What determines mutual fund size?

August 17, 2009

Abstract

The mutual fund industry manages about a quarter of the assets in the U.S. stock market and thus plays an important role in the U.S. economy. The question of how much control is concentrated in the hands of the largest players can be quantitatively discussed in terms of the tail behavior of the mutual fund size distribution. We study the distribution empirically and show that the tail is better described by a log-normal than a power law, indicating less concentration than, for example, personal wealth. We postulate that the reasons for this stem from market efficiency and the stochastic nature of fund returns; they otherwise have very little to do with investor choice. To demonstrate this we study mutual fund entry, exit and growth empirically and develop a stochastic model. Under simplifying assumptions we obtain a time-dependent analytic solution. The distribution evolves from a log-normal into a power law only over long time scales, suggesting that log-normality comes about because the industry is still young. Numerical solutions under more realistic conditions support this conclusion and give good agreement with the data. Our study shows that, while mutual funds behave in many respects like other firms, there some respects in which they are quite unusual. Surprisingly, it appears that transaction costs and investor choice play only a minor role in determining fund size.
Introduction

In the past decade the mutual fund industry has grown rapidly, moving from 3% of taxable household financial assets in 1980, to 8% in 1990, to 23% in 2007\(^1\). In absolute terms, in 2007 this corresponded to 4.4 trillion USD and 24% of U.S. corporate equity holdings. Mutual funds account for a significant fraction of trading volume in financial markets and have a substantial influence on prices. Moreover, large players such as institutional investors are known to play an important role in the market [Corsetti et al., 2001] and it was recently hypothesized that the fund size distribution is of fundamental importance in explaining both the heavy tailed trading volume distribution and the heavy tailed price return distribution observed across different assets\(^2\). This raises the question of who has this influence: Are mutual fund investments concentrated in a few dominant large funds, or spread across many funds of similar size? What are the economic mechanisms that determine this?

This question can be addressed quantitatively in terms of the upper tails of the mutual fund size distribution. The two competing hypotheses usually made in studies of firms in general are a Zipf (power law) distribution, which is so heavy tailed that none of its moments exist, vs. a log-normal distribution, for which all the moments exist. For mutual funds we resolve this empirically using the CRSP data set and find that the equity fund size distribution is much better described by a log-normal distribution. This implies less dominance by the largest funds.

This naturally leads to the question of why this is true. One would naively think (as we did) that investor choice would be a key determinant of fund size. At the level of individual funds this is almost certainly true. Investors make their choice based on factors such as past performance, advertising, fees, and investment fads, that are potentially fertile ground for behavioral economics. This is surely important in determining the size of individual funds. Our results, in contrast, suggest that for the distribution of fund sizes the details of individual choice are not important.

A hypothesis put forth by Berk and Green [2004] is that the sizes of mutual funds are determined by a combination of the distribution of skill


\(^2\)The equity fund size distribution was argued to be responsible for the observed distribution of trading volume [Levy et al., 1996; Solomon and Richmond, 2001], and Gabaix et al. have argued that it is important for explaining the distribution of price returns [Gabaix et al., 2003; Gabaix et al., 2006].
of mutual fund managers and transaction costs. All else being equal, larger funds must make bigger transactions. Due to market impact, bigger transactions incur larger costs per share, and at sufficiently large size this should diminish a fund’s performance and make it less able to attract funds. Under the assumption of market efficiency, a fund with superior performance before transaction costs will grow until its after transaction cost returns are equivalent to those of other funds. This hypothesis is compelling and plausible, but as we discuss in more detail later, the empirical evidence suggests that it is not consistent with the data.

We offer the competing hypothesis that fund size is a diffusive process, and can be well-understood by treating changes in fund size as random. We are working in the tradition of stochastic models for firm size initiated by Gibrat, Simon and Mandelbrot, and elaborating on earlier work by Gabaix et al.3 Our treatment is new in that we find a time-dependent solution. We argue that this is essential for understanding the empirical data. In a sense that we make precise, the diffusion of fund size is slow, and gets even slower as a fund gets bigger. As we show, since funds are entering and exiting, under stationary conditions the distribution will eventually become a power law, but this takes a very long time. This is exacerbated by the fact that the diffusion slows down as a fund gets larger. As a result the typical time to approach the asymptotic heavy tailed power law distribution is the order of a century. In the meantime the transient distribution has an essentially log-normal upper tail.

Our theory suggests that fund size is a robust property for which investor choice plays at most a minor role. A critic might say that this implies that the whole phenomenon of fund size has “no economic content”. We view this as a positive attribute of our theory: It implies that its conclusions are very robust, depending neither on rationality, nor on the idiosyncrasies of human behavior.

Our results are interesting in the broader context of the literature on firm size. Mutual funds provide a particularly good type of firm to study because there are a large number of funds and their size is accurately recorded. It is generally believed that the resulting size distribution from aggregating across industries has a power law tail that roughly follows Zipf’s law, but

3A similar model yielding a steady state power law distribution, was originally proposed by Gabaix et al. [Gabaix et al., 2003]. Their steady state treatment lacks both an empirical verification of the power law hypothesis and a justification for the \( t \to \infty \) limit taken while solving the model. A similar argument for the power law steady state in a simple entry and exit process is described in [Mitzenmacher, 2004]. Neither of these solved the more general time dependent case, as we have here.
for individual industries the tail behavior is debated\textsuperscript{4}. A large number of stochastic process models have been proposed to explain this\textsuperscript{5}. Our empirical results show that in many respects mutual funds behave like typical firms. In particular, they show the same power law decay in the diffusion rate vs. size that has been observed for other types of firms\textsuperscript{6}. However, mutual funds are distinctive in that, unlike other types of firms that have been studied, in the large size limit the mean and standard deviation of the growth rate do not decay to zero, but rather approach a positive limit. This has a simple economic interpretation: In the large size limit, a fund continues to grow in size simply due to its own performance, even without further inflows of capital\textsuperscript{7}. As our analytic treatment of this problem shows, this makes a crucial difference in the asymptotic behavior. It implies that after the passage of a sufficient length of time, the size distribution for mutual funds should approach a power law, whereas other firms should approach a stretched exponential, which has relatively thin tails (thinner than those of a log-normal). Thus for most firms the tails should get thinner with time, but for mutual funds they should become fatter. This does not have much effect on the current empirical distribution, however, since because of the slow diffusion mentioned earlier, over timescales less than a century both mutual funds and other firms should remain roughly log-normal.

The paper is organized as follows. Section 1 describes the data used for the empirical study described in Section 2. The underlying dynamical processes responsible for the size distribution are discussed in Section 3 and are used to develop the model discussed in Section 4. Section 4.1 presents the solution for the number of funds and Section 4.2 presents the solution for the size distribution. In Section 4.3 we present simulation results of the proposed model and compare them to the empirical data. In section 5 we discuss the incompatibility of the Berk and Green [2004] hypothesis

\textsuperscript{4}Some studies have found that the upper tail is a log-normal [Simon, Herbert A. and Bonini, Charles P., 1958; Stanley et al., 1995; Y. and A., 1977; Stanley et al., 1996; Amaral et al., 1997a; Bottazzi and Secchi, 2003a; Dosi, 2005] while others have found a power law [Axtell, 2001; Bottazzi and Secchi, 2003a; Dosi, 2005]

\textsuperscript{5}For past stochastic models see [Gibrat, 1931; Simon, 1955; Simon, Herbert A. and Bonini, Charles P., 1958; Mandelbrot, 1963; Y. and A., 1977; Sutton, 1997; Gabaix et al., 2003; Gabaix et al., 2003]

\textsuperscript{6}For the main body of work on the size dependence of firm growth rate fluctuations see [Stanley et al., 1995, 1996; Amaral et al., 1997a; Bottazzi, 2001a; Bottazzi and Secchi, 2003a, 2005a; Dosi, 2005; De Fabritiis et al., 2003a]

\textsuperscript{7}Under the Berk and Green theory [2004], this should be irrational – investors should remove their funds and reallocate them to smaller funds. Thus in this sense there is some behavioral dependence on the solution, though since the convergence is so slow, it makes no difference in the current mutual fund distribution. See the discussion in Section 5.
with the empirical data and our hypothesis. Finally Section 6 presents our conclusions.

