We investigate the size distribution of equity mutual funds. We show that the heavy tail of the upper size distribution of equity funds is best described by a log-normal as opposed to a power law. The size of an equity fund was taken to be the real dollar value of the Total Assets Managed by the equity fund as given in the CRSP database. Previous analysis of the mutual fund industry 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. Using a stochastic growth model, we argue 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. 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. Using simulations, we show that the modified model does an excellent job of describing the upper tail of the size distribution. As was argued by Herbert Simon, we show by describing the distribution using a random process 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

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 [14, 15]. We are interested in understanding the size distribution of mutual funds for two equally important reasons.

One reason for studying the mutual fund size distribution is part of ongoing attempt to try and relate the distribution of returns in the equity market to the distribution of volume traded in the market through a simple mechanism [14, 16, 21, 31]. It is believed that the relation between these two distributions depends on the distribution of investors wealth. 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.

The second reason for investigating the mutual fund size distribution is that mutual funds are firms belonging to the same industry and as such, using the high quality data available, we are able to investigate the size distribution of firms within a given 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 [28, 36]. 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, 28, 32, 33, 35, 36] 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 [14, 15, 17, 24, 28, 29, 34, 36]. Even though all these processes yield skewed distributions the upper tail of the resulting 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, 28, 32, 33, 36] 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 [14–16] 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. Previous work on the mechanisms of growth of mutual funds include [3, 15]. 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, 18, 27, 30] 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, 19, 20, 22, 25]. The large firms constituting the upper tail of the size distribution are a constant return to scale industry consistent with a Gibrat model for the growth process. We investigate the size dependence of the growth process of mutual fund size and find that it has a similar size dependence to what was observed in firms. 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 [16]. We find this to be in agreement with an efficient market in which we cannot achieve greater return by simply going to larger funds. This observation of efficiency in regard to fund size is complementary to the believed efficiency with respect to performance [23]. 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

---

1 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.
size dependent [1, 4–6, 11, 13, 32, 33]. Not only do we observe a size dependence in the variance of the growth rate but we observe a similar power law dependence as was observed for firms. 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 a guideline, we offer a random growth model that captures the essence of the mechanism responsible for the observed distribution of fund sizes. We define three characteristic processes that govern the evolution; creation of funds, annihilation of funds and size movements of the individual mutual funds. We analyze these processes using the empirical data and model them as stochastic processes. Our modeling approach is that of a top down. We concentrate on the emerging phenomena as it is observed from the empirical data. Our approach is complementary to the more traditional bottom to top approach which builds on investor choice and behavior to explain collective phenomena. Our top down approach is a result of past work [17, 24, 28, 29] where observable attributes of the economy were described using stochastic processes.

We solve the model analytically and 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 than a log-normal. The distribution evolves towards a power law tailed distribution due to the annihilation process of funds. A hand waving argument for the creation of a power law tail from a log normal size distribution due to a random annihilation is as follows; There are relatively more small funds than large funds while the annihilation process results in funds dying with the same probability regardless of their size. As a result, smaller funds die with a higher probability than larger funds. This process results in the fattening of the upper tail. The role of a random annihilation process in creating a steady state power law tailed distribution was studied in a general setting in [26] where it was shown that a power law steady state emerges.

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 large compared to the age of the industry. Thus, our analysis predicts that the distribution will be better described by a log-normal than a power-law, which is verified using the empirical data.

The paper is organized as follows. Section II describes the data used in our empiric investigation.
Using the data set described in Section II we examine the size distribution of mutual funds in Section III. The underlying processes responsible for the size distribution, as they appear in the data, are discussed in Section IV. Using the empirical observation for the dynamical processes Section V lays out our model and the resulting equations that govern the time evolution of the number of funds and their size distribution. Given the model we first examine the time evolution of the number of funds in Section VI. The model is solved and discussed in Section VII. Section VIII presents simulation results of the proposed model and compares them to the empirical data. Section IX presents modification to the model as suggested by the data. The modified model is simulated and compared to the data.

II. DATA SET

Our empirical analysis is carried out using the CRSP Survivor-Bias-Free US Mutual Fund Database. This database 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.

The definition of an equity fund as a fund having a portfolio of at least 80% stock is not cardinal to our work. The results are not qualitatively sensitive to the threshold even if we chose to work with all funds regardless of their portfolio composition. In previous work [14–16] the threshold was chosen to be 95% however we found that this value is too high results in a number of funds which is too small to carry statistically significant tests on the upper tail. Since the threshold did not seem to qualitatively matter we chose a threshold with a large enough number of funds, i.e. 80%.

The data set has monthly values for the Total Assets Managed (TASM) by the fund and the Net Asset Value (NAV). We define the size \( s \) of a fund to be the value of the TASM. 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.
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.

Inset: The upper tail of the CDF for the mutual funds existing in the end of 1998 (dotted line) is compared to an algebraic relation with exponent $-1$ (full line).

III. 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 [14–16]. 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 value of $\zeta_s \approx 1$.

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. In the inset of Figure 1 we reconstruct the graph by plotting the CDF for the year 1998 for funds with sizes $s > 10^2$ (in millions). 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 section we investigate the size distribution and test the hypothesis of a power law tail against the hypothesis that the distribution is actually a lognormal. In subsection
In subsection III B we will test the validity of the power law upper tail hypothesis using statistical tests. In subsection III B we test a log-normal upper tail hypothesis and compare it to the power law hypothesis.

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\(^2\) 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 \(s_{min}\)

\[
p(s) = \frac{\zeta_s}{s_{min}} \left( \frac{s}{s_{min}} \right)^{-(\zeta_s+1)},
\]

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 \(\zeta_s\). This crossover size \(s_{min}\) is chosen such that it minimizes the Kolmogorov-Smirnov (KS) statistic [9]. The KS statistic \(D\) 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 parameter values for \(\zeta_s\) and \(s_{min}\) are summarized in Table I for each of the years. The mean of the yearly values is calculated \(\bar{\zeta}_s = 1.09 \pm 0.04\) and roughly agrees with Zipf’s law for which the exponent is approximately \(\approx 1\). The value of the exponent does not have to be exactly \(\zeta_s = 1\) to be said to follow Zipf’s law but can be any value close to 1 (definitely less than 2). The standard error is calculated by dividing the standard error by the square root of the number of observations (number of years). It is important to note that averaging over the yearly values does not represent a proper estimation for the exponent. We allow ourselves to use it since the exponent estimation is not our main goal and we are arguing that the power law hypothesis is not the best description.

