega} - \tilde{\omega}_0, \tau - \tau_0)\mathrm{d}\tilde{\omega}_0\mathrm{d}\tau_0. \tag{A15}$$

We solve for the Green function using the previous analysis with a source of the form

$$\frac{D}{\mu^2}\nu e^{-\tilde{\omega}/2}f(\tilde{\omega}, \tau) = \delta(\tilde{\omega} - \tilde{\omega}_0)\delta(\tau - \tau_0). \tag{A16}$$

This is a source for funds of size w_0 generating an impulse at $t = t_0$. We will assume that prior to the impulse at $t = 0^-$ there were no funds which means that the initial conditions are $\eta(\tilde{\omega}, 0^-) = 0$ which yields $\psi_0(k) = 0$.

Using (A11) we write

$$\mathcal{F}(k, u) = \frac{1}{\sqrt{2\pi}}\exp\left[-ik\tilde{\omega}_0 - u\tau_0\right] \tag{A17}$$

and $\psi(k, u)$ is given by

$$\psi(k, u) = \frac{1}{\sqrt{2\pi}}\frac{\exp\left[-ik\tilde{\omega}_0 - u\tau_0\right]}{u + \gamma + k^2}. \tag{A18}$$

The inverse Laplace transform yields

$$\phi(k, \tau) = \frac{1}{\sqrt{2\pi}}\exp\left[-(\tau - \tau_0)\left(\gamma + k^2\right) - ik\tilde{\omega}_0\right]\theta(\tau - \tau_0), \tag{A19}$$

where θ is the Heaviside step function. An inverse Fourier transform yields

$$G(\tilde{\omega} - \tilde{\omega}_0, \tau - \tau_0) = \frac{1}{\sqrt{\pi 4(\tau - \tau_0)}}\exp\left[-\frac{(\tilde{\omega} - \tilde{\omega}_0)^2}{4(\tau - \tau_0)} - \gamma(\tau - \tau_0)\right]\theta(\tau - \tau_0). \tag{A20}$$

2. A continuous source of constant size funds

After obtaining the green's function we will look at a continuous source of funds of size w_s starting at a time t_s which can be written as

$$f(\omega, t) = \delta(\omega - \omega_s)\theta(t - t_s). \tag{A21}$$

Rewriting the source using dimensionless variables yields

$$f(\tilde{\omega}, \tau) = \delta(\tilde{\omega} - \tilde{\omega}_s)\theta(\tau - \tau_s) \tag{A22}$$

and using (A15) yields

$$
\begin{aligned}
\eta(\tilde{\omega}, \tau) &= \int \int \mathcal{F}(\tilde{\omega}', \tau')G(\tilde{\omega} - \tilde{\omega}', \tau - \tau')\mathrm{d}\tilde{\omega}'\mathrm{d}\tau' \\
&= \int \int e^{-\tilde{\omega}'/2}\frac{\nu D}{\mu^2}\delta(\tilde{\omega}' - \tilde{\omega}_s)\theta(\tau' - \tau_s)G(\tilde{\omega} - \tilde{\omega}', \tau - \tau')\mathrm{d}\tilde{\omega}'\mathrm{d}\tau' \\
&= \int_{\tau_s}^{\tau} e^{-\tilde{\omega}_s/2}\frac{\nu D}{\mu^2}G(\tilde{\omega} - \tilde{\omega}_s, \tau - \tau')\mathrm{d}\tau' \\
&= e^{-\tilde{\omega}_s/2}\frac{\nu D}{\sqrt{\pi}\mu^2}\int_{\tau_s}^{\tau}\frac{1}{2\sqrt{(\tau - \tau')}}\exp\left[-\frac{(\tilde{\omega} - \tilde{\omega}_s)^2}{4(\tau - \tau')} - \gamma(\tau - \tau')\right]\theta(\tau - \tau')\mathrm{d}\tau'.
\end{aligned}
\tag{A23}
$$