1 Data Set

We analyze the CRSP Survivor-Bias-Free US Mutual Fund Database. Because we have daily data for each mutual fund, this database enables us not only to study the distribution of mutual fund sizes in each year but also to investigate the mechanism of growth. We study the data from 1991 to 2005\(^8\). We define an equity fund as one whose portfolio consists of at least 80% stocks. The results are not qualitatively sensitive to this, e.g. we get essentially the same results even if we use all funds. 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, measured in millions of US dollars and corrected for inflation relative to July 2007. Inflation adjustments are based on the Consumer Price Index, published by the BLS.

2 The observed distribution of mutual fund sizes

Despite the fact that the mutual fund industry offers a large quantity of well-recorded data, the size distribution of mutual funds has not been rigorously studied. This is in contrast with other types of firms where the size distribution has long been an active research subject. The fact that the distribution is highly skewed and heavy tailed can be seen in Figure 1, where we plot the cumulative distribution of sizes \(P(s > X)\) of mutual fund sizes in three different years.

The exact functional form for the distribution of fund size is still debated. The two principal competing hypotheses are that the tail of the size distribution \(p(s)\) is either log-normal or a power law. Log-normality means that \(\log s\) has a normal distribution, while power law means that the cumulative distribution \(P\) for large \(s\) is of the form\(^9\)

\[
P(s > X) \sim X^{-\zeta},
\]

\(^8\)There is data on mutual funds starting in 1961, but prior to 1991 there are very few entries. There is a sharp increase in 1991, suggesting incomplete data collection prior to 1991.

\(^9\)We denote scaling by \(f(x) \sim g(x)\) which means that the two terms are proportional for large \(x\), i.e. \(f(x) \propto g(x)\) for large \(x\).
Figure 1: The CDF for the mutual fund size $s$ (in millions of 2007 dollars) is plotted with a double logarithmic scale. 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 at the end of 1998 (dotted line) is compared to an algebraic relation with exponent $-1$ (solid line).
where $\zeta_s > 0$. In the special case $\zeta_s \approx 1$ a distribution $P$ is said to obey Zipf’s law. Power laws have the property that moments of higher order than $\alpha$ do not exist, so if a distribution follows Zipf’s law its right skewness is so extreme that the mean is on the boundary where it becomes infinite and there is no such thing as an “average fund size”. In contrast, for a log-normal all the moments exist. From the point of view of extreme value theory this distinction is critical, since it implies a completely different class of tail behavior\(^{10}\).

A visual inspection of the mutual fund size distribution suggests that it does not follow Zipf’s law\(^ {11}\). To see this, in the inset of Figure 1 we compare the tail for funds with sizes $s > 10^2$ million to a power law $s^{-\zeta_s}$, with $\zeta_s = -1$. Whereas a power law corresponds to a straight line when plotted on double logarithmic scale, the data show substantial and consistent downward curvature. In the remainder of this section we make more rigorous tests that back up the intuitive impression given by this plot, indicating that the data are not well described by a power law.

### 2.1 Is the tail a power law?

To test the validity of the power law hypothesis we use the method developed by Clauset et al. [2007]. They use the somewhat strict definition\(^ {12}\) that the probability density function $p(s)$ is a power law if there exists an $s_{\text{min}}$ such that for sizes larger than $s_{\text{min}}$, the functional form of the density $p(s)$ can be written

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

(1)

where the distribution is normalized in the interval $[s_{\text{min}}, \infty)$. There are two free parameters $s_{\text{min}}$ and $\zeta_s$. This crossover size $s_{\text{min}}$ is chosen such that it

---

\(^{10}\)According to extreme value theory a probability distribution can have only four possible types of tail behavior. The first three correspond to distributions with finite support, thin tails, and tails that are sufficiently heavy that some of the moments do not exist, i.e. power laws. The fourth category corresponds to distributions that in a certain sense do not converge; it is remarkable that most known distributions fall into one of the first three categories [Embrechts et al., 1997].

\(^{11}\)Previous work on the size distribution of mutual funds by Gabaix et al. [Gabaix et al., 2003; Gabaix et al., 2003; Gabaix et al., 2006] argued for a power law while we argue here for a log-normal.

\(^{12}\)In extreme value theory a power law is defined as any function that in the limit $s \to \infty$ can be written $p(s) = g(s) s^{-\left(\zeta_s + 1\right)}$ where $g(s)$ is a slowly varying function. This means it satisfies $\lim_{s \to \infty} g(t s)/g(s) = C$ for any $t > 0$, where $C$ is a positive constant. The test for power laws in reference [Clauset et al., 2007] is too strong in the sense that it assumes that there exists an $s_0$ such that for $s > s_0$, $g(s)$ is constant.
minimizes the Kolmogorov-Smirnov (KS) statistic $D$, which is the distance between the CDF of the empirical data $P_e(s)$ and that of the fitted model $P_f(s)$, i.e.

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

Using this procedure we estimate $\zeta_s$ and $s_{\text{min}}$ for the years 1991-2005 as shown in Table I. The values of $\zeta_s$ computed in each year range from 0.78 to 1.36 and average $\bar{\zeta}_s = 1.09 \pm 0.04$. If indeed these are power laws this is consistent with Zipf’s law. But of course, merely computing an exponent and getting a low value does not mean that the distribution is actually a power law.

To test the power law hypothesis more rigorously we follow the Monte Carlo method utilized by Clauset et al. Assuming independence, for each year we generate 10,000 synthetic data sets, each drawn from a power law with the empirically measured values of $s_{\text{min}}$ and $\zeta_s$. For each data-set we calculate the KS statistic to its best fit. The $p$-value is the fraction of the data sets for which the KS statistic to its own best fit is larger than the KS statistic for the empirical data and its best fit.

The results are summarized in Table I. The power law hypothesis is rejected with two standard deviations or more in six of the years and rejected at one standard deviation or more in twelve of the years (there are fifteen in total). Furthermore there is a general pattern that as time progresses the rejection of the hypothesis becomes stronger. We suspect that this is because of the increase in the number of equity funds. As can be seen in Table I, the total number of equity funds increases roughly linearly in time, and the number in the upper tail $N_{\text{tail}}$ also increases.

We conclude that the power law tail hypothesis is questionable but cannot be unequivocally rejected in every year. Stronger evidence against it comes from comparison to a log-normal, as done in the next section.