\(^2\) The power law tail hypothesis can be written in a more general form \(p(s) \sim g(s)s^{-(\zeta_s+1)}\), where \(g(s)\) is a slow varying function. However, there are no statistical tests for such a general dependence and as a result we will use the simpler form for which we describe the tests.
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] which has an underlying assumption that the data is independent. Although we did not check rigorously the data for dependence we do not believe it to be a problem. 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 $s \leq s_{\text{min}}$ and for $s > s_{\text{min}}$ we draw from a power law distribution with an exponent that is the best fit to the empirical data given $s_{\text{min}}$. 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 10,000 randomly chosen data sets.

In Figure 3(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 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 2 and so does the number of equity funds in the upper tail \( N_{tail} \) given in Figure 3(b). The number of funds in the upper tail is defined as the number of funds in each of the years with a size \( s \geq s_{\text{min}} \). Figure 3 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, though it cannot be unequivocally rejected. This is the case regardless of the portfolio composition of the fund. The same analysis was carried out for all funds in the CRSP data base in the years 1991 to 2005 and the results are qualitatively the same. 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 upper tail is better described by a log normal distribution. The
TABLE I: Table of yearly parameter values for equity funds defined such that the portfolio contains a fraction of at least 80% stocks. The values for each of the parameters (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.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>$N$</td>
<td>372</td>
<td>1069</td>
<td>1509</td>
<td>2194</td>
<td>2699</td>
<td>3300</td>
<td>4253</td>
<td>4885</td>
<td>5363</td>
<td>5914</td>
<td>6607</td>
<td>7102</td>
<td>7794</td>
<td>8457</td>
<td>8845</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>$\zeta_s$</td>
<td>1.33</td>
<td>1.36</td>
<td>1.19</td>
<td>1.15</td>
<td>1.11</td>
<td>0.78</td>
<td>1.08</td>
<td>1.10</td>
<td>0.95</td>
<td>0.97</td>
<td>1.01</td>
<td>1.07</td>
<td>1.10</td>
<td>1.14</td>
<td>1.09</td>
<td>0.14</td>
<td>0.04</td>
</tr>
<tr>
<td>$s_{\text{min}}$</td>
<td>955</td>
<td>800</td>
<td>695</td>
<td>708</td>
<td>877</td>
<td>182</td>
<td>1494</td>
<td>1945</td>
<td>1147</td>
<td>903</td>
<td>728</td>
<td>836</td>
<td>868</td>
<td>1085</td>
<td>1383</td>
<td>974</td>
<td>408</td>
</tr>
<tr>
<td>$N_{\text{tail}}$</td>
<td>81</td>
<td>129</td>
<td>232</td>
<td>256</td>
<td>280</td>
<td>1067</td>
<td>290</td>
<td>283</td>
<td>557</td>
<td>662</td>
<td>717</td>
<td>494</td>
<td>652</td>
<td>630</td>
<td>550</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>$p$-value</td>
<td>0.58</td>
<td>0.48</td>
<td>0.02</td>
<td>0.45</td>
<td>0.07</td>
<td>0</td>
<td>0.01</td>
<td>0.11</td>
<td>5.10^-4</td>
<td>0.04</td>
<td>0.03</td>
<td>0.07</td>
<td>0.08</td>
<td>0.08</td>
<td>0.15</td>
<td>0.15</td>
<td>0.19</td>
</tr>
<tr>
<td>$\mu(10^{-3})$</td>
<td>26.1</td>
<td>41.8</td>
<td>81.2</td>
<td>46.2</td>
<td>72.7</td>
<td>67.9</td>
<td>57.7</td>
<td>38.8</td>
<td>39.5</td>
<td>19.6</td>
<td>-30</td>
<td>-10</td>
<td>50.2</td>
<td>17.8</td>
<td>30.5</td>
<td>38.5</td>
<td>26.1</td>
</tr>
<tr>
<td>$\sigma$</td>
<td>0.078</td>
<td>0.200</td>
<td>0.260</td>
<td>0.293</td>
<td>0.229</td>
<td>0.282</td>
<td>0.289</td>
<td>0.302</td>
<td>0.283</td>
<td>0.278</td>
<td>0.255</td>
<td>0.242</td>
<td>0.236</td>
<td>0.240</td>
<td>0.246</td>
<td>0.247</td>
<td>0.055</td>
</tr>
<tr>
<td>$R$</td>
<td>-0.50</td>
<td>-1.35</td>
<td>-1.49</td>
<td>-1.71</td>
<td>-3.29</td>
<td>-18.42</td>
<td>-2.25</td>
<td>-1.29</td>
<td>-6.57</td>
<td>-4.96</td>
<td>-2.63</td>
<td>-2.95</td>
<td>-2.00</td>
<td>-1.05</td>
<td>-0.99</td>
<td>-3.43</td>
<td>4.45</td>
</tr>
<tr>
<td>$N_a$</td>
<td>0</td>
<td>41</td>
<td>45</td>
<td>61</td>
<td>139</td>
<td>115</td>
<td>169</td>
<td>269</td>
<td>308</td>
<td>482</td>
<td>427</td>
<td>660</td>
<td>703</td>
<td>675</td>
<td>626</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>$N_c$</td>
<td>185</td>
<td>338</td>
<td>581</td>
<td>783</td>
<td>759</td>
<td>885</td>
<td>1216</td>
<td>1342</td>
<td>1182</td>
<td>1363</td>
<td>1088</td>
<td>1063</td>
<td>1056</td>
<td>796</td>
<td>732</td>
<td>891</td>
<td>346</td>
</tr>
</tbody>
</table>

$N$ - the number of equity funds existing at the end of each year.

$\zeta_s$ - the power law tail exponent (1).

$s_{\text{min}}$ - the lower tail cutoff (in millions of dollars) above which we fit a power law (1).

