Complex Phenomena in Social and Financial Systems: From bird population growth to the dynamics of the mutual fund industry

Thesis by
Yonathan Schwarzkopf

In Partial Fulfillment of the Requirements
for the Degree of
Doctor of Philosophy

California Institute of Technology
Pasadena, California

2011

(Defended July 26, 2010)

To Ma'ayan without you I would not be where I am today nor the man I am today.

Acknowledgements

I would like to thank my adviser Doyne Farmer for taking me under his wing and for allowing me to be his apprentice for a short while. I am forever grateful. I would like to thank my adviser at Caltech Mark Wise who has allowed me to set my own academic path. My advisers Doyne Farmer and Mark Wise have been great academic role models, mentors, advisers, and supporters. I feel very lucky to have had the opportunity to work with them both. I would like to thank Vineer Bhansali who together with Mark Wise have graciously agreed to let me participate in some of their projects and from whom I have learnt a great deal. I would like to thank all the people at the Santa Fe Institute and especially Robert Axtell, Fabrizio Lillo, Bence Toth, Aaron Clauset and Gabriele La Spada. I would like to thank my committee and the researchers at SISL; John Ledyard, Peter Bossaerts and Jaksa Cvitanic. For financial support I would like to thank the Santa Fe Institute and PIMCO. Special thanks are in order to my friends here at Caltech who have provided support and laughs along the way: Or Yogev, Morgan Putnam, Itamar Yaakov, Oren Schaedel and Jonathan Sternberg. On a more personal level, I would like to thank my father Jimmy, my sister Sharon, my brother Ran and most importantly my son Ori and my wife Ma'ayan.

Abstract

This work explores different aspects of the statics and dynamics of the mutual fund industry. In addition, we answer a major question in the field of complex systems; the anomalous growth fluctuations observed for systems as diverse as breeding birds, city population and GDP.

We study how much control is concentrated in the hands of the largest mutual funds by studying the size distribution empirically. We show that it indicates less concentration than, for example, personal income. We argue that the dominant economic factor that determines the size distribution is market efficiency and we show that the mutual fund industry can be described using a random entry, exit and growth process.

Mutual funds face diminishing returns to scale as a result of convex trading costs yet there is no persistence nor a size dependence in their performance. To solve this puzzle we offer a new framework in which skillful profit maximizing fund managers compensate for decreasing performance by lowering their fees. We show that mutual fund behavior depends on size such that bigger funds charge lower fees and trade less frequently in more stocks. We present a reduced form model that is able to describe quantitatively this behavior.

We conclude with an investigation of the growth of mutual funds due to investor funds flows. We show that funds exhibit the same unusual growth fluctuations that have been observed for phenomena as diverse as breeding bird populations, the size of U.S. firms, the GDP of individual countries and the scientific output of universities. To explain this we propose a remarkably simple additive replication model. To illustrate how this can emerge from a collective microscopic dynamics we propose a model based on stochastic influence dynamics over a scale-free contact network.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

> **Truth is ever to be found in simplicity, and not in the multiplicity and confusion of things.** *Isaac Newton.*

1.1 Thesis Overview

This thesis was written as part of my work as a graduate fellow at the Santa Fe Institute under the supervision of J. Doyne Farmer. The work is in the relatively young field of Econophysics and uses the financial market and more specifically mutual funds as an empirical laboratory for the study of complex systems.

In Chapter 2 we examine how much control is concentrated in the hands of the largest mutual funds; a question best 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 much better described by a log-normal than a power law, indicating less concentration than, for example, personal income. The results are highly statistically significant and are consistent across fifteen years. This contradicts a recent theory concerning the origin of the power law tails of the trading volume distribution.

Then, in Chapter 3, we argue that the dominant economic factor that determines the size distribution is market efficiency, which dictates that fund performance is size independent and that fund growth is essentially random. The random process is characterized by entry, exit and growth. We present a new time-dependent solution for the standard equations used in the industrial organization literature and show that relaxation to the steady-state solution is extremely slow. Thus, even if these processes were stationary (which they are not), the steady-state solution, which is a very heavy-tailed power law, is not relevant. The distribution is instead well-approximated by a less heavy-tailed log-normal. Our random model contradicts the predominant belief that investor choice, mediated by trading costs, ensures that the size of a fund is directly related to its manager's skill (her alpha). To this end we investigate two things: the stochastic nature of investor fund flows and the role that managerial skill and trading costs play in mutual fund behavior.

It is well established amongst practitioners that market impact induces diseconomies of scale in mutual funds. Nevertheless, market impact and it detrimental effect on mutual fund performance has yet to be fully accepted in the academic world. In Chapter 4 we show that market impact is not unique to financial markets but rather that for any market the laws of supply and demand dictate its existence. We argue using the CRSP holdings dataset that mutual funds trade mostly with counter-parties outside the industry and assuming that mutual funds act as liquidity takers their performance suffers due to market impact. We further provide an approximate functional form for mutual funds' trading costs.

Mutual funds face diminishing returns to scale as a result of convex trading costs yet there is no persistence nor a size dependence in their performance. To solve this puzzle we offer in Chapter 5 a new framework in which skillful profit maximizing fund managers compensate for decreasing performance by lowering their fees. We show that mutual fund behavior depends on size such that bigger funds charge lower fees and trade less frequently in more stocks. We present a reduced form model that is able to describe quantitatively this behavior. Our model is simple enough to have economic intuition yet rich enough to describe the observed stylized facts; compared to small funds the average large fund decreases the turnover rate by 50%, it increases the number of positions by 30% and reduces the overall expense ratio by approximately 50%. In addition we offer a functional form for the average before costs performance and for the trading costs of mutual funds.

In Chapter 6 we investigate the growth of mutual funds due to investor funds flows. We show that funds exhibit the same unusual growth fluctuations that have been observed for phenomena as diverse as breeding bird populations, the size of U.S. firms, the GDP of individual countries and the scientific output of universities. The fluctuations display characteristic features, including double exponential scaling in the body of the distribution and power law scaling of the standard deviation as a function of size. To explain this we propose a remarkably simple additive replication model: At each step each individual is replaced by a new number of individuals drawn from the same replication distribution. If the replication distribution is sufficiently heavy tailed then the growth fluctuations are Levy distributed. We analyze the mutual fund data as well as data from bird populations and firms and show that our predictions match the data well. To illustrate how this can emerge from a collective microscopic dynamics we propose a model based on stochastic influence dynamics over a scale-free contact network and show that it produces results similar to those observed. We also extend the model to deal with correlations between individual elements. Our main conclusion is that the universality of growth fluctuations is driven by the additivity of growth processes and the action of the generalized central limit theorem.

To conclude, this thesis covers the different aspects of mutual fund dynamics and makes contribution to finance, industrial organization and to the field of complex systems. Section 1.2 provides a short introduction to the mutual fund industry. Section 1.3 offers some motivation to the work discussed in this thesis. We conclude this chapter with a short introduction to the field of Econophysics in Section 1.4.

1.2 Mutual fun(d) facts

Mutual funds offer a way for people, who lack the sufficient funds, to diversify their investments by pooling funds from many investors. The first mutual fund, Massachusetts Investors Trust (now MFS Investment Management), was founded on March 21, 1924. At the end of the first year it had 200 investors and 392,000 USD in assets. The entire mutual fund industry in 1924 managed less than 10 million USD.

The mutual fund industry started growing with a fast pace after the 1975 change in the Internal Revenue Code allowing individuals to open individual retirement accounts (IRAs). While at the end of the 1970s the mutual fund industry was at its infancy, today the industry plays an important role in the world economy. In 2009 [1] there were 7691 mutual funds, out of which 44% are equity funds, with a total asset value of about 11 trillion USD. This is a very large number when compared to the US GDP of 14.4 trillion USD in 2009 or to the entire US market capitalization of approximately 15 trillion USD at the end of 2009. Mutual funds are important not only because of their asset value but also because a large fraction of the country is invested in some way or another in mutual funds. As of 2009, 87 million Americans representing 50.4 million households (43% of all households) invest in mutual funds. Even more impressive is the fact that 51% of all direct contribution (DC)

[1] Data is taken from the Investment Company Institute's 2010 factbook available online.

retirement plan assets are invested through mutual funds. This growth is illustrated in Figure 1.1

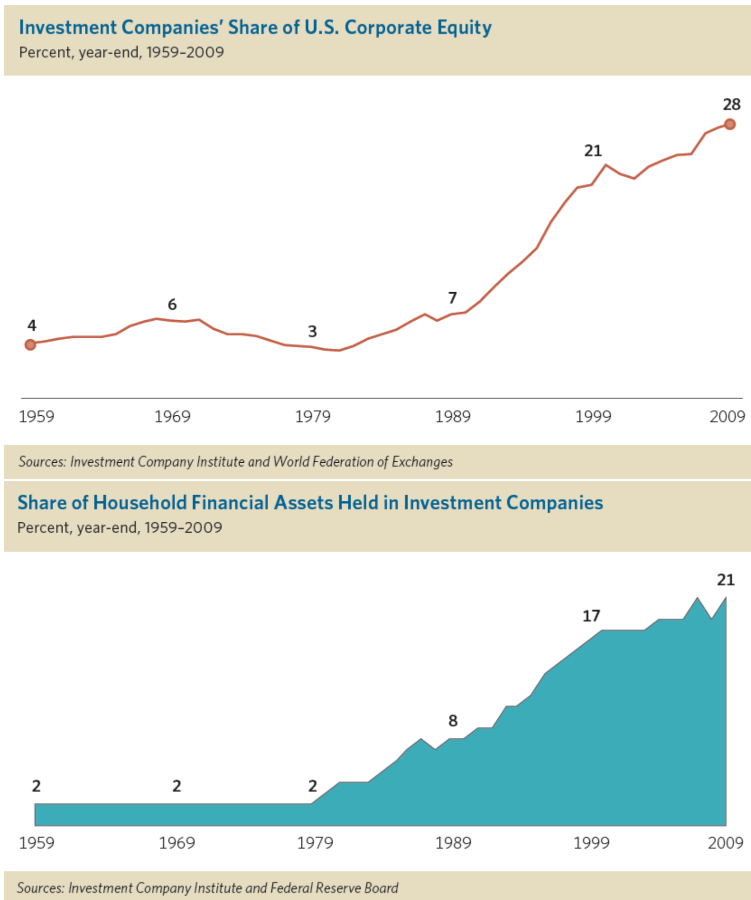


Figure 1.1: Here we illustrate the dramatic growth of the mutual fund industry in the past 50 years. Figures are taken from the Investment Company Institute's 2010 factbook.