### 2.2 Is the tail log-normal?

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

$$P(s' > s) = \frac{1}{2} - \frac{1}{2} \text{erf} \left( \frac{\log(s) - \mu_s}{\sqrt{2}\sigma_s} \right).$$
variable & 91 & 92 & 93 & 94 & 95 & 96 & 97 & 98 & 99 & 00 & 01 & 02 & 03 & 04 & 05 & mean & std \\
$R$ & -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 \\
$E[s]$ (mn) & 810 & 385 & 480 & 398 & 448 & 527 & 559 & 619 & 748 & 635 & 481 & 335 & 425 & 458 & 474 & 519 & 134 \\
$Std[s]$ (bn) & 1.98 & 0.99 & 1.7 & 1.66 & 1.68 & 2.41 & 2.82 & 3.38 & 4.05 & 3.37 & 2.69 & 1.87 & 2.45 & 2.64 & 2.65 & 2.42 & 0.8 \\
$E[\omega]$ & 5.58 & 4.40 & 4.40 & 3.86 & 3.86 & 3.91 & 3.84 & 3.85 & 4.06 & 3.97 & 3.60 & 3.37 & 3.55 & 3.51 & 3.59 & 3.60 & 0.54 \\
$Std[\omega]$ & 1.51 & 1.98 & 2.09 & 2.43 & 2.50 & 2.50 & 2.51 & 2.46 & 2.45 & 2.63 & 2.42 & 2.49 & 2.59 & 2.50 & 2.34 & 0.29 \\
$\zeta$ & 1.33 & 1.36 & 1.39 & 1.15 & 1.11 & 0.78 & 1.08 & 1.10 & 0.95 & 0.97 & 1.01 & 1.07 & 1.10 & 1.14 & 1.09 & 0.14 \\
$s_{\text{min}}$ & 955 & 800 & 695 & 708 & 877 & 182 & 1494 & 1945 & 1147 & 903 & 728 & 836 & 868 & 1085 & 1383 & 974 & 408 \\
p-value & 0.58 & 0.48 & 0.02 & 0.45 & 0.07 & 0.01 & 0.11 & 5\times 10^{-4} & 0.04 & 0.03 & 0.07 & 0.08 & 0.08 & 0.15 & 0.15 & 0.19 \\
$\mu(10^{-4})$ & 26 & 42 & 81 & 46 & 72 & 67 & 58 & 39 & 39 & 20 & -3 & -10 & 50 & 18 & 30.5 & 38.5 & 26 \\
$\sigma(10^{-4})$ & 0.78 & 2.0 & 2.6 & 3.0 & 2.3 & 2.8 & 2.9 & 3.0 & 2.8 & 2.8 & 2.5 & 2.4 & 2.4 & 2.4 & 2.5 & 2.5 & 0.55 \\
$N_{\text{exit}}$ & 0 & 41 & 45 & 61 & 139 & 115 & 169 & 269 & 308 & 482 & 427 & 660 & 703 & 675 & 626 & - & - \\
$N_{\text{enter}}$ & 185 & 338 & 581 & 783 & 759 & 885 & 1216 & 1342 & 1182 & 1363 & 1088 & 1063 & 1056 & 796 & 732 & 891 & 346 \\

Table I: Table of monthly parameter values for equity funds defined such that the portfolio contains a fraction of at least 80% stocks. The values for each of the monthly parameters (rows) were calculated for each year (columns). The mean and standard deviation are evaluated for the monthly values in each year.

$R$ - the base 10 log likelihood ratio of a power law fit relative to a log-normal fit as given by equation (3). A negative value of $R$ indicates that the log-normal hypothesis is a likelier description than a power law. For all years the value is negative meaning that the log-normal distribution is more likely.

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

$E[\omega]$ - the mean log size of funds existing at the end of each year.

$Std[\omega]$ - the standard deviation of log sizes for funds existing at the end of each year.

$E[s]$ - the mean size (in millions) of funds existing at the end of each year.

$Std[s]$ - the standard deviation of sizes (in billions) for funds existing at the end of each year.

$\zeta$ - 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), computed for monthly changes.

$\sigma$ - the standard deviation of the mean zero Wiener process (10), computed for monthly changes.

$N_{\text{exit}}$ - the number of equity funds exiting the industry each year.

$N_{\text{enter}}$ - the number of new equity funds entering the industry in each year.
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 2 compares the two hypotheses, making it clear that the log-normal is a much better fit than the power law. For the log-normal QQ plot most of the large values in the distribution fall on the dashed line corresponding to a log-normal distribution, though the very largest values are somewhat above the dashed line. This says that the empirical distribution decays slightly faster than a log-normal. There are two possible interpretations of this result: Either this is a statistical fluctuation or the true distribution really has slightly thinner tails than a log-normal. In any case, since a log-normal decays faster than
a power law, it strongly suggests that the power law hypothesis is incorrect and the log-normal distribution is a better approximation.

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 [Clauset et al., 2007]. 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

\[ p_{LN}(s) = \frac{p(s)}{1 - P(s_{\min})} \]

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

The more probable that the empirical sample is drawn from a given distribution, the larger the likelihood for that set of observations. The ratio indicates which distribution the data are more likely drawn from. We define the log likelihood ratio as

\[ R = \ln \left( \frac{L_{PL}}{L_{LN}} \right). \quad (3) \]

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 2.1. Using the fit parameters, the log likelihood ratio was computed and the results are summarized graphically in Figure 3 and in Table I. The ratio is always negative, indicating that the likelihood for the log-normal hypothesis is greater than that of the power law hypothesis in every year. It seems clear that tails of the mutual fund data are much better described by a log-normal than by a power law.

### 3 Empirical investigation of size dynamics

Our central thesis in this paper is that the mutual fund size distribution can be explained by a stochastic process governed by three key underlying processes: the size change of existing mutual funds, the entry of new funds
Figure 3: A histogram of the base 10 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. The log likelihood ratio is negative in every year, in several cases strongly so.
Table II: Model parameters as measured from the data in different time periods. $\omega_0$ and $\sigma^2_\omega$ are the mean and variance of the average (log) size of new funds described in (4). $\sigma_0$, $\beta$ and $\sigma_\infty$ are the parameters for the size dependent diffusion and $\mu_0$, $\alpha$ and $\mu_\infty$ are the parameters of the average growth rate (9). The confidence intervals are 95% under the assumption of standard errors. The time intervals were chosen to match the results shown in Fig. 9.

and the exit of existing funds. In this section we empirically investigate each of these three processes, providing motivation for the model developed in the next section.

3.1 Fund entry process

We begin by examining the entry of new funds. We investigate both the number of funds entering each year $N_{\text{enter}}(t)$ and the sizes with which they enter. We perform a linear regression of $N_{\text{enter}}(t)$ against the number of existing funds $N(t-1)$, yielding slope $\alpha = 0.04 \pm 0.05$ and intercept $\beta = 750 \pm 300$. The slope is not statistically significant, and so we approximate fund entry as a Poisson process with a constant rate $\nu$.

The size of entering funds is more complicated. In Figure 4 we compare the distribution of the size of entering funds $f(s)$ to that of all existing funds. The distribution is somewhat irregular, with peaks at round figures such as ten thousand, a hundred thousand, and a million dollars. The average size$^{13}$ of entering funds is almost three orders of magnitude smaller

$^{13}$When discussing the average size one must make sure to account for the difference between the average log size and the average size, since due to the heavy tails the difference is striking. The average entry log size $E[\omega_c] \approx 0$, corresponding to a fund of size one million,
than that of existing funds, making it clear that the typical surviving fund grows significantly after it enters. We compared the distribution of entering funds to a log-normal and found that the tails are substantially thinner than those of the log-normal. When we consider these facts (small size and thin tails) it is clear that the distribution of entering funds cannot be important in determining the upper tail of the fund size distribution.

To conclude, the fund entry process is reasonably well approximated as a Poisson process in which an average of $\nu$ funds enter per month, with the size of each fund drawn from a distribution $f(\omega, t)$. The distribution of new fund sizes is approximated as a log-normal distribution in the fund size $s$, that is a normal distribution in the log size $\omega$

$$f(\omega, t) = \frac{1}{\sqrt{\pi \sigma_\omega^2}} \exp \left( -\frac{(\omega - \omega_0)^2}{\sigma_\omega^2} \right) \theta(t - t_0),$$  \hspace{1cm} (4)

where $\omega_0$ is the mean log size of new funds and $\sigma_\omega^2$ is its variance. $\theta(t - t_0)$ is a unit step function ensuring no funds funds enter the industry before the

while if we average over the entry sizes $E[s_c] = E[e^{\omega c}]$, we get an average entry size of approximately 30 million. For comparison, both the average size and the average log size of existing funds are quoted in Table 1.
Figure 5: The number of equity funds exiting the industry $N_{\text{exit}}(t)$ in the year $t$ as a function of the total number of funds existing in the previous year, $N(t-1)$. The plot is compared to a linear regression (full line). The error bars are calculated for each bin under a Poisson process assumption, and correspond to the square root of the average number of funds exiting the industry in that year.

initial time $t_0$. The value of the mean log size and its variance are calculated from the data for different periods, as summarized in Table II.