$N_{\text{tail}}$ - the number of equity funds belonging to the upper tail s.t. $s \geq s_{\text{min}}$.

$p$-value - the probability of obtaining a goodness of fit at least as bad as the one calculated for the empirical data, under the null hypothesis of a power law upper tail.

$\mu$ - the drift term for the geometric random walk (10).

$\sigma$ - the standard deviation of the mean zero Wiener process (10).

$R$ - the base 10 log likelihood ratio of a power law fit relative to a log-normal fit (3).

$N_a$ - the number of annihilated equity funds in each year.

$N_c$ - the number of new (created) equity funds in each year.

Log normal distribution is defined such that the density function $p_{LN}(s)$ obeys

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

and the CDF is given by

$$P(s' > s) = \frac{1}{2} - \frac{1}{2} \text{erf} \left(\frac{\log(s) - \mu_s}{\sqrt{2}\sigma_s}\right).$$
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 4(a) is a log-normal probability plot for the size distribution of funds existing at the end of the year 1998 while Figure 4(b) is a log-normal probability plot for the size distribution of funds existing at the end of 2005. The empirical probabilities are compared to the theoretical log-normal values. For both years, most of the distribution falls on the dashed line corresponding to a log-normal distribution. However, the largest values in the upper tail of the distribution are above the dashed line for both the years. This means that for very large sizes the empirical distribution decays faster than a log-normal. Since a power law distribution decays slower than the log-normal this hints that a log-normal tail hypothesis is better suited.

A visual comparison between the two hypotheses 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 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 5 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
FIG. 5: 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 III A and the fit parameters are $s_{\text{min}} = 1945$ and $\alpha = 1.107$. The empirical data was truncated from below such that only funds with size $s \geq s_{\text{min}}$ 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 done using the maximum likelihood estimation given $s_{\text{min}}$ (2) yielding $\mu = 2.34$ and $\sigma = 2.5$. The value for $s_{\text{min}}$ is taken from the power law fit evaluation.

The year 1998. In Figure 5(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 using the maximum likelihood described in Section III A and the fit parameters are $s_{\text{min}} = 1945$ and $\alpha = 1.107$. The empirical data was truncated from below such that only funds with size $s \geq s_{\text{min}}$ 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 done using the maximum likelihood estimation given $s_{\text{min}}$ (2) yielding $\mu = 2.34$ and $\sigma = 2.5$. The value for $s_{\text{min}}$ is taken from the power law fit evaluation. In both we can conclude that the tail of the empirical distribution decays faster than 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
FIG. 6: A histogram of the base 10 log likelihood ratios $R$ 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.

data is to compare the likelihood of the observation in both hypotheses. We define the likelihood for the tail of the distribution to be

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

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

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

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 is drawn. We define the log likelihood ratio as

$$R = \ln \left( \frac{L_{\text{PL}}}{L_{\text{LN}}} \right).$$
The sign of the log likelihood is determined by the value of the ratio of likelihoods for the two distributions. If $R$ is positive it implies that the power law hypothesis is more probable than a log-normal. A negative value for $R$ 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 III A. Using the fit parameters, the log likelihood ratio was computed (3) and the results are summarized graphically in Figure 6 and in Table I. The histogram clearly shows that for each of the years (all points) the resulting $R$ is negative, implying that the log-normal hypothesis is more probable.

It seems that this property is not unique to equity funds and that the distribution of all mutual funds regardless of their portfolio decomposition are better described by a log-normal. The log likelihood ratio $R$ was computed for all funds in the data base in the years 1991 to 2005 for which all the calculated values were negative implying that a log-normal distribution is better suited to describe the distribution.

To conclude, we have shown that a power law hypothesis for the upper tail of the size distribution is questionable and that a log-normal tail is better suited to describe the distribution. This was shown to hold for each of the years 1991 to 2005 and even though we treat mostly equity funds in this paper this seem to hold for all funds regardless of their composition.

IV. 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. 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. We treat the systems as opaque in the sense that we ignore the inner mechanisms that drive the phenomena we observe, i.e. behavioral aspects.

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 them in 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 by 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. As explained in Section II, we define the size $s$ of a fund (in millions of real US dollars) at time $t$ as $s(t) \equiv TASM(t)$ and the fractional change in the fund size $s$ as

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

The return $\Delta_r(t)$ defined by measuring the fractional change in the Net Asset Value (NAV) of the fund

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

The fractional change corresponding to net flux of money $\Delta_f$ is defined as

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

The fractional change in size due to return $\Delta_r$ and the fractional change due to money flux $\Delta_f$ are related through the relation

$$\Delta_s(t) = \Delta_f(t) + \Delta_r(t). \quad (7)$$

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. The relation between money flux and fund performance was studied in [18, 27, 30] and found positive correlations. These correlations were asymmetric in the sense that response to loss was lower than the response to gains in mutual funds and the opposite was true for pension funds. In [7, 8] the portfolio too was shown to be correlated to incentives and past performance in order to attract (keep) investors.

Another interesting question is whether there is any size dependence in these changes. To help answer this question, the empirical fractional size changes; $\Delta_s$, $\Delta_r$ and $\Delta_f$ were calculated for the year 2005 and are plotted as a function of the fund size $s$ in Figure 7. The return $\Delta_r$ seems to be independent of size, as can be seen in Figure 7. 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 [16]. 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

\[\textsuperscript{3}\text{This flux can be either negative corresponding to investors leaving the fund or positive corresponding to money invested in the fund. We can only measure the net flux of money.}\]
FIG. 7: The total fractional size change $\Delta_s$ (4), the return $\Delta_r$ (5) and the fractional flux of money $\Delta_f$ (6) are plotted as a function of the fund size (in millions). 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.

fund size is complimentary to the believed efficiency with respect to performance [23]. In contrast to the return, the money flux $\Delta_f$ 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.

The size movements are modeled as an iid random process such that a size change $\Delta_s$ at time $t$ is uncorrelated with previous size changes. A size change consists of return and money flux (7) which were shown to be correlated [7, 8, 18, 27, 30]. For an efficient market the return can be viewed as an iid process. For large funds the return is the dominant source of size change and as such we treat the size changes as an iid process.