We change variables such that $x = \sqrt{\tau - \tau'}$ for which the integral is rewritten as

$$\eta(\tilde{\omega}, \tau) = e^{-\tilde{\omega}_s/2} \frac{\nu D}{\sqrt{\pi}\mu^2} \int_0^{x_{max}=\sqrt{\tau-\tau_s}} \exp\left\{-\frac{(\tilde{\omega}-\tilde{\omega}_s)^2}{4x^2} - \gamma x^2\right\} dx. \tag{A24}$$

We approximate the integral for large sizes using the method of steepest ascent

$$\int_0^{x_{max}} \exp\left\{-g(x)\right\} dx \approx \int_0^{x_{max}} \exp\left\{-g(x*) - \frac{1}{2}g''(x*)(x-x*)^2\right\} dx$$

$$= \sqrt{\frac{\pi}{2g''(x*)}} \exp\left\{-g(x*)\right\} \left[\mathrm{erf}\left(\sqrt{\frac{g''(x*)}{2}}x*\right) + \mathrm{erf}\left(\sqrt{\frac{g''(x*)}{2}}(x_{max}-x*)\right)\right]$$

$$\approx \sqrt{\frac{\pi}{2g''(x*)}} \exp\left\{-g(x*)\right\} \left[1 + \mathrm{erf}\left(\sqrt{\frac{g''(x*)}{2}}(x_{max})\right)\right]$$

$$\approx \sqrt{\frac{2\pi}{g''(x*)}} \exp\left\{-g(x*)\right\}. \tag{A25}$$

The expansion was done under the assumption that $g''(x*) > 0$ where $x*$ is such that $g'(x*) = 0$. In the last part we made the approximation that $x* \gg 0$ and $x_{max} \gg x*$. The approximation that $x* \gg 0$ is valid for times $\tau \gg \tau_s$. If the value $x*$ is not much larger then 0 then the lower bound in the integration will be replace by x_{min}

We define $\Delta\tilde{\omega} = \tilde{\omega} - \tilde{\omega}_s$ and write

$$g(x) = \frac{\Delta\tilde{\omega}^2}{4x^2} + \gamma x^2. \tag{A26}$$

By demanding that the first derivative vanish we get

$$x* = \left(\frac{\Delta\tilde{\omega}^2}{4\gamma}\right)^{1/4}. \tag{A27}$$

For which

$$g(x*) = \sqrt{\gamma}|\Delta\tilde{\omega}|$$
$$g''(x*) = 8\gamma. \tag{A28}$$

As described above the integral is approximated by

$$\eta(\tilde{\omega}, \tau) = \exp\left(-\frac{\tilde{\omega}_s}{2} - \sqrt{\gamma}|\tilde{\omega}-\tilde{\omega}_s|\right) \frac{\nu D}{2\mu^2\sqrt{\gamma}} \left[\mathrm{erf}\left\{2\sqrt{\gamma}\left(\frac{(\tilde{\omega}-\tilde{\omega}_s)^2}{4\gamma}\right)^{1/4}\right\} + \mathrm{erf}\left\{2\sqrt{\gamma}\left(\sqrt{\tau-\tau_s} - \left(\frac{(\tilde{\omega}-\tilde{\omega}_s)^2}{4\gamma}\right)^{1/4}\right)\right\}\right]. \tag{A29}$$

Finally the number density is given by

$$n(\tilde{\omega}, \tau) = \exp\left(\frac{\tilde{\omega}-\tilde{\omega}_s}{2} - \sqrt{\gamma}|\tilde{\omega}-\tilde{\omega}_s|\right) \frac{\nu D}{2\mu^2\sqrt{\gamma}} \left[\mathrm{erf}\left\{2\sqrt{\gamma}\left(\frac{(\tilde{\omega}-\tilde{\omega}_s)^2}{4\gamma}\right)^{1/4}\right\} + \mathrm{erf}\left\{2\sqrt{\gamma}\left(\sqrt{\tau-\tau_s} - \left(\frac{(\tilde{\omega}-\tilde{\omega}_s)^2}{4\gamma}\right)^{1/4}\right)\right\}\right]. \tag{A30}$$