3.2 Fund exit process

As we will show later, fund exit is of critical importance in determining the long-run properties of the fund size distribution. In Figure 5 we plot the number of exiting funds $N_{\text{exit}}(t)$ as a function of the total number of funds existing in the previous year, $N(t-1)$. The linearly increasing trend is clear, and in most cases the number of exiting funds is within a standard deviation of the linear trend line. This suggests that it is a reasonable approximation to assume that the probability for a given fund to exit is independent of time, so that the number of exiting funds is proportional to the number of existing funds, with proportionality constant $\lambda(\omega)$, where $\omega$ is the logarithm of the fund size. Furthermore, it seems to be reasonable to assume that $\lambda(\omega) = \lambda$ is independent of size; where $\lambda$ is the slope of the linear regression in Figure 5. On an annual time scale this gives $\lambda = 0.092 \pm 0.030$. Under the assumption that fund death is a Poisson process the monthly rate is just the yearly rate divided by the number of months per year. Thus, we
approximate the fund exit process as a Poisson process where funds of size \( \omega \) exit the industry with a constant rate \( \lambda \).

### 3.3 Fund growth

Fund growth is a complicated process as it involves not only size changes due to fund performance but also changes due to investors depositing and withdrawing money. The investor behavior induces correlations and size dependence in the random process. It is convenient to define the relative change in a fund’s size \( \Delta s(t) \) as

\[
\Delta s(t) = \frac{s(t+1) - s(t)}{s(t)}. 
\]  

(5)

The relative change can be decomposed into two parts: the return \( \Delta r \) and the fractional investor money flux \( \Delta f(t) \), which are simply related as

\[
\Delta s(t) = \Delta f(t) + \Delta r(t). 
\]

(6)

The return \( \Delta r \) represents the return of the fund to its investors, defined as

\[
\Delta r(t) = \frac{NAV(t+1) - NAV(t)}{NAV(t)}, 
\]

(7)

where \( NAV(t) \) is the Net Asset Value at time \( t \). The fractional money flux \( \Delta f(t) \) is the change in the fund size by investor deposits or withdrawals, defined as

\[
\Delta f(t) = \frac{s(t+1) - [1 + \Delta r(t)]s(t)}{s(t)}. 
\]

(8)

Since the data set only contains information about the size of funds and their returns we are not able to separate deposits from withdrawals, but rather can only observe their net.

Figure 6 gives an overview of the size dependence for both the returns \( \Delta r \) and the money flux \( \Delta f \). The two behave very differently. The returns \( \Delta r \) are essentially independent of size\(^{14} \). This is expected based on market efficiency,

\(^{14}\)The independence of the return \( \Delta r \) on size is verified by performing a linear regression of \( \mu_r \) vs. \( s \) for the year 2005, which results in an intercept \( \beta = 6.7 \pm 0.2 \times 10^{-3} \) and a slope \( \alpha = 0.5 \pm 8.5 \times 10^{-4} \). This result implies a size independent average monthly return of 0.67%.
Figure 6: The average monthly return $\mu_r$ and its volatility $\sigma_r$, and the money flux $\mu_f$ and its volatility $\sigma_f$, are plotted as a function of the fund size (in millions) for the year 2005 (see Eqs. (5 - 8)). The data are binned based on size, using bins with exponentially increasing size. The average monthly return $\mu_r$ is compared to a constant return of 0.008 and the monthly volatility $\sigma_r$ is compared to 0.03. The average monthly flux $\mu_f$ is compared to a slope of -0.5 and the money flux volatility $\sigma_f$ is compared to a linear slope of -0.35. On a double logarithmic scale a line of slope $\alpha$ corresponds to a power law relation $f(s) = s^\alpha$. 

17
as otherwise one could obtain superior performance simply by investing in larger or smaller funds [Malkiel, 1995]. This implies that equity mutual funds can be viewed as a constant return to scale industry [Gabaix et al., 2006]. Both the mean $\mu_r = E[\Delta_r]$ and the standard deviation $\sigma_r = \text{Var}[\Delta_r]^{1/2}$ are constant; the latter is also expected from market efficiency, as otherwise it would be possible to lower one’s risk by simply investing in funds of a different size.

In contrast, the money flow $\Delta f$ decreases with size. Both the mean money flux $\mu_f = E[\Delta f]$ and its standard deviation $\sigma_f = \text{Var}[\Delta f]^{1/2}$ roughly follow a power law over five orders of magnitude in the size $s$. This is similar to the behavior that has been observed for the growth rates of other types of firms [Stanley et al., 1995, 1996; Amaral et al., 1997a; Bottazzi and Secchi, 2003a]. Several theories have been put forth to explain this for firms in general, but at this point there is no consensus view as to its cause$^{15}$. For mutual funds one would expect that in absolute terms, all else being equal, it is harder to raise large amounts of money than small amounts of money, and thus the relative growth rate should decrease with size. This does not, however, explain why the functional form should be a power law. The fact that similar behavior is observed for other types of firms suggests that there should be a common explanation that holds for firms in general, and is not specific to mutual funds.

These two effects can be combined to model the mean total size change $\mu_s$ and the standard deviation $\sigma_s$. The standard deviations are related through the approximate relation

$$\sigma_s^2 \approx \sigma_f^2 + \sigma_r^2.$$  

This is only an approximation since we neglected the correlation term between $\Delta_r$ and $\Delta f$. Based on our empirical studies the correlation is close to zero and this is a good approximation. In Figure 7 we plot both $\mu_s$ and $\sigma_s$ as a function of size.

The decomposition illustrated in Figure 6 suggests that $\mu_s$ and $\sigma_s$ can be reasonably well approximated as

$$\sigma_s(s) = \sigma_0 s^{-\beta} + \sigma_{\infty}$$

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

$^{15}$There has been a significant body of work attempting to explain the observed growth rate of firms. For the main body of work see [Amaral et al., 1997b; Buldyrev et al., 1997; Amaral et al., 1998; De Fabritiis et al., 2003b; Matia et al., 2004; Bottazzi, 2001b; Sutton, 2001; Wyart and Bouchaud, 2002; Bottazzi and Secchi, 2003b, 2005b; Fu et al., 2005; Riccaboni et al., 2008; Podobnik et al., 2008].
Figure 7: (a) The standard deviation \( \sigma \) in the logarithmic size change \( \Delta_s = \Delta(\log s) \) of an equity fund as a function of the fund size \( s \) (in millions of dollars). (b) The mean \( \mu \) logarithmic size change \( \Delta_s = \Delta(\log s) \) of an equity fund as a function of the fund size \( s \) (in millions of dollars). The data for all the funds were divided into 100 equally occupied bins. \( \mu \) is the mean in each bin and \( \sigma \) is the square root of the variance in each bin for the years 1991 to 2005. The data are compared to a fit according to (9) in Figures (a) and (b) respectively.

As illustrated in Figure 7, these functional forms fit the data reasonably well, with only slight variations of parameters in different periods, as shown in Table II.

To summarize: In one respect, mutual funds are like other firms, in that their money flux shows the same power law decay with size that has been observed for the growth rates of other firms. At the same time, mutual funds are unusual in that a component of their growth comes from the returns of the fund itself; this component is independent of size, as expected from market efficiency. Thus mutual funds are unusual in that, due to market efficiency, their relative growth rate does not go to zero in the limit \( s \to \infty \). The upshot is that for small funds money flux is the dominant growth process and for large funds the return is the dominant growth process. This means that for the largest funds we can approximate the mean and standard deviation of the growth rate as being size independent.
Figure 8: The PDF of aggregated monthly log size changes $\Delta \omega = \log_{10}(1 + \Delta s)$ 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 was zero.