The distribution of the aggregated monthly logarithmic size changes in the years 1991 to 2005 is given in Figure 8. The monthly size changes were normalized such that the mean monthly logarithmic size change vanishes

$$\sum_i \Delta_w^{(i)}(t) = 0,$$

where the summation is over all size changes $i$ occurring in month $t$. We use this normalization
FIG. 8: The PDF of aggregated monthly log size changes $\Delta s$ for equity funds in the years 1991 to 2005. The log size changes were binned into 20 bins for positive changes and 20 bins for negative changes. Monthly size changes were normalized such that the average log size change in each month is zero.

FIG. 9: The histogram of fund sizes after dispersing for $t = 0, 2, 4$ and 6 years is given in clockwise order starting at the top left corner. The funds at $t = 0$ were all equity funds with a size between 23 and 94 million dollars at the end of 1998. The size distribution after 2, 4 and 6 years was calculated for the surviving funds.

to ensure that when aggregating over different times the distributions will have the same mean.
The distribution exhibits the same right skewed tent shape as was seen in the study of firm growth rates, see for example [2, 11].

The distribution of size changes seems exponential as can be seen in Figure 8. However, one would expect through the central limit theorem that the aggregated affect of these changes over a longer period of time to be normal. We can qualitatively check this assumption of a normal (in log space) change process by looking at the way these funds disperse in size.
The dispersion of funds sizes is investigated by examining the way a relatively narrow (almost delta like) distribution changes with time. To do this we take all equity funds existing at the end of 1998 with a size between 23 to 94 million dollars. The initial sizes were chosen to be centered symmetrically (in log sizes) around the maximum of the size distribution for that year. We then follow these 1111 funds (some will be annihilated over time) and examine their size distribution over time. The resulting size histograms for $t = 0$, which is the initial distribution and for $t = 2, 4, 6$ years are given in Figure 9. It is clear that with time the distribution drifts to larger sizes and it widens into what appears a bell shape (in log space). To check whether indeed the dispersion is log-normal for each of the $t = 0, 2, 4,$ and 6 years the log-normal probability plot for the resulting distributions is given in Figure 10. It can be seen in that after 6 years the diffusion is approximately log-normal.

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 $\mu$ and standard deviation $\sigma$ 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 $\mu$ and $\sigma$ 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
FIG. 11: (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) A log-normal probability plot for the creation size distribution (in millions of dollars). The quantiles on the x-axis are plotted against the corresponding probabilities on the y-axis. The empirical quantiles ($\times$) are compared to the theoretical values for a log-normal distribution (dashed line).

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 $N_c(t)$ and $N(t-1)$. Thus, we approximate the creation of funds as a Poisson process with a constant rate $\nu$. The number of funds as a function of time is plotted in Figure 2 and suggests a linear time dependence for the number of funds $N(t)$.

In Figure 11(a) the PDF for the size $s_c$ of new funds is compared to the size PDF of funds existing at the end of 2005. The density was estimated using the aggregated data for all created funds in the years 1991 to 2005. With the above plot as motivation we approximate the probability of a new fund to be created with a size $\omega$, $f(\omega)$, 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 11(b). The empirical initial fund size distribution differs mostly in the upper and lower tails as can be seen in Figure 11(b) 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. 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.

The third process is the annihilation of funds. It is in fact the annihilation which is responsible for the formation of a power law tail in the steady state distribution. 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.

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 12 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 $\omega$. If we define the number density as $n(\omega, t)$, the rate $\Lambda$ of annihilated funds is given by

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

If we make the simplifying assumption that the rate is independent of size, the rate $\lambda$ is just the slope of the linear regression in Figure 12 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 months per year.

V. 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 [28] and more recently similar model describing the growth of equity funds was proposed by Gabaix et al [15]. There are key differences
FIG. 12: 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 $N(t-1)$. 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.

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 while they define a higher threshold of 95%\(^4\). 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

\(^4\) 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.
motion
\[ ds(t) = s(t) (\mu_s dt + \sigma_s dW_t), \]
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
\[ d\omega(t) = \mu dt + \sigma dW_t, \]
where \( \mu = \mu_s - \sigma_s^2/2 \) and \( \sigma = \sigma_s \).

The Fokker-Planck equation (also known as the forward Kolmogorov equation) for the number density can be written as
\[ \frac{\partial}{\partial t} n(\omega, t) = \left[ -\mu \frac{\partial}{\partial \omega} + \frac{\sigma^2}{2} \frac{\partial^2}{\partial \omega^2} \right] n(\omega, t). \]

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 \( \omega \) is annihilated. The creation process is a Poisson process with rate \( \nu \) such that at time \( t \) a new fund is created with a probability \( \nu dt \). The size of these new funds is distributed with a distribution \( f(\omega, t) \).

The time evolution of the number density can be written as
\[ \frac{\partial}{\partial t} n(\omega, t) = -\lambda(\omega) n(\omega, t) + \nu f(\omega, t) + \left[ -\mu \frac{\partial}{\partial \omega} + \frac{\sigma^2}{2} \frac{\partial^2}{\partial \omega^2} \right] n(\omega, t). \]
Thus, given the creation and annihilation rates \( \lambda \) and \( \nu \), the size distribution of new funds \( f(\omega, t) \) and the initial condition \( n(\omega, 0) \) we can solve for the number density using (12). The total number of funds is defined as
\[ N(t) = \int_{-\infty}^{\infty} n(\omega', t)d\omega'. \]
Once the number density is known it is straightforward to calculate the probability density as
\[ p(\omega, t) = \frac{n(\omega, t)}{N(t)}. \]

VI. TIME DYNAMICS OF THE TOTAL NUMBER OF FUNDS

In this model the number of funds \( N(t) \) obeys the following time evolution
\[ \frac{dN(t)}{dt} = -\int_{-\infty}^{\infty} \lambda(\omega)n(\omega, t)d\omega + \nu(t). \]
For a creation process starting at time $t = 0$ with a constant rate $\nu$ the time dependence of the number of funds obeys

$$N(t) = \frac{\nu}{\lambda} \left(1 - e^{-\lambda t}\right) \theta(t), \quad (16)$$

where $\theta(t)$ is the Heaviside step function. The steady state number of funds is given by

$$\lim_{t \to \infty} N(t) = \frac{\nu}{\lambda}. \quad (17)$$

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 $\nu = 891 \pm 89$ and an annual annihilation rate of $\lambda = 0.092 \pm 0.015$ the steady state number of funds is expected to be $N = 10,000 \pm 2,000$ which is larger than the 8845 equity funds existing at the end of 2005. The value for $\nu$ was calculated as the mean number of funds created in each year as can be seen in Table I. The value for $\lambda$ was calculated in Section IV.