1.3 Motivation

1.3.1 Regularities and scaling laws in financial markets

In Chapter 2 we analyze empirically the size distribution of the mutual fund industry. The work has originated as an attempt to describe several of the recently observed scaling rules and regularities in financial markets: the high occurrence of large stock price movements, the high occurrence of large trades and the fact that future trades are strongly correlated with past trades.

The distribution of the dollar value of a trade v was observed to have a power law upper tail[2]

$$P(v > X) \sim X^{-\zeta_v}, \tag{1.1}$$

where the tail exponent was observed to be $\zeta_v \approx 3/2$ across different stocks and different markets (Gopikrishnan et al., 2000). This is a surprising observation as it implies that trades do not have a typical size and that the frequency of very large trades is so high that the second moment does not exist. Similarly, price movements were also observed as having extremely large fluctuations. The log return defined as $r_\tau = \log(p_{t+\tau}/p_t)$, where p_t is the price of a stock at time t, was shown to obey

$$P(r > X) \sim x^{-\zeta_r}, \tag{1.2}$$

with $\zeta_r \approx 3$ across different stocks and different markets (Lux, 1996; Longin, 1996; Plerou et al., 1999b). This is a an important result since it implies that there is no typical size for price movements. This is in stark contradiction with most financial models in which the return is assumed to have a normal distribution[3].

The fact that the above regularities were observed across different stocks and in different markets implies that the cause is not necessarily related to a particular stock (or firm) but rather something more general. A plausible explanation for the observed regularities is the existence of large financial players. One can hypothesize that large players trade large quantities corresponding to the observed large trades. Under the assumption that trading affects the price, i.e. buying a stock raises its price, then these large trades might be responsible for the observed large price movements. Moreover, to avoid large price movements these large orders are split into small packages and traded over longer time periods. This in turn can cause the observed correlation between future trades and past trades (Lillo and Farmer, 2004; Bouchaud et al., 2004). This line of thought is very appealing since not only is this a possible explanation for the above regularities but it also connects them using a simple and elegant argument.

[2]The notation $y \sim x$ implies that asymptotically, i.e. for large x, y is proportional to x.

[3]Ever since Bachelier (1900), stock returns have been treated as random variables. However, dating back to R. (1931), the random movement has been categorized as a geometric random walk. This assumption is widely used in the foundations of finance such as the capital asset pricing model (CAPM) (Sharpe, 1964).

This argument was presented by Gabaix et al. (2006) using mutual funds as a proxy for large players. The argument is based on the empirical 'law' that trading a volume v causes an expected price change r that scales with volume as a square root [4]

$$E[r|v] \sim v^{1/2}. \tag{1.3}$$

This, as they argue, forces mutual funds to trade a volume that is related to its size s (in US dollars) to the power of two thirds

$$v \sim s^{2/3}. \tag{1.4}$$

This results in a scaling relation between the size of mutual funds, the volume they trades and the resulting price movements. Under the assumption that mutual funds have the same power law size distribution as firms [5]

$$P(s > X) \sim X^{-\zeta_s} \tag{1.5}$$

with a tail exponent $\zeta_s \approx 1$ [6], one gets that $\zeta_v = 3/2$ and $\zeta_r = 3$ [7], in agreement with the observed exponents. Thus, as a first step in determining the feasibility of such an argument is to study the size distribution of mutual funds, which we do in Chapter 2.

1.3.2 Market concentration

The study of mutual funds presented in this thesis is important in and of itself since mutual funds represent a large fraction of the financial market and thus play an important role in the U.S. economy (as described in Section 1.2). Naturally this leads to the question of whether the large influence that mutual funds assert on the U.S. financial system spread across many funds, or is it concentrated in only a few? In view of the latest financial crisis

[4] The effect of trading an asset on the price of the asset has been studied extensively in recent years. Even though the argument Gabaix et al. give for the square root law is debatable (Gillemot et al., 2006), the empirical evidence does not reject such a hypothesis. A more detailed discussion and literature review can be found in Chapter 4.

[5] Axtell (2001) was first to show that the size distribution of US firms has a power law upper tail.

[6] A power law distribution with a tail exponent ≈ 1 is commonly referred to as a Zipf distribution.

[7] For random variables a and b that obey $P(a > X) \sim X^{\zeta_a}$ and $P(a > X) \sim X^{\zeta_a}$, that obey the scaling relation $a \sim b^{\alpha}$ one can show that the tail exponents are related through $\zeta_a = \zeta_b/\alpha$.

one might be concerned with the relative size of the largest funds.

While it is standard in economics to describe distributional inequalities in terms of statistics such as the Gini or Herfindahl indices, discussed in Chapter 3, this approach is inadequate to describe the concentration in the tail. Instead, the best way to describe the concentration of assets is in terms of the functional form of the tail. As is well-known in extreme value theory (Embrechts et al., 1997), the key distinction is whether all the moments of the distribution are finite. If the tail is truly concentrated, the tail is a power law, and all the moments above a given threshold, called the tail exponent, are infinite. Thus power law tails imply a very high degree of concentration. We show in Chapter 2 that the tail of the mutual fund size distribution is not a power law, and is well-approximated by a lognormal, for which all of the moments exist. indicating less concentration than, for example, personal income (Silva and Yakovenko, 2005) or US firms (Axtell, 2001).

In Chapter 3, motivated by the results described in Chapter 2, we investigate what economic factors determine the tail properties of the mutual fund distribution. There are two basic types of explanation. One type of explanation is based on a detailed description of investor choice, and another is based on efficient markets, which predicts that growth should be random, and that the causes can be understood in terms of a simple random process description of entry, exit and growth. Of course market efficiency depends on investor choice, but the key distinction is that the random process approach does not depend on any of the details, but rather only requires that no one can make superior investments based on simple criteria, such as size. Our work shows that a simple model based on market efficiency provides a good explanation of the concentration of assets, suggesting that other effects, such as transaction costs or the behavioral aspects of investor choice, play a smaller role.

Explanations based on investor choice can in turn be divided into two types: rational and behavioral. For example, Berk and Green (2004) have proposed that investors are rational, making investments based on past performance. Their theory implies that the distribution of fund size is determined by the skill of mutual fund managers and the dependence of transaction costs on size. This view is somewhat in disagreement with behavioral studies, such as the work by Barber et al. (2005), where they have shown that investors are influenced by other factors such as marketing and advertising. These observations play an important role in both Chapter 5 and Chapter 6.

Another motivation comes from industrial organization since, after all, mutual funds are firms and so why is it that they do not exhibit the same power law concentration that has been observed for firms? The question is especially relevant since the random process approach was originally pioneered as an explanation for firm size by Gibrat, Simon and Mandelbrot (Gibrat, 1931; Simon, 1955; Simon, H. A. and Bonini, Charles P., 1958; Mandelbrot, 1963; Ijiri and Simon, 1977). While mutual funds are in many respects like other firms, we show in Chapter 3 that market efficiency introduces effects that make their growth process distinctly different.

1.3.3 Mutual Fund Flows

Investors play an important role in the dynamics of mutual funds. In fact, we show in Chapter 3 that for small funds investor fund flows are the primary source of growth. Therefore it is important to understand the interaction between mutual funds and investors [8]. The questions that are most relevant to us are: how do investors choose which fund to put their money into, what parameters are they taking into account when they make this decision and more importantly, are they rational?

Typically the answers to these questions belong to one of two types; investors are rational or investors are irrational and their behavior is a mystery best left for neuroscientists and psychologists. This rational investor view backed by the theoretical work of Berk and Green (2004), discussed in Chapter 5. Even though this line of thought has a big following, there are many, such as Fama and French (2010) and others, that show that actively managed mutual funds underperform the benchmark (on average) and therefore argue that investors are not rational. In Chapter 5 we discuss other arguments against the Berk and Green point of view. For most of us, intuition alone suggests that the way investors choose where to invest their money is far from the Berk and Green rational view.

Rationality aside, there is no reason to believe that the mechanism in which investors choose which fund to invest in is unique. While Berk and Green have nicely shown that a mutual fund can not be treated simply like any other firms, the investors are the same human beings that buy the product of all the other firm. After all, for many investors choosing a mutual fund is one of many choices they face on a regular basis. Moreover, this choice is similar to many other choices such as: deciding on a mortgage, buying a new car,

[8] This can be viewed under as part of the study of delegated portfolio management. For a review see (Stracca, 2006).

buying a new camera, etc. What most of these choices have in common is that there is a vast amount of products to chose from, the choice involves a lot of technical know-how that the consumer/investor does not necessarily have and most importantly, many other consumers/investors face the same dilemma.

Even though our choices are not always rational, it does not mean we cannot explain them using a simple mathematical model. In fact, in Chapter 6 we do just that. Our model makes the natural assumption that the choices we make, including which fund to invest in, are influenced by our surroundings. Not surprisingly Barber et al. (2005) have shown that choosing a fund is influenced by marketing and advertising. Still, there is another environmental factor that can be as influential or even more influential than marketing; our social contacts, i.e. friends, family, co-workers etc. [9]

Thus, one can model investor choice as a contagion process taking place on a social network. What this model implies is that the fund I choose to invest in will be influenced by the recommendations of my social contacts. This is a general model that is as suited to describe firm sales (cameras, cars, etc.) as it is suited to describe mutual fund investor flows. When simplified to its bare essence this model is a very general additive replication process. In Chapter 6 we test this model on the observed growth of mutual funds and find that the model is in superb agreement with the data.

Moreover, in Chapter 6 we show that the growth of mutual funds due to investor fund flows is similar to the growth of many other systems: firm sales, GDP growth, urban population growth, scientific output growth and bird population growth, among others. All these systems exhibit growth fluctuations that display characteristic features, including double exponential scaling in the body of the distribution and power law scaling of the standard deviation as a function of size.