3.4 Diffusion model

We approximate the growth of existing funds as a multiplicative Gibrat-like process\(^\text{16}\) satisfying a stochastic evolution equation of the form

$$ds = s [\mu_s(s)dt + \sigma_s(s)dW_t],$$  \hspace{1cm} (10)

where $W_t$ is a mean zero and unit variance normal random variable, $\mu_s$ is the drift and $\sigma_s$ is the volatility. The use of an underlying Wiener process in the stochastic evolution equation (10) assumes implicitly that the relative size change distributions of funds of a given size is an i.i.d normal random variable.

We first discuss the normality assumption. As can be seen in Figure 8, the relative size change distribution has a tent shape when plotted in semi-logarithmic scale\(^\text{17}\). It can be approximated as a double exponential, also called a Laplace distribution, which has heavier tails than a normal distribution. Nonetheless, due to the fact that the logarithmic increments are

\(^\text{16}\) A Gibrat-like process is a multiplicative process in which the size of the fund at any given time is given as a multiplicative factor times the size of the fund at a previous time. In Gibrat’s law of proportionate effect [Gibrat, 1931] the multiplicative term depends linearly on size while here we allow it to have any size dependence.

\(^\text{17}\) A similar phenomenon was originally observed by Stanley et al for the distribution of firm growth [Stanley et al., 1996; Axtell, 2001; Dosi, 2005].
Table III: Cross-sectional regression coefficients of the monthly fund flow, computed for several months, against the performance in past months, as indicated in Eq. 11. The regression was computed cross-sectionally using data for 6189 equity funds. For example the entry for $\beta_1$ in the first (from the left) column represents the linear regression coefficient of the money flux at the end of 2005 on the previous month’s return. The errors correspond to the 95% confidence intervals.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>$\beta_1$</td>
<td>0.10 ± 0.16</td>
<td>0.40 ± 0.98</td>
<td>0.27 ± 0.68</td>
<td>1.17 ± 4.68</td>
<td>$-0.23 ± 1.24$</td>
</tr>
<tr>
<td>$\beta_2$</td>
<td>0.14 ± 0.27</td>
<td>0.36 ± 1.20</td>
<td>0.48 ± 0.83</td>
<td>$-0.79 ± 3.13$</td>
<td>$-0.65 ± 2.31$</td>
</tr>
<tr>
<td>$\beta_3$</td>
<td>0.28 ± 0.45</td>
<td>0.01 ± 1.07</td>
<td>0.33 ± 0.83</td>
<td>1.79 ± 3.24</td>
<td>0.60 ± 2.57</td>
</tr>
<tr>
<td>$\beta_4$</td>
<td>0.56 ± 0.40</td>
<td>$-0.28 ± 0.85$</td>
<td>0.24 ± 1.27</td>
<td>$-0.28 ± 1.65$</td>
<td>0.44 ± 2.32</td>
</tr>
<tr>
<td>$\beta_5$</td>
<td>0.24 ± 0.43</td>
<td>$-0.25 ± 1.13$</td>
<td>0.21 ± 0.90</td>
<td>$-0.24 ± 2.95$</td>
<td>0.43 ± 2.49</td>
</tr>
<tr>
<td>$\beta_6$</td>
<td>0.48 ± 0.38</td>
<td>$-0.02 ± 1.03$</td>
<td>0.30 ± 0.92</td>
<td>1.27 ± 3.50</td>
<td>0.31 ± 2.09</td>
</tr>
<tr>
<td>$\beta$</td>
<td>$-0.02 ± 0.02$</td>
<td>0.03 ± 0.05</td>
<td>0.01 ± 0.05</td>
<td>0.14 ± 0.21</td>
<td>0.06 ± 0.15</td>
</tr>
</tbody>
</table>

additive, after several steps the central limit theorem causes this to quickly converge to a normal distribution. We have explicitly verified this by tracking a group of funds in a given size range over time. Thus even though the normality assumption is not true on short timescales it rapidly becomes valid on longer timescales.

The second approximation is that of an independent process. Strictly speaking this approximation is also violated, due to the fact that investors react to past performance. This causes correlations in the money flux $\Delta f$, which in turns induces correlations in the total size change $\Delta s$. We find, however, that these correlations are small. This can be seen in Table III. We perform cross sectional regressions for several different times $t$ of the form

$$\Delta f(t) = \beta + \beta_1 \Delta r(t - 1) + \beta_2 \Delta r(t - 2) + \ldots + \beta_6 \Delta r(t - 6).$$  \hspace{1cm} (11)

The results are extremely noisy; for example, eight of the thirty possible coefficients $\beta_i$ are negative and only two of them are significant at the two standard deviation level. The independent process approximation is particularly true for large funds where the size growth is dominated by the return

\footnote{For empirical evidence that investors react to past performance see Remolona et al., 1997; Busse, 2001; Chevalier, Judith and Ellison, Glenn, 1997; Sirri, Erik R. and Tufano, Peter, 1998; Guercio, Diane Del and Tkac, Paula A., 2002; Bollen, 2007.}
\( \Delta_r \), which is essentially uncorrelated due to market efficiency [Bollen and Busse, 2005; Carhart, 1997].

4 An analytical model for the mutual fund size distribution

The empirical evidence described in the previous section motivates a simple stochastic growth model for the equity mutual fund industry. The aim of the model is to describe the time evolution of the size distribution, that is, to solve for the probability density function \( p(\omega, t) \) of funds with (log) size \( \omega \) existing at time \( t \). The size distribution can be written as

\[
p(\omega, t) = \frac{n(\omega, t)}{N(t)}, \tag{12}
\]

where \( n(\omega, t) \) is the number of funds at time with logarithmic size \( \omega \) and \( N(t) = \int n(\omega, t)d\omega \) is the total number of funds at time \( t \). To simplify the analysis we solve separately for the total number of funds \( N(t) \) and for the number density \( n(\omega, t) \).

4.1 Dynamics of the total number of funds

Based on the empirical observations of Section 3, we have argued that it is natural to model the total number of funds as a function of time as

\[
\frac{dN}{dt} = \nu - \lambda N \tag{13}
\]

where \( \nu \) is the rate of creating new funds and \( \lambda \) is the exit rate of existing funds. Under the assumption that \( \nu \) and \( \lambda \) are constant in time and independent of \( N \) this has the solution

\[
N(t) = \frac{\nu}{\lambda} \left( 1 - e^{-\lambda t} \right) \theta(t), \tag{14}
\]

where \( \theta(t) \) is a unit step function at \( t = 0 \), the year in which the first funds enter. This solution has the surprising property that the dynamics only depend on the fund exit rate \( \lambda \), with a characteristic timescale \( 1/\lambda \). For example, for \( \lambda \approx 0.09 \), as estimated in Section 3, the timescale for \( N(t) \) to reach its steady state is only roughly a decade. An examination of Table I makes it clear, however, that \( \nu = \text{constant} \) is not a very good approximation. Nonetheless, if we crudely use the mean creation rate \( \nu \approx 900 \) from Table I
and the fund exit rate $\lambda \approx 0.09$ estimated in Section 3, the steady state number of funds should be about $N \approx 10,000$, compared to the 8,845 funds that actually existed in 2005. Thus this gives an estimate with the right order of magnitude.

The important point to stress is that the dynamics for $N(t)$ operate on a different timescale than that of $n(\omega, t)$. As we will show in the next section the characteristic timescale for $n(\omega, t)$ is much longer than that for $N(t)$.