The behavior predicted does not agree with the almost linear dependence we observe empirically in Figure 2. The time scale for the convergence of the number of funds to the steady state value (17) is given by $1/\lambda \approx 11 \pm 2$. This time scale is not large enough to explain this linear growth regime. However, the creation rate $\nu(t)$ is not constant as can be seen in Table I. This is reasonable given the assumption that the number of created funds each year correlate to the market performance in the previous years. We believe that the change in $\nu$ over the years is responsible for the linear increase and indeed by solving for $\nu(t) \sim t$ one gets the observed linear dependence. It is important to note that whether or not $\nu$ is constant or changes with time does not affect the solution for the size distribution.

The time scale for the convergence of the number of funds is different than the time scales governing the dynamics of the size distribution. We will show in the following section that this process is much faster than the one governing the probability distribution and as a result the number of fund will reach a steady state while the probability distribution is still in a transient state.

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

We begin our analysis by simplifying the Fokker-Plank through a change of variables. A convenient method is one in which we transform them into dimensionless variables. To do this we begin by identifying the dimensions of the problem. There are two; the first is log-size and the second is
time. We continue by identifying the dimensions of the parameters of the problem. We concentrate
on the parameters of the diffusion process, a choice that will prove to be useful. The drift term
\( \mu \) has dimensions of log-size/time. The diffusion parameter defined as \( D = \sigma^2/2 \) has dimensions
of log – size\(^2\)/time. Using these two parameters we can construct a time unit \( t_d = D/\mu^2 \) and a
log-size unit \( \omega_0 = D/\mu \).

We define the dimensionless size
\[
\tilde{\omega} = \omega/\omega_0 = (\mu/D)\omega, \tag{18}
\]
the dimensionless time
\[
\tau = t/t_0 = (\mu^2/D)t \tag{19}
\]
and the dimensionless parameter
\[
\gamma = 1/4 + (D/\mu^2)\lambda. \tag{20}
\]

By transforming to the above dimensionless variables the Fokker-Plank equation is written in the
simple form
\[
\left[ \frac{\partial}{\partial \tau} + \gamma - \frac{\partial^2}{\partial \tilde{\omega}^2} \right] \eta(\tilde{\omega}, \tau) = \frac{D}{\mu^2} \nu e^{-\tilde{\omega}/2} f(\tilde{\omega}, \tau). \tag{21}
\]

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

A. An impulse response (Green’s function)

The Fokker-Planck equation in (21) is given in a linear form
\[
\mathcal{L}\eta(\tilde{\omega}, \tau) = \mathcal{S}(\tilde{\omega}, \tau), \tag{22}
\]
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
\[
\mathcal{L}G(\tilde{\omega} - \tilde{\omega}_0, \tau - \tau_0) = \delta(\tilde{\omega} - \tilde{\omega}_0)\delta(\tau - \tau_0), \tag{23}
\]
where \( \delta \) is the Dirac delta function. Using the Green’s function, the number density for any general
source can be written as
\[
\eta(\tilde{\omega}, \tau) = \int \int \mathcal{S}(\tilde{\omega}_0, \tau_0)G(\tilde{\omega} - \tilde{\omega}_0, \tau - \tau_0)d\tilde{\omega}_0d\tau_0. \tag{24}
\]
We solve for the Green function using a source of the form

\[ \frac{D}{\mu^2} \nu e^{-\tilde{\omega}^2/2} f(\tilde{\omega}, \tau) = \delta(\tilde{\omega} - \tilde{\omega}_0) \delta(\tau - \tau_0). \] (25)

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

\[ G(\tilde{\omega} - \tilde{\omega}_0, \tau - \tau_0) = \frac{1}{\sqrt{\pi(\tau - \tau_0)}} \exp\left[ -\frac{(\tilde{\omega} - \tilde{\omega}_0)^2}{4(\tau - \tau_0)} - \frac{\gamma}{4(\tau - \tau_0)} \right] \theta(\tau - \tau_0). \] (26)

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 \( \omega_s \) starting at a time \( t_s \) which can be written as

\[ f(\omega, t) = \delta(\omega - \omega_s) \theta(t - t_s). \] (27)

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

\[ n(\tilde{\omega}, \tau) = \frac{\nu D}{2\mu^2 \sqrt{\gamma}} \exp\left( \frac{\omega_s}{2} - \sqrt{\gamma} |\tilde{\omega} - \tilde{\omega}_s| \right) \exp\left[ -\frac{(\tilde{\omega} - \tilde{\omega}_0)^2}{4(\tau - \tau_0)} - \frac{\gamma}{4(\tau - \tau_0)} \right] \theta(\tau - \tau_0). \] (28)

Since \( \gamma > 1/4 \) (20) the density vanishes for both \( \tilde{\omega} \to \infty \) and \( \tilde{\omega} \to -\infty \).

We can now determine a more quantitative notion of large sizes and large times given the above asymptotic solution (28). We define a large sizes such that the error function becomes constant and we are left with the exponential tail. The error function is approximately constant for large arguments which yields

\[ 2\sqrt{\gamma} \left( \frac{(\tilde{\omega} - \tilde{\omega}_s)^2}{4\gamma} \right)^{1/4} \gg 1, \] (29)

which can be rewritten as

\[ 4\gamma (\tilde{\omega} - \tilde{\omega}_s)^2 \gg 1, \] (30)

yielding

\[ \omega - \omega_s \gg \frac{D}{4\mu \gamma} = \frac{D}{\mu + 4D\lambda/\mu}. \] (31)
Using the average 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 sizes to be

$$s \gg s_s,$$

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 is time independent, i.e. the argument in the error function is large, 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}}.