3. A normal source of funds

The above analysis can be easily applied to a source of the type

$$f(\tilde{\omega}, \tau) = \frac{1}{\sqrt{\pi\sigma_s^2}} \exp\left(-\frac{(\tilde{\omega}-\tilde{\omega}_s)^2}{\sigma_s^2}\right) \theta(\tau - \tau_s), \tag{A31}$$

for which we write

$$\eta(\tilde{\omega}, \tau) = \int\int e^{-\tilde{\omega}'/2} \frac{\nu D}{\mu^2\sqrt{\pi\sigma_s^2}} \exp\left(-\frac{(\tilde{\omega}'-\tilde{\omega}_s)^2}{\sigma_s^2}\right) \theta(\tau' - \tau_s)G(\tilde{\omega} - \tilde{\omega}', \tau - \tau')d\tilde{\omega}'d\tau'. \tag{A32}$$

Substituting the form for the green function from (A20) into the above equation for η yields

$$\eta(\tilde{\omega}, \tau) = \int \int e^{-\tilde{\omega}'/2} \frac{\nu D}{\mu^2 \sqrt{\pi \sigma_s^2}} \exp\left[-\frac{(\tilde{\omega}' - \tilde{\omega}_s)^2}{\sigma_s^2}\right] \frac{1}{\sqrt{\pi 4(\tau - \tau')}} \exp\left[-\frac{(\tilde{\omega} - \tilde{\omega}')^2}{4(\tau - \tau')} - \gamma(\tau - \tau')\right] \theta(\tau - \tau')\theta(\tau' - \tau_s) d\tilde{\omega}' d\tau'.$$
(A33)

Performing the integration on $\tilde{\omega}'$ in (A33) we solve for $\eta(\tilde{\omega}, \tau)$ and the transformation (A2) yields the solution for the number density

$$n(\tilde{\omega}, \tau) = \int_{\tau_s}^{\tau} \frac{\nu D}{\mu^2 \sqrt{\pi (\sigma_s^2 + 4(\tau - \tau'))}} \exp\left[\frac{(\tau - \tau')(\sigma_s^2(1 - 4\gamma) + 8\tilde{\omega}) - 4(4\gamma(\tau - \tau')^2 + (\tilde{\omega} - \tilde{\omega}_s)^2 + 2(\tau - \tau')\tilde{\omega}_s)}{4(\sigma_s^2 + 4(\tau - \tau'))}\right] d\tau'.$$
(A34)

There is no closed form (that we could get) for (A34) and we leave the solution in the integral form.

Appendix B: The Simulation Model

We define three rates:

1. The rate of size change taken to be 1 for each fund and N for the entire population.
   Thus, each fund changes size with a rate taken to be unity.

2. The rate of annihilation of funds of size ω defined as $\lambda(\omega)n(\omega, t)$.
   Each fund is annihilated with a rate λ which can be depend on the fund size.

3. The rate of creation of new funds ν.
   Each new fund is created with a size ω with a probability density $f(\omega)$.

In order to simulate the process we define probabilities for the occurrence of each of the above with probabilities that are defined in such a way that the ratio of any pair is equal to the ratio of the corresponding rates. At every simulation time step, with a probability $\frac{\nu}{1+\lambda+\nu}$ a new fund is created and we proceed to the next simulation time step. If a fund was not created then the following is repeated $(1 + \lambda)N$ times. We pick a fund at random. With a probability $\frac{1}{1+\lambda}$ we change the fund size and with a probability of $\frac{\lambda}{1+\lambda}$ the fund is annihilated. The simulation time can be compared to 'real' time if every time a fund is not created we add $1/(1 + \lambda)$ to the clock. The time is then measured in what ever units our rates are measured in. In our simulation we use monthly rates and as such a unit time step corresponds to one month.

---

[1] L. A. N. Amaral, S. V. Buldyrev, S. Havlin, P. Maass, M. A. Salinger, H. E. Stanley, and M. H. R. Stanley. Scaling behavior in economics: The problem of quantifying company growth.