1.4 Econophysics

Econophysics, a term coined by Stanley et al. (1996), is a discipline in which the methodology and tools of statistical physics are applied to the study of economical and financial systems. Econophysics is regarded as an empirical discipline that seeks to discover empirical

[9]There is a growing branch in marketing science that is dedicated to understanding the role our social network plays in shaping our decisions. See Watts and Dodds (2007) and references within.

regularities, define empirical 'laws' and develop theories to explain these regularities[10]. This is part of a larger effort to study Complex Adaptive Systems (CAS), a term that was first defined and used by the Santa Fe Institute. We define CAS as systems of adapting agents with non-linear microscopic interactions, which lead to the emergence of macroscopic behavior. Usually these systems are out of equilibrium. Economic and financial systems serve as perfect examples of CAS and offer abundant high quality data. Consequently a large fraction of past and current research is focused on financial markets since the large amounts of quantitative data facilitate the construction of data driven and falsifiable theories.

[10] Econophysics has gained much attention in the past decade and was the subject of numerous editorial and commentary articles appearing in some of academia's top journals. For an example of such articles see Bouchaud (2008); Farmer and Foley (2009); Buchanan (2009); Roehner (2010).

Chapter 2

An Empirical Study of the Tails of Mutual Fund Size

[1] 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 is best 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 much better described by a log-normal than a power law, indicating less concentration than, for example, personal income. The results are highly statistically significant and are consistent across fifteen years. This contradicts a recent theory concerning the origin of the power law tails of the trading volume distribution. Based on the analysis in a companion paper, the log-normality is to be expected, and indicates that the distribution of mutual funds remains perpetually out of equilibrium.

2.1 Introduction

As of 2007 the mutual fund industry controlled 23% of household taxable assets in the United States[2]. In absolute terms this corresponded to 4.4 trillion USD and 24% of U.S. corporate equity holdings. Large players such as institutional investors are known to play an important role in the market (Corsetti et al., 2001). 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? Are there mutual funds that are so large that they are "too big to fail"?

This question is best addressed in terms of the behavior of the upper tail of the mutual

[1]Published work: Yonathan Schwarzkopf and J.Doyne Farmer, Empirical study of the tails of mutual fund size, Phys. Rev. E 81, 066113 (2010)."Copyright (2010) by the American Physical Society."

[2]Data is taken from the Investment Company Institute's 2007 fact book available at www.ici.org.

fund size distribution. The two competing hypotheses usually made in studies of firms are Zipf's law vs. a lognormal. Zipf's law means that the distribution of the size s is a power law with tail exponent $\zeta_s \approx 1$, i.e.

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

Log-normality means that $\log s$ has a normal distribution, i.e. the density function $p_{LN}(s)$ obeys

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

From the point of view of extreme value theory this distinction is critical, since it implies a completely different class of tail behavior[3]. These are both heavy tailed, but Zipf's law is much more heavy tailed. For a log-normal all the moments exist, whereas for Zipf's law none of the moments exist. For Zipf's law an estimator of the mean fails to converge. In practical terms, for mutual funds this would imply that for any sample size N, with significant probability an individual fund can be so large that it is bigger than all other $N - 1$ firms combined. In contrast, for a log-normal, in the limit as $N \to \infty$ the relative size of a single fund becomes negligible.

This question takes on added meaning because the assumption that mutual funds follow Zipf's law has been argued to be responsible for the observed power law distribution of trading volume (Levy et al., 1996; Solomon and Richmond, 2001). Gabaix et al. have also asserted that the mutual fund distribution follows Zipf's law and have used this in a proposed explanation for the distribution of price returns (Gabaix et al., 2003; Gabaix et al., 2006).

We resolve this empirically using the Center for Research in Security Prices (CRSP) dataset and find that the equity fund size distribution is much better described by a log-normal distribution.

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

[3] 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).

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 for individual industries the tail behavior is debated[4]. A large number of stochastic process models have been proposed to explain this[5]. Our results add support to the notion that for single industries the distribution is log-normal.

The log-normality of the distribution of mutual funds is also interesting for what it suggests about the underlying processes that determine mutual fund size. In a companion paper (Schwarzkopf and Farmer, 2010a) we develop a model for the random process of mutual fund entry, exit and growth under the assumption of market efficiency, and show that this gives a good fit to the data studied here. We show that while the steady-state solution is a power law, the timescale for reaching this solution is very slow. Thus given any substantial non-stationarity in the entry and exit processes the distribution will remain in its non-equilibrium log-normal state. See the discussion in Section 2.5.

2.2 Data Set

We analyze the Center for Research in Security Prices (CRSP) Survivor-Bias-Free US Mutual Fund Database[6]. The database is survivor bias free as it contains historical performance data for both active and inactive mutual funds. We study monthly data from 1991 to 2005[7] on all reported equity funds. 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

[4]Some studies have found that the upper tail is a log-normal (Simon, H. A. and Bonini, Charles P., 1958; Stanley et al., 1995; Ijiri and Simon, 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).

[5]For past stochastic models see (Gibrat, 1931; Simon, 1955; Simon, H. A. and Bonini, Charles P., 1958; Mandelbrot, 1963; Ijiri and Simon, 1977; Sutton, 1997; Gabaix et al., 2003; Gabaix et al., 2003).

[6]The US Mutual Fund Database can be purchased from the Center for Research in Security Prices (www.crsp.com).

[7]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.

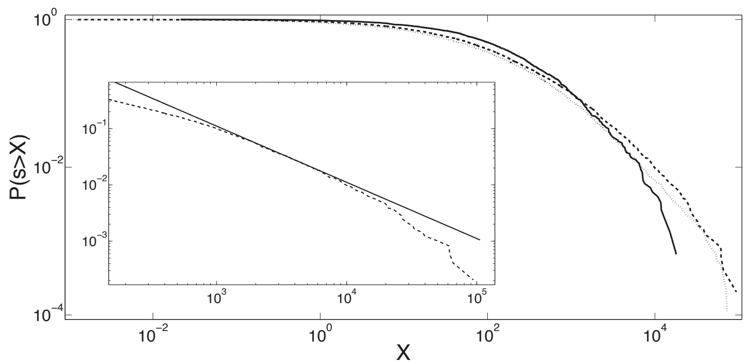


Figure 2.1: The CDF for the mutual fund size s (in millions of 2007 dollars) is plotted with a double logarithmic scale. The cumulative distribution for funds existing at the end of the years 1993, 1998 and 2005 are given by the full, dashed and dotted lines respectively. Inset: The upper tail of the CDF for the mutual funds existing at the end of 1998 (dotted line) is compared to an algebraic relation with exponent -1 (solid line).

for inflation relative to July 2007. Inflation adjustments are based on the Consumer Price Index, published by the BLS.

2.3 Is the tail a power law?

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 2.1, where we plot the cumulative distribution of sizes $P(s > X)$ of mutual fund sizes in three different years.

A visual inspection of the mutual fund size distribution suggests that it does not follow Zipf's law[8]. In the inset of Figure 2.1 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. The main point of this paper is to make more rigorous tests of the

[8] 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.

power law vs. the log-normal hypothesis. These back up the intuitive impression given by this plot, indicating that the data are not well described by 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[9] that the probability density function $p(s)$ is a power law if there exists an s_{min} such that for sizes larger than s_{min}, the functional form of the density $p(s)$ can be written

$$p(s) = \frac{\zeta_s}{s_{min}} \left(\frac{s}{s_{min}} \right)^{-(\zeta_s+1)}, \tag{2.1}$$

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

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

Using this procedure we estimate ζ_s and s_{min} for the years 1991- 2005 as shown in Table 2.1. The values of ζ_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_{min} and ζ_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 2.1. The power law hypothesis is rejected with two standard deviations or more in six of the years and rejected at one standard deviation or

[9] In extreme value theory a power law is defined as any function that in the limit $s \to \infty$ can be written $p(s) = g(s)s^{-(\zeta_s+1)}$ where $g(s)$ is a slowly varying function. This means it satisfies $\lim_{s\to\infty} g(ts)/g(s) = C$ 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.

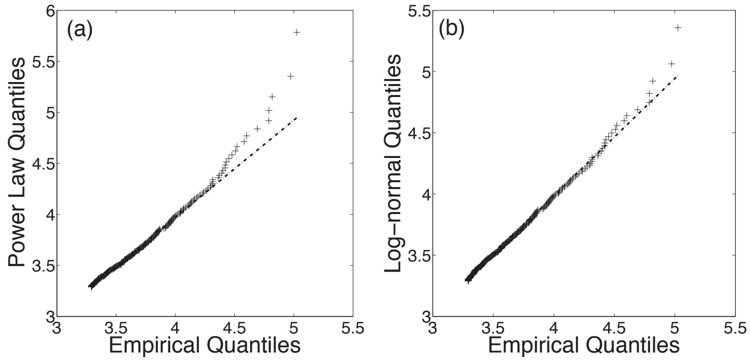


Figure 2.2: A Quantile-Quantile (QQ) plot for the upper tail of the size distribution of equity funds. The quantiles are the base ten logarithm of the fund size, in millions of dollars. The empirical quantiles are calculated from the size distribution of funds existing at the end of the year 1998. The empirical data were truncated from below such that only funds with size $s \geq s_{min}$ were included in the calculation of the quantiles. (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 2.3, yielding $s_{min} = 1945$ and $\alpha = 1.107$. (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, with the same s_{min} as in (a). The log-normal fit for the data was done used the maximum likelihood estimation given s_{min} (2.2) yielding $\mu = 2.34$ and $\sigma = 2.5$.

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 2.1, the total number of equity funds increases roughly linearly in time, and the number in the upper tail N_{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.

variable	91	92	93	94	95	96	97	98	99	00	01	02	03	04	05	mean	std
$\mathcal{R}$	-0.50	-1.35	-1.49	-1.71	-3.29	-18.42	-2.25	-1.29	-6.57	-4.96	-2.63	-2.95	-2.00	-1.05	-0.99	-3.43	4.45
N	372	1069	1509	2194	2699	3300	4253	4885	5363	5914	6607	7102	7794	8457	8845	-	-
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[ω]	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	3.96	0.54
Std[ω]	1.51	1.98	2.09	2.43	2.50	2.46	2.50	2.51	2.46	2.45	2.63	2.42	2.49	2.59	2.50	2.34	0.29
ζ_s	1.33	1.36	1.19	1.15	1.11	0.78	1.08	1.10	0.95	0.97	1.01	1.07	1.07	1.10	1.14	1.09	0.14
s_{min}	955	800	695	708	877	182	1494	1945	1147	903	728	836	868	1085	1383	974	408
N_{tail}	81	129	232	256	280	1067	290	283	557	662	717	494	652	630	550	-	-
p-value	0.58	0.48	0.02	0.45	0.07	0	0.01	0.11	$5\,10^{-4}$	0.04	0.03	0.07	0.08	0.08	0.15	0.15	0.19

Table 2.1: 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.