Using the average monthly rates in Table I and the value for $\lambda = 0.092$ given in Section IV 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|.

Since we use $t$ is units of months, this means that even for a fund of log size (in millions) $\omega = 1$ 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).

Since the steady state 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).

The upper tail distribution is given by

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

Since $s = \exp(\omega)$ the CDF for $s$ has a power law tail with an exponent $\zeta_s$, i.e.

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

Substituting for the parameter $\gamma$ (20) yields

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

26
Using the average parameter values in Table I the asymptotic exponent has the value

\[ \zeta_s = 0.18 \pm 0.04. \]  

(40)

The asymptotic value for the exponent is smaller than 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 blindly attempt to measure the exponent for a power law tail with a finite data set, this would result in a larger value. This is what is observed.

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 \( \omega \). The above analysis can be easily applied to a source of the type

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

(41)

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

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

(42)

There is no closed form (that we could get) for (42) and we leave the solution in the integral form. The number density can be calculated numerically using (42) and indeed the resulting distribution is normal for short time scales and becomes exponential in \( \omega \) (power law in \( s \)) for long times. In Figure 13 the CDF for the fund size \( s \) was calculated numerically using (42) for time horizons of 5, 50 and 100 years. As expected, the size distribution seems to converge towards a power law. Nevertheless, the time scales are quite large and are of the order of 100 years. 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.
FIG. 13: 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.

VIII. 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 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 12. The annihilation rate was taken to be $\lambda = 0.092/12$ per month. The creation rate was taken to be $\nu = 500/12$ per month. As was described in Section VI, the value of $\nu$ affects only the number of funds in the simulation and does not affect the resulting distribution. The drift $\mu = 0.038$ was calculated as the mean of the logarithmic size changes and $\sigma = 0.247$ was calculated as the standard deviation of these logarithmic size changes as give in Table I.

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.

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. The analytical solution for the CDF and the CDF calculated from simulations are in good agreement as can be seen in Appendix B.

**B. Simulation Results**

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.

**IX. A MORE REALISTIC MODEL**

In the construction of the simplifies model we made some crude assumption which we will now refine. The simplified model given is able to 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 which we now integrate into the model.

The first phenomena is that the variation in growth rate $\sigma$ seems to decay with size. A similar phenomena was observed for the processes of firm growth [1, 4, 32, 33]. 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 funds are firms belonging to the same industry. In Section IX A we modify the simplistic model to incorporate the size dependence of the diffusion constant $D$. The second interesting observation is that the average growth rate $\mu$ depends on size. We find that the drift decays with size and becomes constant for large funds. In Section IX B
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 VIII. The second row, Figures (c) and (d) are simulation results of the model with a size dependent diffusion constant (44). The third row, Figures (e) and (f) are simulation results of the model with a size dependent diffusion constant (44) and drift term (46).
FIG. 15: The standard deviation in the logarithmic size change of an equity fund $\sigma$ as a function of the fund's size $s$. The standard deviation was binned into five logarithmically spaced bins with respect to the fund size. 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.

we incorporate the decaying growth rate into our model. The modified models are simulated and compared to the empirical distribution. The resulting distributions (from simulations) for different years are compared to the corresponding empirical distributions with good agreement.

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. 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.

A. Size Dependent Drift Model

A striking result is that the variance in the logarithmic size change is size dependent and can be approximated by a power law dependence of the form

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

(43) 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 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 $\sigma$ 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.

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

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

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 constant standard deviation. This can be seen using the QQ-plot given in Figure 14(d).

The value of the asymptotic exponent to which the distribution converges to is also affected by the size dependence of $\sigma$. Since $\sigma$ decays with size, we approximate the tail exponent by taking $\sigma \to 0$ in (39) which yields,

$$\zeta_s = \frac{\lambda}{\mu}.$$  

For the measured annihilation and drift rates the asymptotic tail exponent is $\zeta_s \approx 0.2 \pm 0.05$. The value for the annihilation rate was taken as $\lambda = 0.092 \pm 0.015$ given in Section IV and for the drift we use $\mu = 0.038 \pm 0.007$ given in Table I.

B. Size Dependent Drift + Diffusion Model

Another aspect of the growth process that we have neglected so far is the dependence of the drift term $\mu$ on the size of the fund as can be seen in Figure 7. We now wish to further enhance
The mean logarithmic size change of an equity fund $\mu$ as a function of the fund's size $s$ (in millions). The data is compared to a power law relation given by (46). The mean logarithmic size change was binned into five logarithmically spaced bins with respect to the fund size. (a) The mean was calculated for the aggregated data for the years 1991 to 2005. (b) The mean was calculated for the aggregated data for the years 1991 to 1998.

our model to take this size dependence into account. As a first approximation we fit a power law relation between the drift $\mu$ and the size $s$ (exponential in the log size $\omega$) of the form

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

(46)

This relation implies that the growth rate decays with the size of the fund. Smaller funds grow in relative terms faster than larger funds. For very large sizes the growth rate becomes approximately constant with a rate $\mu_\infty$. Similar to what we have done previously, we compute the mean of the log size change and check for a size dependence in the drift term $\mu$. 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. Fitting (46) for the data, aggregated over different years, we get slightly different values for the fit parameters given in Table II.

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.


\[
\begin{array}{|c|c|c|c|c|}
\hline
\hline
\sigma_0 & 0.21 \pm 0.07 & 0.26 \pm 0.06 & 0.32 \pm 0.08 & 0.35 \pm 0.07 \\
\beta & 0.29 \pm 0.03 & 0.25 \pm 0.03 & 0.21 \pm 0.03 & 0.21 \pm 0.03 \\
\mu_0 & 0.38 \pm 0.14 & 0.33 \pm 0.11 & 0.22 \pm 0.04 & 0.19 \pm 0.05 \\
\alpha & 0.69 \pm 0.26 & 0.7 \pm 0.24 & 0.76 \pm 0.14 & 0.74 \pm 0.21 \\
\mu_\infty & 0.016 \pm 0.01 & 0.014 \pm 0.008 & 0.0015 \pm 0.0025 & 0.005 \pm 0.004 \\
\hline
\end{array}
\]

TABLE II: \( \sigma_0 \) and \( \alpha \) - The fitted parameter values for the size dependence of the variance of the growth process (43).
\( \mu_0, \alpha \) and \( \mu_\infty \) - 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.