[2] Robert L. Axtell. Zipf distribution of u.s. firm sizes. *Science*, 293(5536):1818–1820, September 2001.

[3] Nicolas P.B. Bollen. Mutual fund attributes and investor behavior. *Journal of Financial and Quantitative Analysis*, 42(3):683–708, September 2007.

[4] Giulio Bottazzi and Angelo Secchi. Common properties and sectoral specificities in the dynamics of u.s. manufacturing companies. *Review of Industrial Organization*, 23(3-4), December 2003.

[5] Giulio Bottazzi and Angelo Secchi. Explaining the distribution of firms growth rates. LEM Papers Series 2005/16, Laboratory of Economics and Management (LEM), Sant'Anna School of Advanced Studies, Pisa, Italy, July 2005. available at http://ideas.repec.org/p/ssa/lemwps/2005-16.html.

[6] Giulio Bottazzi. Firm diversification and the law of proportionate effect, 2001.

[7] Jeffrey A. Busse. Another look at mutual fund tournaments. *The Journal of Financial and Quantitative Analysis*, 36(1):53–73, mar 2001.

[8] Chevalier, Judith and Ellison, Glenn. Risk taking by mutual funds as a response to incentives. *The Journal of Political Economy*, 105(6):1167–1200, dec 1997.

[9] Aaron Clauset, Cosma Rohilla Shalizi, and M. E. J. Newman. Power-law distributions in empirical data, 2007.

[10] Giancarlo Corsetti, Paolo A. Pesenti, and Nouriel Roubini. The Role of Large Players in Currency Crises. *SSRN eLibrary*, 2001.

[11] Giovanni Dosi. Statistical regularities in the evolution of industries. a guide through some evidence and challenges for the theory. LEM Papers Series 2005/17, Laboratory of Economics and Management (LEM), Sant'Anna School of Advanced Studies, Pisa, Italy, July 2005. available at http://ideas.repec.org/p/ssa/lemwps/2005-17.html.

[12] David S. Evans. The relationship between firm growth, size, and age: Estimates for 100 manufacturing industries. *The Journal of Industrial Economics*, 35(4):567–581, jun 1987.

[13] J Doyne Farmer and Fabrizio Lillo. On the origin of power-law tails in price fluctuations. *Quantitative Finance*, 4(1):C7–C11, February 2004.

[14] De Fabritiis G. On size and growth of business firms. *Physica A*, 324:38–44(7), 1 June 2003.

[15] Xavier Gabaix, Parameswaran Gopikrishnan, Vasiliki Plerou, and H. Eugene. Stanley. A theory of power-law distributions in financial market fluctuations. *Nature (London)* , 423:267–270, May 2003.

[16] Xavier Gabaix, Rita Ramalho, and Jonathan Reuter. Power laws and mutual fund dynamics. Technical report, Massachusetts Institute of Technology, 2003.

[17] Xavier Gabaix, Parameswaran Gopikrishnan, Vasiliki Plerou, and H. Eugene Stanley. Institutional investors and stock market volatility. *The Quarterly Journal of Economics*, 121(2):461–504, May 2006. available at http://ideas.repec.org/a/tpr/qjecon/v121y2006i2p461-504.html.

[18] Xavier Gabaix. Zipf's law for cities: An explanation. *The Quarterly Journal of Economics*, 114(3):739–767, August 1999.

[19] R. Gibrat. Les inegalites economiques, 1931.

[20] Parameswaran Gopikrishnan, Vasiliki Plerou, Amaral Nunes, A. Meyer Martin Luís, and Stanley H. Eugene. Scaling of the distribution of fluctuations of financial market indices. *Phys. Rev. E*, 60(5):5305–5316, Nov 1999.

[21] Parameswaran Gopikrishnan, Vasiliki Plerou, Xavier Gabaix, and Stanley H. Eugene. Statistical properties of share volume traded in financial markets. *Phys. Rev. E*, 62(4):R4493–R4496, Oct 2000.

[22] Guercio, Diane Del and Tkac, Paula A. The determinants of the flow of funds of managed portfolios: Mutual funds vs. pension funds. *The Journal of Financial and Quantitative Analysis*, 37(4):523–557, dec 2002.