$\mathcal{R}$ - the base 10 log likelihood ratio of a power law fit relative to a log-normal fit as given by equation (2.3). A negative value of $\mathcal{R}$ 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.

ζ_s - the power law tail exponent (2.1).

s_{min} - the lower tail cutoff (in millions of dollars) above which we fit a power law (2.1).

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

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.

2.4 Is the tail log-normal?

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 2.2 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 (2.1). The lognormal likelihood is defined as $L_{LN} = \prod_{s_j \geq s_{min}} p_{LN}(s_j)$ with the probability density of the lognormal tail given by

$$\begin{aligned} p_{LN}(s) &= \frac{p(s)}{1 - P(s_{min})} \\ &= \frac{\sqrt{2}}{s\sqrt{\pi}\sigma} \left[\operatorname{erfc}\left(\frac{\ln s_{min} - \mu}{\sqrt{2}\sigma} \right) \right]^{-1} \exp\left[-\frac{(\ln s - \mu)^2}{2\sigma^2} \right]. \end{aligned} \tag{2.2}$$

The more probable it is that the empirical sample is drawn from a given distribution, the larger the likelihood for that set of observations. The ratio indicates which distribution

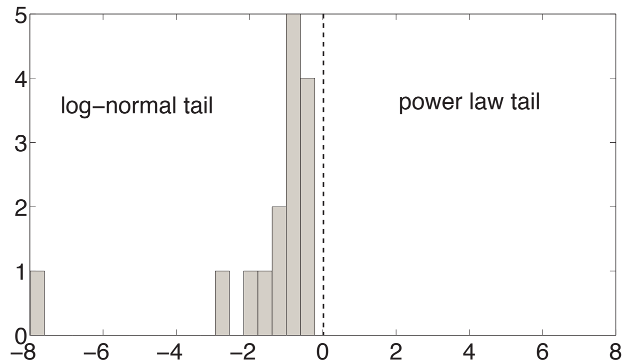


Figure 2.3: A histogram of the base 10 log likelihood ratios $\mathcal{R}$ computed using (2.3) for each of the years 1991 to 2005. A negative log likelihood ratio implies that it is more likely that the empirical distribution is log-normal then a power law. The log likelihood ratio is negative in every year, in several cases strongly so.

the data are more likely drawn from. We define the log likelihood ratio as

$$\mathcal{R} = \ln\left(\frac{L_{PL}}{L_{LN}}\right). \tag{2.3}$$

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 2.3. Using the fit parameters, the log likelihood ratio was computed and the results are summarized graphically in Figure 2.3 and in Table 2.1. 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.

2.5 Implications of log-normality

The log-normal nature of the size distribution has important implications on the role investor behavior plays in the mutual fund industry. Is the size distribution of mutual funds, i.e. the concentration of assets, determined through investor choice or is it just a consequence of the random nature of the market? In a companion paper (Schwarzkopf and Farmer, 2010a) we propose that the size distribution can be explained by a simple random process model. This model, characterizing the entry, exit and growth of mutual funds as a random process, is based on market efficiency, which dictates that fund performance is size independent and fund growth is essentially random. This model provides a good explanation of the concentration of assets, suggesting that other effects, such as transaction costs or the behavioral aspects of investor choice, play a smaller role.

The fact that the fund distribution is a log-normal is interesting because, as we argue in the companion paper, this indicates a very slow convergence toward equilibrium. There we find a time-dependent solution for the underlying random process of mutual fund entry, exit, and growth, and show that the size distribution evolves from a log-normal towards a Zipf power law distribution. However, the relaxation to the steady-state solution is extremely slow, with time scales on the order of a century or more. Given that the mutual fund industry is still young, the distribution remains in its non-equilibrium state as a log-normal. Furthermore, given that the properties of the entry and exit processes are not stable over long periods of time, the non-equilibrium log-normal state will very likely persist indefinitely.

2.6 Conclusions

We have shown in unequivocal terms that the mutual fund size distribution is much closer to a log-normal than to a power law. Thus, while the distribution is concentrated, it is not nearly as concentrated as it might be. Among other things this suggests that that the power law distribution observed for trading volume by Gopikrishnan et al. (2000) cannot be explained based on a power law distribution for funds. The companion paper discussed in the previous section (Schwarzkopf and Farmer, 2010a) constructs a theory that explains the log-normality based on the random nature of the mutual fund entry, exit and growth, and the very long-time scales required for convergence to the steady-state power law solution.

Chapter 3

What Drives Mutual Fund Asset Concentration?

Is the large influence that mutual funds assert on the U.S. financial system spread across many funds, or is it is concentrated in only a few? We argue that the dominant economic factor that determines this is market efficiency, which dictates that fund performance is size independent and fund growth is essentially random. The random process is characterized by entry, exit and growth. We present a new time-dependent solution for the standard equations used in the industrial organization literature and show that relaxation to the steady-state solution is extremely slow. Thus, even if these processes were stationary (which they are not), the steady-state solution, which is a very heavy-tailed power law, is not relevant. The distribution is instead well-approximated by a less heavy-tailed log-normal. We perform an empirical analysis of the growth of mutual funds, propose a new, more accurate size-dependent model, and show that it makes a good prediction of the empirically observed size distribution. While mutual funds are in many respects like other firms, market efficiency introduces effects that make their growth process distinctly different. Our work shows that a simple model based on market efficiency provides a good explanation of the concentration of assets, suggesting that other effects, such as transaction costs or the behavioral aspects of investor choice, play a smaller role.

3.1 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. This raises the question of who has this influence:

[1]Data is taken from the Investment Company Institute's 2007 fact book available at www.ici.org.

Are mutual fund investments concentrated in a few dominant large funds, or spread across many funds of similar size? Do we need to worry that a few funds might become so large that they are "too big to fail"? What are the economic mechanisms that determine the concentration of investment capital in mutual funds?

Large institutional investors are known to play an important role in the market (Corsetti et al., 2001). Gabaix et al. recently hypothesized that the fund size distribution plays a central role in explaining the heavy tails in the distribution of both trading volume and price returns[2]. If their theory is true this would imply that the heavy tails in the distribution of mutual fund size play an important role in determining market risk.

While it is standard in economics to describe distributional inequalities in terms of statistics such as the Gini or Herfindahl indices, as we show in Appendix A, this approach is inadequate to describe the concentration in the tail. Instead, the best way to describe the concentration of assets is in terms of the functional form of the tail. As is well-known in extreme value theory (Embrechts et al., 1997), the key distinction is whether all the moments of the distribution are finite. If the tail is truly concentrated, the tail is a power law, and all the moments above a given threshold, called the tail exponent, are infinite. So, for example, if the tail of the mutual fund size distribution follows Zipf's law as hypothesized by Gabaix et al., i.e. if it were a power law with tail exponent one, this would imply nonexistence of the mean. In this case the sample estimator fails to converge because the tails are so heavy that with significant probability a single fund can be larger than the rest of the sample combined. This is true even in the limit as the sample size goes to infinity. Thus power law tails imply a very high degree of concentration.

Instead, empirical analysis shows that the tail of the mutual fund size distribution is not a power law, and is well-approximated by a lognormal (Schwarzkopf and Farmer, 2010b). Thus, while the distribution is heavy tailed, it is not as heavy tailed as it would be if the distribution were a power law. The key difference is that for a log-normal all of the moments exist.

This naturally leads to the question of what economic factors determine the tail properties of the mutual fund distribution. There are two basic types of explanation. One type

[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 explanation is based on a detailed description of investor choice, and another is based on efficient markets, which predicts that growth should be random, and that the causes can be understood in terms of a simple random process description of entry, exit and growth. Of course market efficiency depends on investor choice, but the key distinction is that the random process approach does not depend on any of the details, but rather only requires that no one can make superior investments based on simple criteria, such as size.

Explanations based on investor choice can in turn be divided into two types: rational and behavioral. For example, Berk and Green [2004] have proposed that investors are rational, making investments based on past performance. Their theory implies that the distribution of fund size is determined by the skill of mutual fund managers and the dependence of transaction costs on size. If we assume, for example, that the transaction cost is a power law (which includes linearity) if the distribution of fund size is log-normal, then it is possible to show that the distribution of mutual fund skill must also be log-normal. Unfortunately, without a method of measuring skill this is difficult to test.

Another type of explanation is behavioral, i.e. that investors are strongly influenced by factors such as advertising, fees, and investment fads[3]. We strongly suspect that this is true, and that they play an important role in determining the size of individual funds. The question we investigate here is not whether such effects exist, but whether they are essential to explain the form of the distribution.

The alternative is that the details of investor choice don't matter, and that the distribution of fund size is driven by market efficiency, which dictates an approach based on the random process of entry, exit and growth. The random process approach was originally pioneered as an explanation for firm size by Gibrat, Simon and Mandelbrot, and is popular in the industrial organization literature[4]. The basic idea is that while details of investor choice are surely important in determining the size of *individual* funds, the details may average out or be treatable as noise, so that in aggregate they do not matter in shaping the overall size distribution.

On the face of it, however, there seems to be a serious problem with this approach. Under simple assumptions about the entry, exit and growth of fund size, Gabaix et al.

[3]Barber et al. (2005) have found that investors flows are correlated to marketing and advertising while they are not correlated to the expense ratio.

[4]For past stochastic models see (Gibrat, 1931; Simon, 1955; Simon, H. A. and Bonini, Charles P., 1958; Mandelbrot, 1963; Ijiri and Simon, 1977; Sutton, 1997; Gabaix et al., 2003; Gabaix et al., 2003).

(2003) showed that the steady state solution is a power law; a similar argument is described in Montroll and Shlesinger (1982) and Reed (2001)[5]. As already mentioned, however, the upper tail of the empirical distribution is a log-normal, not a power law. Thus there would seem to be a contradiction. Apparently either the correct random process is more complicated, or this whole line of attack fails.