To verify that the model captures the evolution of the size distribution the parameters of the size dependent drift + diffusion model were calculated from the aggregated data for several periods. Using the power law relations for the decaying variance in growth rates (43) 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 the size dependent drift + diffusion model 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 captures 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.

The value of the asymptotic exponent calculated in the previous section is affected by the size dependence of \( \mu \). Since \( \mu \) decays with size towards a constant value \( \mu_\infty \), we approximate the tail exponent by taking \( \mu \rightarrow \mu_\infty \) in (45) which yields,

\[
\zeta_s = \frac{\lambda}{\mu_\infty}.
\]

Using the value for \( \mu_\infty = 0.01 \pm 0.14 \) in Table II, corresponding to aggregation over the years 1991-1998, the steady state power law exponent is \( \zeta_s = 0.8 \pm 0.4 \). This predicted asymptotic value is significantly larger than the values predicted for the simplified model and the size dependent drift model. This value predicts the the distribution will evolve towards a steady state in which
FIG. 17: The size dependent drift + diffusion model 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 the model (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.

the tail will follow Zipf’s law. Nevertheless, it will take a long time for the distribution to reach it’s steady state.

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 $\mu$ and the variance of the growth process $\sigma$. 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 than 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.

X. 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 than a power law as our model predicts. In contrast with past work on 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. This allowed us to 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.

We were able to describe the size distribution in a rather unorthodox manner, not by investigating investor choice but rather by building a Zero Intelligence (ZI) model. This ZI model relies on a surprising result that selection doesn’t seem to play a major role and that investor behavior is not rational. When looking at flux of money in/out of a fund with respect to trailing return we found very weak correlations contrary to what we would expect from a rational and competitive market. Selection seems to be a very weak and noisy process.

The model we offer for explaining the size distribution can be extended to the problem of specie allometry in which one investigates the physical size distribution. Such a physical size distribution will portray the same characteristics but with a cutoff corresponding to physical limitations on possible sizes.

The question of whether a size limit exists is important to scholars and practitioners alike. Such a size limit might be the result of transaction costs that will induce a cutoff on profitable fund sizes. The slight deviation in the upper tail from a log-normal might be an indication that funds are approaching such a size limitation. This deviation is too small to indicate whether such a limit exists or not. Whether this cutoff exists and whether it is a hard or a soft cutoff is an open question.

The results and model have implications also to the field of industrial economics. If we consider mutual funds as firms belonging to the same industry this allows us to examine an important
question in industrial organization; the distribution of firm sizes. There too the shape of the
distribution within an industry is a subject of ongoing debate. Our work has revealed many
similarities between the growth processes of mutual funds and that what was found for firms, i.e.
the size dependent mean and standard deviation of growth rates. Applying the methodology of
this paper to the study of firms can help shed light on the time scales of the problem and hopefully
provide insight on whether the power law of firm size distribution is the consequence of aggregation
or is it an inherent property of the process.
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

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

By defining the function

\[
\eta(\tilde{\omega}, \tau) = e^{-\tilde{\omega}/2} n(\tilde{\omega}, \tau)
\]

The Fokker-Plank equation (A1) is rewritten as

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

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

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

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

\[
\phi(\tilde{\omega}, u) = \int_{0}^{\infty} \eta(\tilde{\omega}, \tau) e^{-u\tau} d\tau \quad (A5)
\]

and the Fourier transformed functions

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

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

For which (A4) is written as

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

with the source distribution transformed as follows

\[
\tilde{f}(k, u) = \int_{0}^{\infty} \int_{-\infty}^{\infty} e^{-\tilde{\omega}/2} f(\tilde{\omega}, \tau) \frac{e^{-ik\tilde{\omega}-u\tau}}{\sqrt{2\pi}} d\tilde{\omega} d\tau \quad (A10)
\]

We define a generalized source as

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

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

\[
\psi(k, u) = \frac{\mathcal{F}(k, u)}{u + \gamma + k^2} \quad (A12)
\]

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

\[ \mathcal{L}\eta(\tilde{\omega}, \tau) = S(\tilde{\omega}, \tau), \]  

(A13)

where \( \mathcal{L} \) is a linear operator and \( S \) 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

\[ \mathcal{L}G(\tilde{\omega} - \tilde{\omega}_0, \tau - \tau_0) = \delta(\tilde{\omega} - \tilde{\omega}_0)\delta(\tau - \tau_0), \]  

(A14)

where \( \delta \) is the Dirac delta function. Using the Green’s function, the number density for any general source can be written as

\[ \eta(\tilde{\omega}, \tau) = \int \int S(\tilde{\omega}_0, \tau_0)G(\tilde{\omega} - \tilde{\omega}_0, \tau - \tau_0)d\tilde{\omega}_0d\tau_0. \]  

(A15)

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

\[ \frac{D}{\mu^2}e^{-\tilde{\omega}/2}f(\tilde{\omega}, \tau) = \delta(\tilde{\omega} - \tilde{\omega}_0)\delta(\tau - \tau_0). \]  

(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

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

(A17)

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

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

(A18)

The inverse Laplace transform yields

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

(A19)

where \( \theta \) is the Heaviside step function. An inverse Fourier transform yields

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

(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 \( \omega_s \) starting at a time \( t_s \) which can be written as

\[
f(\omega, t) = \delta(\omega - \omega_s)\theta(t - t_s).
\] (A21)