4.2 Analytical solution for the number density $n(\omega, t)$

We define and solve the time evolution equation for the number density $n(\omega, t)$ using the empirically motivated approximations described earlier. These approximations are: The entry process is modeled as a Poisson process with rate $\nu$, such that at time $t$ a new fund enters the industry with a probability $\nu dt$ and (log) size $\omega$ drawn from a distribution $f(\omega, t)$. The entry size distribution is approximated as normal distribution in log size given by equation (4). The exit process is modeled as an individual Poisson process such that at any time time $t$ a fund exits the industry with a size independent probability $\lambda dt$. The size change is approximated as a (log) Brownian motion with a size dependent drift and diffusion term given by equation (9). Under these assumptions the forward Kolmogorov equation (also known as the Fokker-Plank equation) defining the time evolution of the number density [Gardiner, 2004] is given by

$$\frac{\partial}{\partial t} n(\omega, t) = \nu f(\omega, t) - \lambda n(\omega, t) - \frac{\partial}{\partial \omega} [\mu(\omega)n(\omega, t)] + \frac{\partial^2}{\partial \omega^2} [D(\omega)n(\omega, t)],$$

where $D(\omega) = \sigma(\omega)^2/2$ is the size diffusion coefficient. The first term on the right describes the entry process, the second describes the fund exit process and the third and fourth terms describe the change in size of an existing fund. Under the change of variables $s = \exp(\omega)$ in (9) the size dependent drift and diffusion terms are given by

$$\mu(\omega) = \mu_0 e^{-\alpha \omega} + \mu_\infty$$

$$\sigma(\omega) = \sigma_0 e^{-\beta \omega} + \sigma_\infty.$$ (16)

4.2.1 Approximate solution for large funds

Eqs. (15) and (16) cannot be solved analytically, but a great deal of insight can be gained by going to the large size limit. As previously described, due
to the dominance of the returns, in this limit $\Delta_s \approx \Delta_r$, and $\mu(\omega)$ and $\sigma(\omega)$ are constant, i.e. $\mu = \mu_\infty$ and $\sigma = \sigma_\infty$. The evolution equation becomes

$$\frac{\partial}{\partial t} n(\omega, t) = \nu f(\omega, t) - \lambda n(\omega, t) - \mu \frac{\partial}{\partial \omega} n(\omega, t) + D \frac{\partial^2}{\partial \omega^2} n(\omega, t), \quad (17)$$

where we used the notation $D = \sigma^2_\infty/2$. The upper tail behavior of the distribution described by (17) strongly depends on the exit and entry process. If for example, funds only change in size and no funds enter or exit, then the resulting distribution is normal

$$\tilde{n}(\omega, t) = \frac{1}{\sqrt{4\pi Dt}} \exp \left[ -\frac{(\omega - \mu t)^2}{4Dt} \right], \quad (18)$$

which corresponds to a size distribution $p(s)$ with a lognormal upper tail.

The exit process plays a major role and is responsible for the thickening with time of the upper tail of the distribution. The intuition is as follows: Since each fund exits the industry with the same probability and since there are more smaller sized funds than larger ones the number of small funds exiting the industry is larger. This results in an increasing fraction of larger funds. As we will now show this results in the distribution evolving from a log-normal upper tail to a power law upper tail. In contrast, the entry process is not as important and determines mainly the number of funds in the industry. This is true as long as the entry size distribution $f(\omega, t)$ is not heavier-tailed than a lognormal, which is supported by the empirical data.

In the large size limit the solution for an arbitrary entry size distribution $f$ is given by

$$n(\omega, t) = \nu \int_{-\infty}^{\infty} \int_0^t \exp^{-\lambda t'} \frac{1}{\sqrt{4\pi Dt'}} \exp \left[ -\frac{(\omega - \omega' - \mu t)^2}{4Dt'} \right] f(\omega', t-t') \, dt' \, d\omega'. \quad (19)$$

Stated in words, a fund of size $\omega'$ enters at time $t - \tau$ with probability $f(\omega', t - \tau)$. The fund will survive to time $t$ with a probability $\exp(\lambda \tau)$ and will have a size $\omega$ at time $t$ with a probability according to (18).

If funds enter the industry with a constant rate $\nu$ beginning at $t = 0$, with a log-normal entry size distribution $f(\omega, t)$ centered around $\omega_s$ with width $\sigma_s$ as given by (4), the size density can be shown to be

$$n(\omega, t) = \frac{\nu \mu}{4\sqrt{\gamma D}} \exp \left[ \left( \frac{\gamma + 1}{4} \right) \frac{\sigma_s^2}{\gamma} - \sqrt{\gamma} \frac{\sigma_s^2}{\gamma} + \frac{\mu}{D} (\omega - \omega_s) + \frac{\mu}{2D} (\omega - \omega_s) \right] \times \left( A + \exp \left[ \frac{\sqrt{\gamma} \sigma_s^2}{2} + 2 \frac{\mu}{D} (\omega - \omega_s) \right] B \right), \quad (20)$$
where the drift and diffusion terms are approximated as $\mu = \mu_\infty$ and $D = \sigma_\infty^2/2$. The parameters $A$, $B$ and $\gamma$ are defined as

$$
\gamma = \sqrt{\frac{1}{4} + \frac{\lambda D}{\mu^2}},
$$

\begin{align}
A &= \text{Erf} \left[ \frac{\sigma_s^2 + \frac{\mu}{D} (\omega - \omega_s)}{\sqrt{2}\sigma_s} \right] \\
&- \text{Erf} \left[ \frac{\sigma_s^2 + \frac{\mu}{D} (\omega - \omega_s) - \sqrt{\gamma} \sigma_s^2}{\sqrt{2}\sigma_s + 2\mu^2/D} \right]
\end{align}

and

\begin{align}
B &= \text{Erf} \left[ \frac{\sqrt{\gamma} \left( \frac{\sigma_s^2}{2} + \frac{\mu^2}{D} t \right) + |\frac{\sigma_s^2}{2} + \frac{\mu}{D} (\omega - \omega_s)|}{\sqrt{2}\sigma_s + 2\mu^2/D} \right] \\
&- \text{Erf} \left[ \frac{\sqrt{\gamma} \sigma_s^2 + \left( \frac{\sigma_s^2}{2} + \frac{\mu}{D} (\omega - \omega_s) \right)}{\sqrt{2}\sigma_s} \right]
\end{align}

where Erf is the error function, i.e. the integral of the normal distribution.

### 4.2.2 Steady state solution for large funds

Making a further approximation simplifies the solution considerably. Let us define a large fund as one with $\omega \gg \omega_s$, where $\omega_s$ is the logarithm of the typical entry size of one million USD. For large funds we can approximate the lognormal distribution as having zero width, in other words we assume that all new funds have the same size $\omega_s$. The number density is then given by

$$
n(\omega, t) = \frac{\nu D}{4\sqrt{\gamma} \mu^2} e^{\frac{\mu}{D} (\omega - \omega_s)} \left[ e^{-\sqrt{\gamma} \frac{\mu}{D} |\omega - \omega_s|} \left( 1 + \text{erf} \left[ \sqrt{\gamma} \frac{\mu^2 t}{D} - \frac{|\omega - \omega_s|}{2\sqrt{D} t} \right] \right) \\
- e^{\sqrt{\gamma} \frac{\mu}{D} |\omega - \omega_s|} \left( 1 - \text{erf} \left[ \mu \sqrt{\frac{t}{D}} \left( \frac{1}{2} + \sqrt{\gamma} \right) \right] \right) \right],
$$

where we have approximated the drift and diffusion as $\mu = \mu_\infty$ and $D = \sigma_\infty^2/2$. Since $\gamma > 1/4$ (21), the density vanishes for both $\omega \to \infty$ and $\omega \to -\infty$.