We show here that the central problem comes from considering only the steady state (i.e. infinite time) solution. We study the same equations considered by Gabaix et al. and Reed, but we find a more general time-dependent solution, and show that the time required to reach steady state is very long. The mutual fund industry is rapidly growing and, even if the growth process had been stationary over the last few decades, not enough time has elapsed to reach the stationary solution for the fund size distribution. In the meantime the solution is well approximated by a log-normal. This qualitative conclusion is very robust under variations of the assumptions. In contrast to the hypothesis of Berk and Green, it does not depend on details such as the distribution of investor skill – the log-normal property emerges automatically from market efficiency and the random multiplicative nature of fund growth.

To test our conjectures more quantitatively we study the empirical properties of entry, exit and growth of mutual funds, propose a more accurate model than those previously studied, and show it makes a good prediction of the empirically observed fund size distribution. The model differs from previous models in that it incorporates the fact that the relative growth rate of funds slows down as they get bigger[6]. This makes the time needed to approach the steady state solution even longer: Whereas the relaxation time for the size-independent diffusion model is several decades, for the more accurate size-dependent model it is more than a century.

Market efficiency is the key economic principle that makes the random process model work, and dictates many of its properties. It enters the story in several ways. (1) The fact that stock market returns are essentially random implies that growth fluctuations are random, for two reasons: (a) Without inflows and outflows, under the principle that past returns are not indicative of future returns, fund growth is random. (b) Although investors

[5]For a review on similar generative models see Mitzenmacher (2003).

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

chase past returns, since what they are chasing is random, fund growth due to inflow and outflow is random on sufficiently long time scales. (2) Efficiency dictates that mutual fund performance must be independent of size. Thus as mutual funds randomly diffuse through the size space, there is no pressure pushing them toward a particular size. (3) Efficiency, together with the empirical fact that the relative importance of fund inflows and outflows diminishes as funds get bigger, implies that the mean growth rate and the growth diffusion approach a constant in the large size limit. As we show, this shapes the long-term properties of the size distribution. All of these points are explained in more detail in Section 3.4.3.

Market efficiency makes mutual funds unusual relative to most other types of firms. For most firms, in the large size limit the mean and standard deviation of the growth rate are empirically observed to decay to zero. For mutual funds, in contrast, due to market efficiency they both approach a positive limit. This potentially affects the long-term behavior: Most firms approach a solution that is thinner than a log-normal, i.e under stationary growth conditions their tails are getting thinner with time, whereas mutual funds approach a power law, so their tails are getting fatter with time. Nonetheless, as we have already mentioned, even under stationary growth conditions the approach to steady-state takes so long that this is a moot point.

At a broader level our work here shows how the non-stationarity of market conditions can prevent convergence to an "equilibrium" solution. Nonetheless, even under stationary conditions the random process model usefully describes the time-dependent relationships between entry, exit and growth phenomena on one hand and size on the other hand. While we cannot show that the random process model is the only possible explanation, we do show that it provides a good explanation[7]. The conditions for this are robust, depending only on market efficiency, without the stronger requirements of perfect rationality, or the complications of mapping out the idiosyncrasies of human behavior.

The paper is organized as follows. In Section 3.2 we develop the standard exit and entry model. Section 3.2.1 presents the time-dependent solution for the number of funds, Section 3.2.2 presents the time-dependent solution for the size distribution, and Section 3.2.3 intro-

[7]While variations in the assumptions about the random process preserve certain qualitative conclusions, such as the log-normal character of the upper tail, we found that getting a good fit to the data requires a reasonable degree of fidelity in the modeling process. The size-dependent nature of the diffusion process, for example, is quite important.

duces a size-dependent model. Section 3.3 describes our data. In Section 3.4 we perform an empirical analysis to justify our assumptions and to calibrate the model. In Section 3.5 we present simulation results of the proposed model and compare them to the empirical data. Finally Section 3.6 presents our conclusions.

3.2 Model

Our central thesis in this paper is that due to market efficiency 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 and the exit of existing funds. In this section we introduce the standard diffusion model and derive a time-dependent solution for the special case when the diffusion process has constant mean and variance. We then make a proposal for how to model the more general case where the mean and variance depend on size.

The aim of the model we develop here 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 size s at time t, where $\omega = \log s$. The size distribution can be written as

$$p(\omega, t) = \frac{n(\omega, t)}{N(t)}, \tag{3.1}$$

where $n(\omega, t)$ is the number of funds at time with logarithmic size ω and $N(t) = \int n(\omega, t)\mathrm{d}\omega$ 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 $n(\omega, t)$.

3.2.1 Dynamics of the total number of funds

As we will argue in Section 3.4, the total number of funds as a function of time can be modeled as

$$\frac{dN}{dt} = \nu - \lambda N \tag{3.2}$$

where ν is the rate of creating new funds and λ is the exit rate of existing funds. Under the assumption that ν and λ are constants this has the solution

$$N(t) = \frac{\nu}{\lambda}\left(1 - \mathrm{e}^{-\lambda t}\right)\theta(t), \tag{3.3}$$

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 λ, with a characteristic timescale $1/\lambda$. For example, for $\lambda \approx 0.09$, as estimated in Section 3.4, the timescale for $N(t)$ to reach its steady state is only roughly a decade. An examination of Table 3.3 makes it clear, however, that $\nu =$ constant is not a very good approximation. Nonetheless, if we crudely use the mean creation rate $\nu \approx 900$ from Table 3.3 and the fund exit rate $\lambda \approx 0.09$ estimated in Section 3.4, 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 $N(t)$.

3.2.2 Solution for the number density $n(\omega, t)$

We define and solve the time evolution equation for the number density $n(\omega, t)$. The empirical justification for the hypotheses of the model will be given in Section 3.4. The hypotheses are:

- The entry process is a Poisson process with rate ν, such that at time t a new fund enters the industry with a probability $\nu \mathrm{d}t$ and (log) size ω drawn from a distribution $f(\omega, t)$. We approximate the entry size distribution as a log-normal distribution in the fund size s, that is a normal distribution in ω given by

$$f(\omega, t) = \frac{1}{\sqrt{\pi \sigma_\omega^2}} \exp\left(-\frac{(\omega - \omega_0)^2}{\sigma_\omega^2}\right) \theta(t - t_0), \tag{3.4}$$

  where ω_0 is the mean log size of new funds and σ_ω^2 is its variance. $\theta(t - t_0)$ is a unit step function ensuring no funds funds enter the industry before the initial time t_0.

- The exit process is a Poisson process such that at any time time t a fund exits the industry with a size independent probability $\lambda \mathrm{d}t$.

- The size change is approximated as a (log) Brownian motion with a size dependent drift and diffusion term

$$\mathrm{d}\omega = \mu(\omega)\mathrm{d}t + \sigma(\omega)\mathrm{d}W, \tag{3.5}$$

where dW is an i.i.d random variable drawn from a zero mean and unit variance normal distribution.

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

$$\frac{\partial}{\partial t}n(\omega,t) = \nu f(\omega,t) - \lambda n(\omega,t) - \frac{\partial}{\partial\omega}[\mu(\omega)n(\omega,t)] + \frac{\partial^2}{\partial\omega^2}[D(\omega)n(\omega,t)], \tag{3.6}$$

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 a existing funds.

Approximate solution for large funds

To finish the model it is necessary to specify the functions $\mu(\omega)$ and $D(\omega)$. It is convenient to define the relative change in a fund's size $\Delta_s(t)$ as

$$\Delta_s(t) = \frac{s(t+1) - s(t)}{s(t)}, \tag{3.7}$$

such that drift and diffusion parameters in our model are given by

$$\mu(\omega) = \mathrm{E}[\log(1+\Delta_s)] \qquad D(\omega) = \frac{1}{2}\mathrm{Var}[\log(1+\Delta_s)].$$

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

$$\Delta_s(t) = \Delta_f(t) + \Delta_r(t). \tag{3.8}$$

The return Δ_r represents the return of the fund to its investors, defined as

$$\Delta_r(t) = \frac{NAV(t+1) - NAV(t)}{NAV(t)}, \tag{3.9}$$

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

$$\Delta_f(t) = \frac{s(t+1) - [1+\Delta_r(t)]s(t)}{s(t)}. \tag{3.10}$$

In Section 3.4 we will demonstrate empirically that the returns Δ_r are independent of size, as they must be for market efficiency. In contrast the money flux Δ_f decreases monotonically with size. In the large size limit the returns Δ_r dominate, and thus it is reasonable to treat $\mu(s)$ as a constant, $\mu = \mu_\infty$. Market efficiency also implies that in the large size limit the standard deviation $\sigma(s)$ is a constant, i.e. $\sigma = \sigma_\infty$. Otherwise investors would be able to improve their risk adjusted returns by simply investing in larger funds.

With these approximations the evolution equation becomes

$$\frac{\partial}{\partial t} n(\omega,t) = \nu f(\omega,t) - \lambda n(\omega,t) - \mu \frac{\partial}{\partial \omega} n(\omega,t) + D \frac{\partial^2}{\partial \omega^2} n(\omega,t), \tag{3.11}$$

In this and subsequent equations, to keep things simple we use the notation $D = \sigma_\infty^2/2$ and $\mu = \mu_\infty$.

The exit process is particularly important, since it is responsible for thickening 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 small funds than large funds, more small funds exit the industry. This results in relatively more large funds, making the distribution heavy-tailed. 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 important for determining the shape of the distribution, and influences only the total number of funds N. 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

$$n(\omega,t) = \nu \int_{-\infty}^{\infty} \int_0^t \exp^{-\lambda t'} \frac{1}{\sqrt{4\pi D t'}} \exp\left[-\frac{(\omega - \omega' - \mu t')^2}{4Dt'} \right] f(\omega', t-t')\, \mathrm{d}t' \mathrm{d}\omega'. \tag{3.12}$$

Stated in words, a fund of size ω' 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 ω at time t with a probability according to (3.23).