Rewriting the source using dimensionless variables yields

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

and using (A15) yields

\[
\eta(\tilde{\omega}, \tau) = \int \int \mathcal{F}(\omega', \tau')G(\tilde{\omega} - \omega', \tau - \tau')d\omega'd\tau'
\]

\[= \int \int e^{-\omega'/2} \frac{\nu D}{\mu^2} \delta(\omega' - \omega_s)\theta(\tau' - \tau_s)G(\tilde{\omega} - \omega', \tau - \tau')d\omega'd\tau'
\]

\[= \int_{\tau_s}^{\tau} e^{-\tilde{\omega}_s/2} \frac{\nu D}{\mu^2} G(\tilde{\omega} - \tilde{\omega}_s, \tau - \tau')d\tau'
\]

\[= e^{-\tilde{\omega}_s/2} \frac{\nu D}{\sqrt{\pi} \mu^2} \int_{\tau_s}^{\tau} \frac{1}{2\sqrt{(\tau - \tau')}} \exp \left[ -\frac{(\tilde{\omega} - \tilde{\omega}_s)^2}{4(\tau - \tau')} \right] \theta(\tau - \tau')d\tau'.
\] (A23)

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

\[
\eta(\tilde{\omega}, \tau) = e^{-\tilde{\omega}_s/2} \frac{\nu D}{\sqrt{\pi} \mu^2} \int_{0}^{\frac{\tau}{\tau_s}} \exp \left\{ -\frac{(\tilde{\omega} - \tilde{\omega}_s)^2}{4x^2} - \gamma x^2 \right\} dx.
\] (A24)

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

\[
\int_{0}^{x_{max}} \exp \{-g(x)\} \, dx \approx \int_{0}^{x_{max}} \exp \left\{ -g(x^*) - \frac{1}{2} g''(x^*)(x^* - x^*)^2 \right\} \, dx
\]

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

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

\[\approx \sqrt{\frac{2\pi}{g''(x^*)}} \exp \{-g(x^*)\}.
\] (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

\[
g(x) = \frac{\Delta \tilde{\omega}^2}{4x^2} + \gamma x^2.
\] (A26)
By demanding that the first derivative vanish we get
\[ x* = \left( \frac{\Delta \tilde{\omega}^2}{4\gamma} \right)^{1/4}. \] (A27)

For which
\[ g(x*) = \sqrt{\gamma} |\Delta \tilde{\omega}| \]
\[ g''(x*) = 8\gamma. \] (A28)

As described above the integral is approximated by
\[ \eta(\tilde{\omega}, \tau) = \frac{\nu D}{2\mu^2 \sqrt{\gamma}} \exp \left( -\frac{\tilde{\omega}_s}{2} - \sqrt{\gamma}|\tilde{\omega} - \tilde{\omega}_s| \right) \]
\[ \times \left[ \text{erf} \left( 2\sqrt{\gamma} \left( \frac{(\tilde{\omega} - \tilde{\omega}_s)^2}{4\gamma} \right)^{1/4} \right) + \text{erf} \left( 2\sqrt{\gamma} \left( \frac{(\tilde{\omega} - \tilde{\omega}_s)^2}{4\gamma} \right)^{1/4} \right) \right]. \] (A29)

Finally the number density is given by
\[ n(\tilde{\omega}, \tau) = \frac{\nu D}{2\mu^2 \sqrt{\gamma}} \exp \left( \frac{(\tilde{\omega} - \tilde{\omega}_s)^2}{2} - \sqrt{\gamma}|\tilde{\omega} - \tilde{\omega}_s| \right) \]
\[ \times \left[ \text{erf} \left( 2\sqrt{\gamma} \left( \frac{(\tilde{\omega} - \tilde{\omega}_s)^2}{4\gamma} \right)^{1/4} \right) + \text{erf} \left( 2\sqrt{\gamma} \left( \frac{(\tilde{\omega} - \tilde{\omega}_s)^2}{4\gamma} \right)^{1/4} \right) \right]. \] (A30)

3. A normal source of funds

The above analysis can be easily applied to a source of the type
\[ f(\tilde{\omega}, \tau) = \frac{1}{\sqrt{\pi\sigma_s^2}} \exp \left( -\frac{(\tilde{\omega} - \tilde{\omega}_s)^2}{\sigma_s^2} \right) \theta(\tau - \tau_s), \] (A31)
for which we write
\[ \eta(\tilde{\omega}, \tau) = \int \int e^{-\tilde{\omega}'^2/2} \frac{\nu D}{\mu^2 \sqrt{\pi\sigma_s^2}} \exp \left( -\frac{(\tilde{\omega}' - \tilde{\omega}_s)^2}{\sigma_s^2} \right) \theta(\tau' - \tau_s)G(\tilde{\omega} - \tilde{\omega}', \tau - \tau')d\tilde{\omega}'d\tau'. \] (A32)

Substituting the form for the green function from (A20) into the above equation for \( \eta \) yields
\[ \eta(\tilde{\omega}, \tau) = \int \int e^{-\tilde{\omega}'^2/2} \frac{\nu D}{\mu^2 \pi\sigma_s \sqrt{4(\tau - \tau')}} \]
\[ \times \exp \left[ -\frac{(\tilde{\omega}' - \tilde{\omega}_s)^2}{\sigma_s^2} - \frac{(\tilde{\omega} - \tilde{\omega}_s)^2}{4(\tau - \tau')} - \gamma(\tau - \tau') \right] \theta(\tau - \tau')\theta(\tau' - \tau_s)d\tilde{\omega}'d\tau'. \] (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
\[ n(\tilde{\omega}, \tau) = \int_{\tau_s}^{\tau} \frac{\nu D}{\mu^2 \sqrt{\pi(\sigma_s^2 + 4(\tau - \tau'))}} \]
\[ \times \exp \left[ \frac{(\tau - \tau')(\sigma_s^2(1 - 4\gamma) + 8\tilde{\omega}) - 4(4\gamma(\tau - \tau')^2 + (\tilde{\omega} - \tilde{\omega}_s)^2 + 2(\tau - \tau')\tilde{\omega}_s)}{4(\sigma_s^2 + 4(\tau - \tau'))} \right] d\tau'. \] (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 $\omega$ defined as $\lambda(\omega)n(\omega, t)$.

   Each fund is annihilated with a rate $\lambda$ which can be depend on the fund size.

3. The rate of creation of new funds $\nu$.

   Each new fund is created with a size $\omega$ 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.

In Figure 18 the CDF obtained from simulation is compared to the numerical calculation using (42) calculated for the same time horizon. The comparison is given for varying time horizons of 1,5,10,20 and 50 years.

---


FIG. 18: The cumulative size density $P(s > x)$ from simulations (○) 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.