The steady state solution for large times is achieved by taking the $t \to \infty$ limit of (24), which gives

$$
n(\omega) = \frac{\nu}{2\mu \sqrt{\gamma}} \exp \frac{\mu}{D} \left( \frac{\omega - \omega_s}{2} - \sqrt{\gamma} |\omega - \omega_s| \right). \quad (24)
$$
Since the log size density (24) has an exponential upper tail \( p(\omega) \sim \exp(-\zeta_s \omega) \) and \( s = \exp(\omega) \) the CDF for \( s \) has a power law tail with an exponent\(^{19} \) \( \zeta_s \), i.e.

\[
P(s > X) \sim X^{-\zeta_s}. \tag{25}
\]

Substituting for the parameter \( \gamma \) using Eq. (21) for the upper tail exponent yields

\[
\zeta_s = -\mu + \sqrt{\mu^2 + 4D\lambda} \quad \frac{2D}{\mu}. \tag{26}
\]

Note that this does not depend on the creation rate \( \nu \). Using the average parameter values in Table II the asymptotic exponent has the value

\[
\zeta_s = 1.2 \pm 0.6. \tag{27}
\]

This suggests that when the distribution will reach its steady state the tails will be in agreement with those expected under Zipf’s law, \( \zeta_s \approx 1 \).

4.2.3 The importance of \( \mu_\infty \) and \( \sigma_\infty \)

The non vanishing drift \( \mu_\infty > 0 \) and diffusion terms \( \sigma_\infty > 0 \) are essential for the distribution to evolve towards a power law. As can be seen from the fit parameters, based on data for \( \Delta_s \) alone, we cannot strongly reject the hypothesis that the drift and diffusion rates vanish for large sizes, i.e. \( \mu_\infty \to 0 \) and \( \sigma_\infty \to 0 \). However, because the size change \( \Delta_s \) can be decomposed into \( \Delta_r \) and \( \Delta_f \), and because the mean of \( \Delta_r > 0 \) due to the growth of the market, we are confident that neither \( \mu_\infty \) nor \( \sigma_\infty \) are zero.

This distinguishes mutual funds from other types of firms, which are typically observed empirically to have \( \sigma_\infty = 0 \) [Stanley et al., 1996; Matia et al., 2004]. It suggests that in a sufficiently stationary situation and after the passage of sufficient time their distributions should be different. Previous empirical studies for other firms found \( \mu_\infty = 0 \) and \( \sigma_\infty = 0 \). Assuming that other types of firms obey similar diffusion equations to those used here, it can be shown that the resulting distribution has a stretched exponential upper tail, which is much thinner than a power law\(^{20} \).

\(^{19}\)To calculate the tail exponent of the density correctly one must change variables through \( p(s) = p(\omega) \frac{d\omega}{ds} \sim s^{-\zeta_s - 1} \). This results in a CDF with a tail exponent of \( \zeta_s \).

\(^{20}\)A stretched exponential is of the form \( p(x) \sim \exp(ax^{-b}) \), where \( a \) and \( b \) are positive constants. There is some evidence in the empirical data that the death rate \( \lambda \) also decays with size. However, in our simulations we found that this makes very little difference for the size distribution as long as it decays slower than the distribution of entering funds, and so in the interest of keeping the model parsimonious we have not included this in our model.
4.2.4 Timescale to reach steady state

The most useful aspect of solving for the time dependent solution is that it provides the ability to estimate the timescale to reach the steady state solution. The time dependence in Eq. 24 is contained in the arguments of the error function terms on the right. When these arguments become large, say larger than 3, the solution is roughly time independent, and can be written as

\[ t > \frac{9D}{4\gamma\mu^2} \left( 1 + \sqrt{1 + \frac{2 \sqrt{\gamma\mu^2}}{9D} |\omega - \omega_s|} \right)^2. \] (28)

Using the values obtained by fitting the whole data set, in units of months \( \mu = \mu_\infty \approx 0.005, D = \sigma_\infty^2/2 \) and \( \sigma_\infty \approx 0.05 \). This gives

\[ t > 180 \left( 1 + \sqrt{1 + 0.7 |\omega - \omega_s|} \right)^2, \]

where the time is in months. Plugging in some numbers makes it clear that the time scale to reach steady state is very long. For instance, for funds of a billion dollars it will take about 170 years for their distribution to come within 1 percent of its steady state. This agrees with the observation that there seems to be no significant fattening of the tail over nearly two decades since 1991. Note that the time required for the distribution \( n(\omega, t) \) to reach steady state for large values of \( \omega \) is much greater than that for the total number of funds \( N(t) \) to become constant.

4.3 Comparisons to empirical data

We are unable to find an analytic solution for the general case of Eq. (15) including the size dependence of the diffusion and drift terms, and so we employ a simulation. This was done on a monthly time scale, averaging over 1000 different runs to estimate the final distribution. As we have emphasized in the previous discussion the time scales for relaxation to the steady state distribution are long. We thus take the evidence for the huge increase in the number of new funds seriously. We begin the simulation in 1991 and simulate the process for varying periods of time, making our target the empirical data at the end of the period. In each case we assume the size distribution for injecting funds is log-normal, as discussed in Section 3.1.

To compare our predictions to the empirical data we measure the parameters for each of the processes of fund creation, entry and exit using data from the same period as the simulation, summarized in Table II. A
key point is that we are not fitting these parameters on the target data for fund size\(^{21}\), but rather are fitting them on the individual creation, exit and diffusion processes and then simulating the corresponding model to predict fund size. One of our main predictions is that the time dependence of the solution is important. In Figure 9 we compare the predictions of the simulation to the empirical data at two different ending times. The model fits quite well at all time horizons, though the fit in the tail is somewhat less good at the longest time horizon. This indicates that even for this model there is a slight tendency to estimate a tail that is heavier than that of the data. Note, however, that the simulations also make it clear that the fluctuations in the tail are substantial, and this deviation is very likely due to chance – many of the individual runs of the simulation deviate from the mean of the

\(^{21}\)It is not our intention to claim that the processes describing fund size are constant or even stationary. Thus, we would not necessarily expect that parameters measured on periods outside of the sample period will be a good approximation for those in the sample period. Rather, our purpose is to show that these processes can explain the distribution of fund sizes.
5 Comparison with the hypothesis of Berk and Green

As mentioned in the introduction, Berk and Green [2004] have offered an alternative hypothesis to ours, based on the distribution of fund manager skill and the size dependence of transaction cost. Under market efficiency the after-transaction returns of funds should be independent of size. Transaction costs are known to increase with trading size. Thus, they argue, the difference between a large fund and small fund is that larger funds have higher before-transaction cost returns, corresponding to superior skill. This skill allows the funds to grow larger and make better profits by collecting more fees.

Given a distribution $P(\alpha)$ of investor skill $\alpha$ and a functional form $g(s)$ for transaction costs, the Berk and Green model makes a prediction about the distribution of fund sizes. Berk and Green somewhat arbitrarily assumed that $P(\alpha)$ is normal and that $g(s)$ is linear. This implies a normal distribution of fund size, which is clearly very different than what is observed. A more realistic assumption for transaction costs would be $g(s) \sim s^b$, with $b \approx 0.5$. Under this assumption it is possible to show that $P(\alpha)$ must have a log-normal upper tail in order to be consistent with the data. (This is true for any $b > 0$ including the Berk and Green choice of $b = 1$). Whether or not this is plausible is not clear; certainly it means that the distribution of skill must be very heavy tailed. What is clear is that their theory has little predictive value for fund size as it stands, since the prediction of fund size is itself based on an unknown distribution for skill, which requires as much information to specify as the prediction itself.