If funds enter the industry with a constant rate ν beginning at $t = 0$, with a log-normal entry size distribution $f(\omega,t)$ centered around ω_0 with width σ_ω as given by (3.4), the size density can be shown to be

$$\begin{aligned} n(\omega,t) &= \frac{\nu\mu}{4\sqrt{\gamma}D}\exp\left[(\gamma+\frac{1}{4})\frac{\sigma_\omega^2}{2}-\sqrt{\gamma}\left|\frac{\sigma_\omega^2}{2}+\frac{\mu}{D}(\omega-\omega_0)\right|+\frac{\mu}{2D}(\omega-\omega_0)\right] \\ &\quad\times\left(A+\exp\left[\sqrt{\gamma}|\sigma_\omega^2+2\frac{\mu}{D}(\omega-\omega_0)\,|\right]B\right). \end{aligned} \tag{3.13}$$

The parameters A, B and γ are defined as

$$\gamma=\sqrt{\frac{1}{4}+\frac{\lambda D}{\mu^2}}, \tag{3.14}$$

$$\begin{aligned} A &= \mathrm{Erf}\left[\frac{\left|\frac{\sigma_\omega^2}{2}+\frac{\mu}{D}(\omega-\omega_0)\right|-\sqrt{\gamma}\sigma_\omega^2}{\sqrt{2}\sigma_\omega}\right] \\ &\quad-\mathrm{Erf}\left[\frac{\left|\frac{\sigma_\omega^2}{2}+\frac{\mu}{D}(\omega-\omega_0)\right|-\sqrt{\gamma}\left(\sigma_\omega^2+2\frac{\mu^2}{D}t\right)}{\sqrt{2}\sqrt{\sigma_\omega^2+2\frac{\mu^2}{D}t}}\right] \end{aligned} \tag{3.15}$$

and

$$\begin{aligned} B &= \mathrm{Erf}\left[\frac{\sqrt{\gamma}\left(\frac{\sigma_\omega^2}{2}+\frac{\mu^2}{D}t\right)+|\frac{\sigma_\omega^2}{2}+\frac{\mu}{D}(\omega-\omega_0)\,|}{\sqrt{2}\sqrt{\sigma_\omega^2+2\frac{\mu^2}{D}t}}\right] \\ &\quad-\mathrm{Erf}\left[\frac{\sqrt{\gamma}\sigma_\omega^2+\left|\frac{\sigma_\omega^2}{2}+\frac{\mu}{D}(\omega-\omega_0)\right|}{\sqrt{2}\sigma_\omega}\right], \end{aligned} \tag{3.16}$$

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

Approximating the distribution of entering funds as having zero width simplifies the solution. Let us define a large fund as one with $\omega \gg \omega_0$, where ω_0 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, i.e. all new funds have the same size ω_0. The number density is then given by

$$\begin{aligned} n(\omega,t) &= \frac{\nu D}{4\sqrt{\gamma}\mu^2}e^{\frac{1}{2}\frac{\mu}{D}(\omega-\omega_0)}\left[e^{-\sqrt{\gamma}\frac{\mu}{D}|\omega-\omega_0|}\left(1+\mathrm{erf}\left[\sqrt{\frac{\gamma\mu^2 t}{D}}-\frac{|\omega-\omega_0|}{2\sqrt{Dt}}\right]\right)\right. \\ &\quad\left.-e^{\sqrt{\gamma}\frac{\mu}{D}|\omega-\omega_0|}\left(1-\mathrm{erf}\left[\mu\sqrt{\frac{t}{D}}(\frac{1}{2}+\sqrt{\gamma})\right]\right)\right]. \end{aligned} \tag{3.17}$$

Since $\gamma > 1/4$ (3.14), the density vanishes for both $\omega \to \infty$ and $\omega \to -\infty$.

Steady state solution for large funds

The steady state solution for large times is achieved by taking the $t \to \infty$ limit of (3.17), which gives

$$n(\omega) = \frac{\nu}{2\mu\sqrt{\gamma}} \exp \frac{\mu}{D} \left(\frac{\omega - \omega_0}{2} - \sqrt{\gamma}|\omega - \omega_0| \right). \tag{3.18}$$

Since the log size density (3.18) 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[8] ζ_s, i.e.

$$P(s > X) \sim X^{-\zeta_s}. \tag{3.19}$$

Substituting for the parameter γ using Eq. (3.14) for the upper tail exponent yields

$$\zeta_s = \frac{-\mu + \sqrt{\mu^2 + 4D\lambda}}{2D}. \tag{3.20}$$

Note that this does not depend on the creation rate ν. Using the average parameter values in Table 3.4.3 the asymptotic exponent has the value

$$\zeta_s = 1.2 \pm 0.6. \tag{3.21}$$

This suggests that if the distribution reaches steady state it will follow Zipf's law, which is just the statement that it will be a power law with $\zeta_s \approx 1$. As discussed in the introduction, this creates a puzzle, as the empirical distribution is clearly log-normal (Schwarzkopf and Farmer, 2010b).

Timescale to reach steady state

Since we have a time dependent solution we can easily estimate of the timescale to reach steady state. The time dependence in Eq. 3.17 is contained in the arguments of the error function terms on the right. When these arguments become large, say larger than 3, the

[8] 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 ζ_s.

solution is roughly time independent, and can be written as

$$t > \frac{9D}{4\gamma\mu^2}\left(1+\sqrt{1+\frac{2}{9}\frac{\sqrt{\gamma\mu^2}}{D}\left|\omega-\omega_0\right|}\right)^2. \tag{3.22}$$

Using the average values in Table 3.4.3 in units of months $\mu = \mu_\infty \approx 0.005$, $D = \sigma_\infty^2/2$ and $\sigma_\infty \approx 0.05$. This gives

$$t > 180\left(1+\sqrt{1+0.7\left|\omega-\omega_0\right|}\right)^2,$$

where the time is in months. Plugging in some numbers from Table 3.4.3 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 empirical observation that there seems to be no significant fattening of the tail in the fifteen years from 1991 - 2005. Note that the time required for the distribution $n(\omega, t)$ to reach steady state for large values of ω is much greater than that for the total number of funds $N(t)$ to become constant.

During the transient phase the solution remains approximately log-normal for a long time. If funds only change in size and no funds enter or exit, then the resulting distribution is normal

$$\tilde{n}(w,t) = \frac{1}{\sqrt{4\pi Dt}}\exp\left[-\frac{(\omega-\mu t)^2}{4Dt}\right], \tag{3.23}$$

which corresponds to a size distribution $p(s)$ with a lognormal upper tail. While the exit process acts quickly in changing the total number of funds, it acts slowly in changing the shape. This is the key reason why the distribution remains approximately log-normal for so long.

3.2.3 A better model of size dependence

The mean rate of growth and diffusion are in general size dependent. We hypothesize that the mean growth rate $\mu(s)$ and the standard deviation $\sigma(s)$ are the sum of a power law and a constant, of the form

$$\begin{aligned} \sigma_s(s) &= \sigma_0 s^{-\beta} + \sigma_\infty \qquad (3.24)\\ \mu_s(s) &= \mu_0 s^{-\alpha} + \mu_\infty. \end{aligned}$$

The constant terms come from mutual fund returns (neglecting inflow or outflow of funds), and must be constant due to market efficiency, as explained in more detail in Section 3.4.3. The power law terms, in contrast, are due to the flow of funds in and out of the market. There is a substantial literature of proposed theories for this, including ours[9]. We present the empirical evidence for the power law hypothesis and explain the role of efficiency in more detail in Section 3.4.3.

The functional form given above for the size dependence can be used to make a more accurate diffusion model. 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 already mentioned, due to market efficiency $E[\Delta_r(s)]$ must be independent of s, and since $E[\Delta_f(s)]$ is a decreasing function of s, for large s $\mu(s) = E[\Delta_r(s)] + E[\Delta_f(s)] = \mu_\infty > 0$. This distinguishes mutual funds from other types of firms, which are typically observed empirically to have $\mu_\infty = \sigma_\infty = 0$ (Stanley et al., 1996; Matia et al., 2004). 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[10].

3.3 Data Set

We test our model against the CRSP Survivor-Bias-Free US Mutual Fund Database. Because we have daily data for each mutual fund, this database enables us to investigate the

[9]There has been a significant body of work attempting to explain the heavy tails in the growth rate of firms and the associated size dependence in the diffusion rate. See (Amaral et al., 1997a; Buldyrev et al., 1997; Amaral et al., 1998; De Fabritiis et al., 2003; Matia et al., 2004; Bottazzi, 2001; Sutton, 2001; Wyart and Bouchaud, 2003; Bottazzi and Secchi, 2003b, 2005; Fu et al., 2005; Riccaboni et al., 2008; Podobnik et al., 2008). Our theory argues for an additive replication model, and produces predictions that fit the data extremely well for a diverse set of different phenomena, including mutual funds (Schwarzkopf et al., 2010). We argue that the fundamental reason for the power tails is the influence network of investors.

[10]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 λ 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 effect.

mechanism of fund entry, exit and growth to calibrate and test our model[11]. We study the data from 1991 to 2005[12]. 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. In Table 3.3 we provide summary statistics of the data set and as seen there the total number of equity funds increases roughly linearly in time, and the number of funds in the upper tail N_{tail} also increases.

3.4 Empirical investigation of size dynamics

In this section we empirically investigate the processes of entry, exit and growth, providing empirical justification and calibration of the model described in Section 3.2.

3.4.1 Fund entry

We begin by examining the entry of new funds. We investigate both the number $N_{\mathtt{enter}}(t)$ and size s of funds entering each year. We perform a linear regression of $N_{\mathtt{enter}}(t)$ 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, justifying the approximation of entry as a Poisson process with a constant rate ν, independent of $N(t)$.

The size of entering funds is more complicated. In Figure 3.1 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

[11] Note that we treat mergers as the dissolution of both original firms followed by the creation of a new (generally larger) firm. This increases the size of entering firms but does not make a significant difference in our conclusions.

[12] 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.

[13] When discussing the average size one must account for the difference between the average log size and the average size: Due to the heavy tails the difference is striking. The average entry log size $\mathrm{E}[\omega_c] \approx 0$, corresponding to a fund of size one million, while if we average over the entry sizes $\mathrm{E}[s_c] = \mathrm{E}[\mathrm{e}^{\omega_c}]$, we get

variable	91	92	93	94	95	96	97	98	99	00	01	02	03	04	05	mean	std
N	372	1069	1509	2194	2699	3300	4253	4885	5363	5914	6607	7102	7794	8457	8845	-	-
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[ω]	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	3.96	0.54
Std[ω]	1.51	1.98	2.09	2.43	2.50	2.46	2.50	2.51	2.46	2.45	2.63	2.42	2.49	2.59	2.50	2.34	0.29
$\mu(10^{-3})$	26	42	81	46	72	67	58	39	39	20	-3	-10	50	18	30.5	38.5	26
$\sigma(10^{-1})$	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
$N_{\texttt{exit}}$	0	41	45	61	139	115	169	269	308	482	427	660	703	675	626	-	-
$N_{\texttt{enter}}$	185	338	581	783	759	885	1216	1342	1182	1363	1088	1063	1056	796	732	891	346

Table 3.1: Summary statistics and 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) were calculated for each year (columns). The mean and standard deviation are evaluated across the different years.