[23] Bronwyn H. Hall. The relationship between firm size and firm growth in the us manufacturing sector. *The Journal of Industrial Economics*, 35(4):583–606, jun 1987.

[24] M. S. Kumar. Growth, acquisition activity and firm size: Evidence from the united kingdom. *The Journal of Industrial Economics*, 33(3):327–338, mar 1985.

[25] M. Levy, S. Solomon, and G. Ram. Dynamical Explanation for the Emergence of Power Law in a STOCK Market Model. *International Journal of Modern Physics C*, 7:65–72, 1996.

[26] Francesca Lotti, Enrico Santarelli, and Marco Vivarelli. Does gibrat's law hold among young, small firms? *Journal of Evolutionary Economics*, 13(3):213–235, August 2003. available at http://ideas.repec.org/a/spr/joevec/v13y2003i3p213-235.html.

[27] Burton G. Malkiel. Returns from investing in equity mutual funds 1971 to 1991. *The Journal of Finance*, 50(2):549–572, jun 1995.

[28] Benoit Mandelbrot. The variation of certain speculative prices. *The Journal of Business*, 36(4):394–419, October 1963.

[29] Edwin Mansfield. Entry, gibrat's law, innovation, and the growth of firms. *The American Economic Review*, 52(5):1023–1051, dec 1962.

[30] William J. Reed and Barry D. Hughes. From gene families and genera to incomes and internet file sizes: Why power laws are so common in nature. *Phys. Rev. E*, 66(6):067103, Dec 2002.

[31] Eli M. Remolona, Paul Kleiman, and Debbie Gruenstein. Market returns and mutual fund flows. *Economic Policy Review*, (Jul):33–52, 1997. available at http://ideas.repec.org/a/fip/fednep/y1997ijulp33-52nv.3no.2.html.

[32] Simon, Herbert A. and Bonini, Charles P. The size distribution of business firms. *The American Economic Review*, 48(4):607–617, sep 1958.

[33] H. A. Simon. On a class of skew distribution functions. *Biometrika*, 42(3/4):425–440, December 1955.

[34] Sirri, Erik R. and Tufano, Peter. Costly search and mutual fund flows. *The Journal of Finance*, 53(5):1589–1622, oct 1998.

[35] S. Solomon and P. Richmond. Power laws of wealth, market order volumes and market returns. *Physica A Statistical Mechanics and its Applications*, 299:188–197, October 2001.

[36] M. H. R. Stanley, L. A. N. Amaral, S. V. Buldyrev, S. Havlin, H. Leschhorn, P. Maass, M. A. Salinger, and H. E. Stanley. Scaling behaviour in the growth of companies. *Nature (London)* , 379:804–806, February 1996.

[37] M. H. R. Stanley, S. V. Buldyrev, S. Havlin, R. N. Mantegna, M. A. Salinger, and H. E. Stanley. Zipf plots and the size distribution of firms. *Economics Letters*, 49:453–457(5), October 1995.

[38] John Sutton. Gibrat's legacy. *Journal of Economic Literature*, 35(1):40–59, mar 1997.

[39] Misako Takayasu Takayuki Mizuno and Hideki Takayasu. The mean-field approximation model of company's income growth. *Physica A: Statistical Mechanics and its Applications*, 332:403–411, February 2004.

[40] Xavier Gabaix Parameswaran Gopikrishnan Vasiliki Plerou, H Eugene Stanley. On the origin of power-law fluctuations in stock prices. *Quantitative Finance*, 4(1):C11–C15, February 2004.

[41] Ijiry Y. and Simon H. A. *Skew Distribution and the Size of Business Firms*. 1977.

[42] By using the CRSP Survivor-Bias-Free US Mutual Fund Database we are able to study not only the size distribution but the mechanism of growth.

[43] The size s of a mutual fund is given in millions of US dollars and is corrected for inflation relative to July 2007. The inflation was calculated using the Consumer Price Index published by the BLS.