In contrast, the prediction made by our hypothesis is very robust. While getting a good fit to the true fund size distribution requires several assumptions, as we have spelled out here, under our model the log-normal upper tail depends only on the fact that returns are essentially random, and that the timescales to achieve the asymptotic distribution are extremely long. The essence of the answer emerges from the simple fact that returns are multiplicative.

\[\text{For a select body of work on the functional form of transaction cost see [Torre, 1997; Kempf and Korn, 1999; Farmer and Lillo, 2004; Gabaix et al., 2003; Gabaix et al., 2006; Hopman, 2007; Bouchaud et al., 2008] and references therein.}\]
This brings up the question of whether or not our theory can be compatible with the Berk and Green hypothesis. In our theory the aspect that transaction costs might affect is the mean inflow of money to funds as a function of size, \( \mu_f(s) \). This is where investor choice might enter our model. For example, one can hypothesize that the net rate of deposits to larger funds decreases with size due to the fact that investors correctly understand that there are increased transaction costs. Alternatively, fund managers may understand that they have higher costs and close their funds or take less relative effort in marketing them as they get bigger\(^{23}\).

While both of these effects are plausible, they do not qualitatively change the results of our model. The lognormal upper tail is very robust under variations of the investor money flux – the upper tail of the distribution is essentially independent of \( \mu_f(s) \) and \( \sigma_f(s) \), as long as they stay within reasonable bounds\(^{24}\). Thus, it seems that the diffusive nature of fund size dominates investor choice in determining the fund size distribution.

We are in the process of making a more detailed study quantitatively testing the Berk and Green hypothesis, under several additional criteria that we do not have the space to explain here. We believe, though, that the data presented here already strongly suggest that their hypothesis cannot provide the correct explanation for fund size, while ours does.

### 6 Conclusions

In this paper we have shown that the empirical evidence strongly favors the hypothesis that the upper tail of the distribution of fund size is not a power law and is better approximated by a log-normal. Even though the log-normal distribution is strongly right skewed, and thus in some respects is a heavy-tailed distribution, it is still much thinner-tailed than a power law, as evidenced by the fact that all its moments exist. As a consequence there are fewer extremely large funds than one would expect if it were a power law.

We have argued that the mutual fund size distribution is naturally explained in terms of the stochastic nature of the growth process. The essential elements of the growth process are multiplicative random changes in the size

\(^{23}\)One must bear in mind that the rate only drops relative to size; in absolute terms the rate is increasing, so perhaps what one needs to explain is why there is an increase with fund size.

\(^{24}\)The requirement is that \( \mu_s(s) \) remain greater than zero. It fails if either \( \mu_f(s) \) or \( \sigma(s) \) increases with size, which seems implausible, or if \( \mu_f(s) \) becomes sufficiently negative that \( \mu_s(s) < 0 \). Neither of these are consistent with the data.
of existing funds, entry of new funds, and exit of existing funds as they go out of business. We find, however, that entry plays no role at all other than setting the scale; exit plays a role in thickening the tails of the distribution, but this acts only on a very slow timescale. The log-normality comes about because the industry is young and still in a transient state. In the future, if the conditions for fund growth and removal remain stationary for more than a century, the distribution should become a power law. The thickening of the tails happens from the body of the distribution outward, as the power law tail extends to successively larger funds.

There is also an interesting size dependence in the growth rate of mutual fund size, which is both like and unlike that of other types of firms. Mutual funds are distinctive in that their overall growth rates can be decomposed as a sum of two terms, \( \Delta s = \Delta f + \Delta r \), where \( \Delta f \) represents the flow of money in and out of funds, and \( \Delta r \) the returns on money that is already in the fund. The money flow \( \Delta f \) decreases as a power law as a function of size, similar to what is widely observed in the overall growth rates for other types of firms. Furthermore the exponents are similar to those observed elsewhere. The returns \( \Delta r \), in contrast, are essentially independent of fund size, as they must be under market efficiency. As a result, for large sizes the mean and variance of the overall growth are constant – this is unlike other firms, for which the mean and variance appear to go to zero in the limit. As we discuss here, this makes an important difference in the long-term evolution: While mutual funds should eventually evolve toward a power law, other firms should evolve toward a thinner-tailed, stretched exponential distribution.

Our analysis here suggests that investor preference has a negligible influence on the upper tail of the mutual fund size distribution. Investor preference enters our analysis only through \( \Delta f \), the flow of money in and out of the fund. Since \( \Delta f \) becomes relatively small in the large size limit, in this limit the growth of funds is dominated by the returns \( \Delta r \), whose mean and variance are constant. Thus the upper tail of the size distribution is determined by market efficiency, which dictates both that returns are essentially random, and thus diffusive, and that there is no dependence on size. Thus for large fund size investor preference doesn’t have much influence on the growth process; this is reinforced by the fact that the statistical properties of the money flux \( \Delta f \) are essentially like those of the growth of other firms. This, as well as other arguments presented in the previous section, indicates that transaction costs do not play a major role in determining fund size, as suggested by Berk and Green (2004).

Our results present a puzzle as to what determines the distribution of trading volume. The existence of a power law for large trading volumes
is now fairly well documented\textsuperscript{25}. A possible explanation for the volume distribution assumes that fund trading is proportional to its size and relies on the assumption that funds size is described by a power law obeying Zipf’s law. Our analysis here shows that this argument fails for two reasons: (1) Mutual funds sizes do not follow Zipf’s law, and are not even power law distributed and (2) transaction costs are not important in determining fund size. Thus the puzzle of what determines the distribution of trading volume remains an open question.

We have argued here that the cause of the log-normal upper tail of the mutual fund size distribution is essentially market efficiency, combined with the fact that fund growth is a slow-moving, diffusive process that takes place on time scales measured in decades. Investors are not making rational allocations as posited by Berk and Green; rather they move their money in and out of funds slowly, and with a great deal of inertia. This explanation might be disappointing to some as having “little economic content”; its strength is that it explains an important economic phenomenon in a manner that is very robust, and largely independent of the details of human choice.

\textsuperscript{25}For work arguing for a power law volume distribution see [Gopikrishnan et al., 2000; Gabaix et al., 2003; Lillo et al., 2005; Gabaix et al., 2006], while Eisler and Kertesz [2006] argue against one.
A Simulation model

We simulate a model with three independent stochastic processes. These processes are modeled as Poisson process and as such are modeled as having at each time step a probability for an event to occur. The simulation uses asynchronous updating to mimic continuous time. At each simulation time step we perform one of three events with an appropriate probability. These probabilities will determine the rates in which that process occurs. The probability ratio between any pair of events should be equal to the ratio of the rates of the corresponding processes. Thus, if we want to simulate this model for given rates our probabilities are determined.

These processes we simulate are:

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 fund exit rate \( \lambda \) which can depend on the fund size.

3. The rate of creation of new funds \( \nu \).
   Each new fund enters with a size \( \omega \) with a probability density \( f(\omega) \).

Since some of these processes are defined per firm as opposed to the creation process, the simulation is not straightforward. We offer a brief description of our simulation procedure.

1. At every simulation time step, with a probability \( \nu / (1 + \lambda + \nu) \) a new fund enters and we proceed to the next simulation time step.

2. If a fund did not enter then the following is repeated \( (1 + \lambda)N \) times.
   a. We pick a fund at random.
   b. With a probability of \( \lambda / (1 + \lambda) \) the fund enters.
   c. If it is not annihilated, which happens with a probability of \( 1 / (1 + \lambda) \), we change the fund size.

We are interested in comparing the simulations to both numerical and empirical results. The comparisons with analytical results are done for specific times and for specific years when comparing to empirical data. In order to do so, we need to convert simulation time to "real" time. The simulation time can be compared to ‘real’ time if every time a fund does not enter we add a time step. Because of the way we defined the probabilities each
simulation time step is comparable to $1/(1 + \lambda)$ in "real" time units. The resulting "real" time is then measured in what ever units our rates were measured in. In our simulation we use monthly rates and as such a unit time step corresponds to one month.

References


Eisler, Z. and Kertesz, J. (2006). Why do hurst exponents of traded value increase as the logarithm of company size?