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.

μ - the drift term for the geometric random walk (3.5), computed for monthly changes.

σ - the standard deviation of the mean zero Wiener process (3.5), computed for monthly changes.

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

$N_{\texttt{enter}}$ - the number of new equity funds entering the industry in each year.

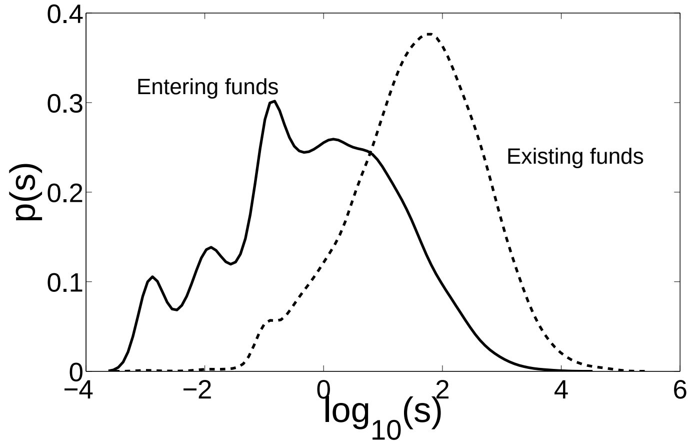


Figure 3.1: The probability density for the size s of entering funds in millions of dollars (solid line) compared to that of all funds (dashed line) including all data for the years 1991 to 2005. The densities were estimated using a Gaussian kernel smoothing technique.

smaller than that of existing funds, making it clear that the typical surviving fund grows significantly after it enters. It is clear that the distribution of entering funds is not important in determining the upper tails[14]. The value of the mean log size and its variance are calculated from the data for each period as summarized in Table 3.4.3.

Thus the empirical data justifies the approximation of entry as a Poisson process in which an average of ν funds enter per month, with the size of each fund drawn from a distribution $f(\omega, t)$.

3.4.2 Fund exit

Unlike entry, fund exit is of critical importance in determining the long-run properties of the fund size distribution. In Figure 3.2 we plot the number of exiting funds $N_{\texttt{exit}}(t)$ as a function of the total number of funds existing in the previous year, $N(t-1)$. There is a good fit to a line of slope λ, which on an annual time scale is $\lambda = 0.092 \pm 0.030$. This justifies our assumption that fund exit is a Poisson process with constant rate λ.

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 3.3.

[14]In Section 3.2.2 we showed that the entry process is not important as long as the tails of the entry distribution f are sufficiently thin. We compared the empirical f to a log-normal and found that the tails are substantially thinner.

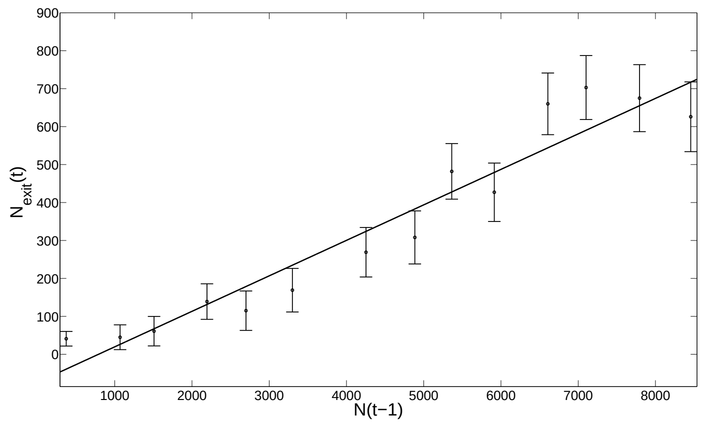


Figure 3.2: The number of equity funds exiting the industry $N_{\texttt{exit}}(t)$ in the year t as a function of the total number of funds existing in the previous year, $N(t-1)$. The plot is compared to a linear regression (full line). The error bars are calculated for each bin under a Poisson process assumption, and correspond to the square root of the average number of funds exiting the industry in that year.

3.4.3 Fund growth

We first test the i.i.d and normality assumptions of the diffusion growth model, and then test to demonstrate the size dependence of the growth process that we proposed in Section 3.2.3. We also discuss the diverse roles that efficiency plays in shaping the random process for firm growth in more detail.

Justification for the diffusion model

In the absence of entry or exit we have approximated the growth of existing funds as a multiplicative Gibrat-like process[15] satisfying a random walk in the log size ω. This implicitly assumes that Δ_s is an i.i.d normal random variable.

The assumption of independence is justified by market efficiency, which requires that the returns Δ_r of a given fund should be random (Bollen and Busse, 2005; Carhart, 1997). Under the decomposition of the total growth as $\Delta_s = \Delta_r + \Delta_f$, as demonstrated in the next

[15] 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.

sub-section, in the large size limit the returns Δ_r dominate, so under market efficiency the i.i.d. assumption is automatically valid.

This is not so obvious for smaller size firms, where the money flux Δ_f dominates the total growth Δ_s. It is well known that investors chase past performance[16]. Even though the past performance they are chasing is random, if they track a sufficiently long history of past returns, this can induce correlations. This causes correlations in the money flux Δ_f, which in turn induces correlations in the total size change Δ_s.

To test whether such correlations are strong enough to cause problems with the random process hypothesis, we perform cross-sectional regressions of the form

$$\Delta_f(t) = \beta + \beta_1 \Delta_r(t-1) + \beta_2 \Delta_r(t-2) + \ldots + \beta_6 \Delta_r(t-6) + \xi(t), \tag{3.25}$$

where $\xi(t)$ is a noise term. The results are extremely noisy; for example, when we perform separate regressions in five different periods, eight of the thirty possible coefficients β_i shown in Table 3.4.3 are negative and only two of them are significant at the two standard deviation level. We also perform direct tests of the correlations in Δ_f and we find that they are small. This justifies our use of the i.i.d. hypothesis.

The normality assumption is also not strictly true. Here we are saved by the fact that the money flux Δ_f is defined in terms of a logarithm, and while it has heavy tails, they are not sufficiently heavy to prevent it from converging to a normal. We have explicitly verified this by tracking a group of funds in a given size range over time and demonstrating that normality is reached within 5 months. Thus even though the normality assumption is not true on short timescales it rapidly becomes valid on longer timescales.

Size dependence of the growth process

We now test the model for the size dependence of the growth process proposed in Section 3.2.3. We also discuss the crucial role of the decomposition into returns and money flux in determining the size dependence.

[16] 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).

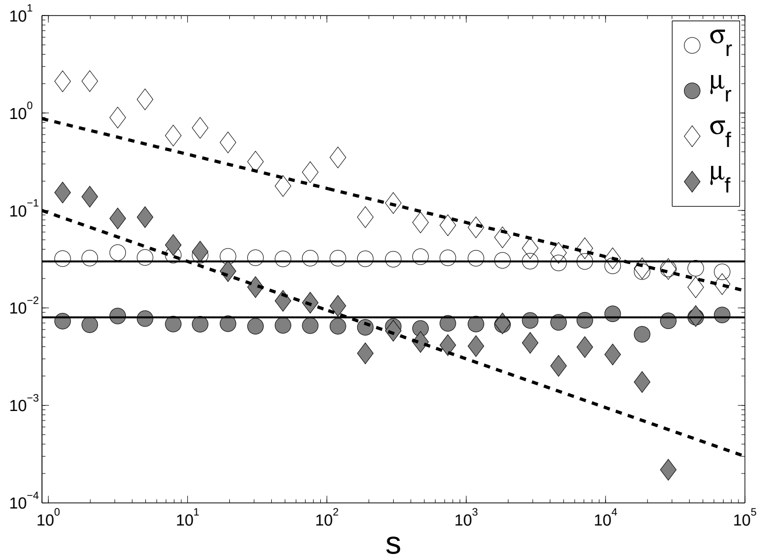


Figure 3.3: A summary of the size dependence of mutual fund growth. The average mean μ_r and volatility σ_r of fund returns, as well as the average μ_f and volatility σ_f of money flux (i.e. the flow of money in and out of funds), are plotted as a function of the fund size (in millions) for the year 2005 (see Eqs. (3.7 - 3.10)). The data are binned based on size, using bins with exponentially increasing size; we use monthly units. The average monthly return μ_r is compared to a constant return of 0.008 and the monthly volatility σ_r is compared to 0.03. The average monthly flux μ_f is compared to a line of slope of -0.5 and the money flux volatility σ_f is compared to a line of slope -0.35. Thus absent any flow of money in or out of funds, performance is independent of size, as dictated by market efficiency. In contrast, both the mean and the standard deviation of the money flows of funds decrease roughly as a power law as a function of size.

date	12/2005	9/2005	6/2005	3/2005	12/2004
β_1	0.10 ± 0.16	0.40 ± 0.98	0.27 ± 0.68	1.17 ± 4.68	-0.23 ± 1.24
β_2	0.14 ± 0.27	0.36 ± 1.20	0.48 ± 0.83	-0.79 ± 3.13	-0.65 ± 2.31
β_3	0.28 ± 0.45	0.01 ± 1.07	0.33 ± 0.83	1.79 ± 3.24	0.60 ± 2.57
β_4	0.56 ± 0.40	-0.28 ± 0.85	0.24 ± 1.27	-0.28 ± 1.65	0.44 ± 2.32
β_5	0.24 ± 0.43	-0.25 ± 1.13	0.21 ± 0.90	-0.24 ± 2.95	0.43 ± 2.49
β_6	0.48 ± 0.38	-0.02 ± 1.03	0.30 ± 0.92	1.27 ± 3.50	0.31 ± 2.09
β	-0.02 ± 0.02	0.03 ± 0.05	0.01 ± 0.05	0.14 ± 0.21	0.06 ± 0.15

Table 3.2: Cross-sectional regression coefficients of the monthly fund flow, computed for several months, against the performance in past months, as indicated in Eq. 3.25. The regression was computed cross-sectionally using data for 6189 equity funds. For example the entry for β_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 are 95% confidence intervals.

Figure 3.3 gives an overview of the size dependence for both the returns Δ_r and the money flux Δ_f. The two behave very differently. The returns Δ_r are essentially independent of size[17]. This is expected based on market efficiency, as otherwise one could obtain superior performance simply by investing in larger or smaller funds (Malkiel, 1995b). 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 $\sigma_r = \mathrm{Var}[\Delta_r]^{1/2}$ 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 flux Δ_f decreases with size. Both the mean money flux $\mu_f = E[\Delta_f]$ and its standard deviation $\sigma_f = \mathrm{Var}[\Delta_f]^{1/2}$ 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). As already discussed in footnote 9, there is a large body of theory attempting to explain this (and we believe our own theory presented elsewhere provides the correct explanation (Schwarzkopf et al., 2010)).

[17]The independence of the return Δ_r on size is verified by performing a linear regression of μ_r vs. s for the year 2005, which results in an intercept $\beta = 6.7 \pm 0.2 \times 10^{-3}$ and a slope $\alpha = 0.5 \pm 8.5 \times 10^{-8}$. This result implies a size independent average monthly return of 0.67%.

variable	1991- 1998	1991- 2005
ω_0	0.14	-0.37
σ_ω	3.02	3.16
σ_0	0.35 ± 0.02	0.30 ± 0.02
β	0.31 ± 0.03	0.27 ± 0.02
σ_∞	0.05 ± 0.01	0.05 ± 0.01
R^2	0.93	0.96
μ_0	0.15 ± 0.01	0.08 ± 0.05
α	0.48 ± 0.03	0.52 ± 0.04
μ_∞	0.002 ± 0.008	0.004 ± 0.001
R^2	0.98	0.97

Table 3.3: Model parameters as measured from the data in different time periods. ω_0 and σ_ω^2 are the mean and variance of the average (log) size of new funds described in (3.4). σ_0, β and σ_∞ are the parameters for the size dependent diffusion and μ_0, α and μ_∞ are the parameters of the average growth rate (3.24). The confidence intervals are 95% under the assumption of standard errors. The adjusted R^2 is given for the fits for each period. The time intervals were chosen to match the results shown in Fig. 3.5.

As explained in Section 3.2.3, the steady state solution is qualitatively different depending on whether the parameters μ_∞ and σ_∞ in Eq. 3.24 are positive. As can be seen from the fit parameters in Table 3.4.3, based on data for Δ_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 Δ_s can be decomposed as $\Delta_s = \Delta_r + \Delta_f$, efficiency dictates that Δ_r is independent of size, and since $E[\Delta_r] > 0$, we are confident that neither μ_∞ nor σ_∞ are zero.

As we showed in Table 3.4.3, the correlation between the returns Δ_r and the money flux Δ_f is small. This implies that the standard deviations can be written as a simple sum. Since Δ_r is independent of size and both the mean and standard deviation of Δ_f are power laws, this indicates that Eq. 3.2.3 is a good approximation, and that μ_∞ and σ_∞ are both greater than zero. As illustrated in Figure 3.4, these functional forms fit the data reasonably well, with only slight variations of parameters in different periods, as shown in

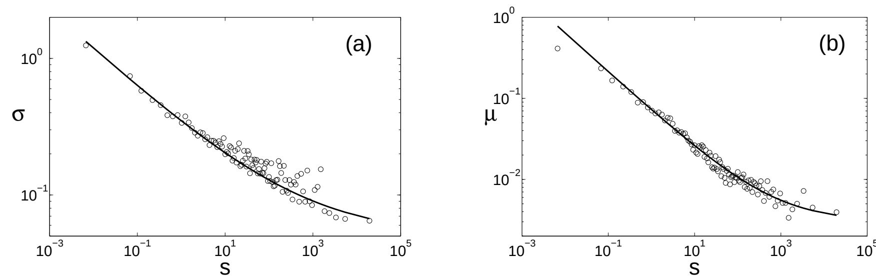


Figure 3.4: An illustration that the empirical power law-based model provides a good fit to the distribution of mutual funds. (a) The standard deviation σ of the logarithmic size change $\Delta_s = \Delta(\log s)$ of an equity fund as a function of the fund size s (in millions of dollars). (b) The mean μ of $\Delta_s = \Delta(\log s)$ 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. μ is the mean in each bin and σ 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 (3.24) in Figures (a) and (b) respectively.

Table 3.4.3.

3.5 Testing the predictions of the model

In this section we use our calibrated model of the entry, exit and size-dependent growth processes to simulate the evolution of the firm size distribution through time. We are forced to use a simulation since, once we include the size dependence of the diffusion and drift terms as given in equation (3.24), we are unable to find an analytic solution for the general diffusion equation (Eq. (3.6)). The analytic solution of the size independent case (Eq. (3.17)) gives the correct qualitative behavior, but the match is much better once one includes the size dependence.

The simulation 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. It is therefore necessary to take 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 distribution for fund size at the end of each period. In each case we assume the size

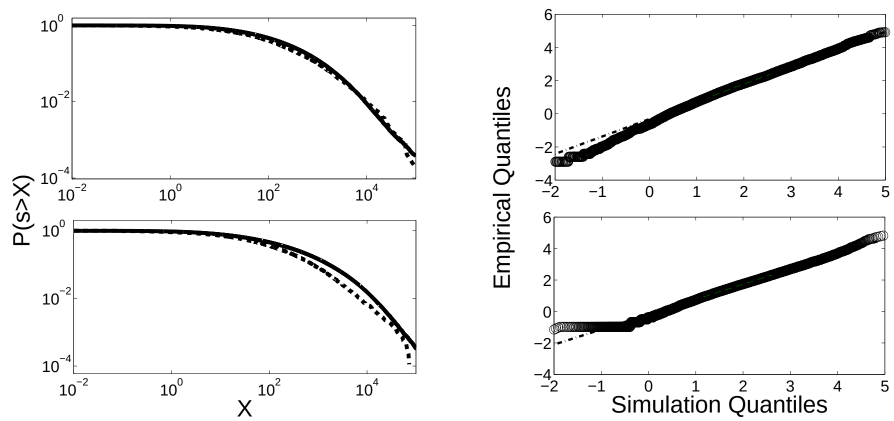


Figure 3.5: The model is compared to the empirical distribution at different time horizons. The left column compares CDFs from the simulation (full line) to the empirical data (dashed line). The right column is a QQ-plot comparing the two distributions. In each case the simulation begins in 1991 and is based on the parameters in Table 3.4.3. The first row corresponds to the years 1991-1998 and the second row to the years 1991-2005 (in each case we use the data at the end of the quoted year).

distribution for injecting funds is log-normal, as discussed in Section 3.4.1.

To compare our predictions to the empirical data we measure the parameters for fund entry, exit and growth using data from the same period as the simulation, summarized in Table 3.4.3. A key point is that we are not fitting these parameters on the target data for fund size[18], but rather are fitting them on the individual entry, 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 3.5 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. Note, that our simulations make it clear that the fluctuations in the tail are substantial. The deviations between the prediction and the data are thus very likely due to chance – many of the individual runs of the simulation deviate

[18]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 the random model for the entry, exit and growth processes can explain the distribution of fund sizes.

from the mean of the 1000 simulations more than the empirical data does.

3.6 Conclusions

We have argued that the mutual fund size distribution is driven by market efficiency, which gives rise to a random growth process. The essential elements of the growth process are multiplicative random changes in the size 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 small 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, and the exit process has not had a sufficient time to act. In the future, if the conditions for fund growth and exit were to remain stationary for more than a century, the distribution would 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. We suspect that the conditions are highly unlikely to remain this stationary, and that the fund size distribution will remain indefinitely in its current log-normal, out of equilibrium state.

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 Δ_f represents the flow of money in and out of funds, and Δ_r the returns on money that is already in the fund. The money flow Δ_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 Δ_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 a difference in the long-term evolution: While the exit process is driving mutual funds to evolve toward a heavier-tailed distribution, other firms are evolving toward a thinner-tailed distribution. Again, though, due to the extremely slow relaxation times, we suspect this makes little or no difference.

Our analysis here suggests that the details of investor preference have a negligible influence on the upper tail of the mutual fund size distribution, except insofar as investors choose funds so as to enforce market efficiency. Investor preference enters our analysis only through Δ_f, the flow of money in and out of the fund. Since Δ_f becomes relatively small in the large size limit, the growth of large funds is dominated by the returns Δ_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. As a result, 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 Δ_f are essentially like those of the growth of other firms.

How can size-dependent transaction costs be compatible with our results here? We have performed an empirical study, which we will report elsewhere, that demonstrates that as size increases fund managers maintain constant after-transaction cost performance by lowering fees, reducing trading and diversifying investments. This is in contrast to the theory proposed by Berk and Green (2004) that fund size is determined by the skill of fund managers, i.e. that better managers attract more investment until increased transaction cost causes excess returns to disappear. Both our theory and that of Berk and Green are based on market efficiency. The key difference is that we find that the flatness of performance vs. size is enforced by simple actions taken by fund managers that do not influence the diffusion of fund size. In contrast, the Berk and Green theory requires choices by investors that directly influence fund size, and thus is not compatible with the free diffusion that we have prevented empirical evidence for here. In their theory transaction costs and investor skill determine fund size; in our theory, neither plays a role.

We would like to stress that, while we are fitting econometric models to the entry, exit and growth processes, and calibrating these models against the data, we are not fitting any parameters on the size data itself. This makes it challenging to get a model that fits as well as the model shown in Fig. 3.5. Of course, we have only demonstrated that the random process model is sufficient to explain fund size; we cannot demonstrate that other explanations might not also be able to explain it. However, the assumptions that we make here are simple and natural. The stochastic nature of fund growth is not surprising: It is well known that past returns do not predict future returns. Thus even if investors chase returns, they are chasing something that is inherently random. We believe that this is at the core of why our model works so well. Our demonstration that a good explanation can be obtained based on market efficiency alone, which requires weaker assumptions than full rationality, provides a theory that is robust and largely independent of the details of human choice.

3.A Inadequacy of Gini coefficients to characterize tail behavior

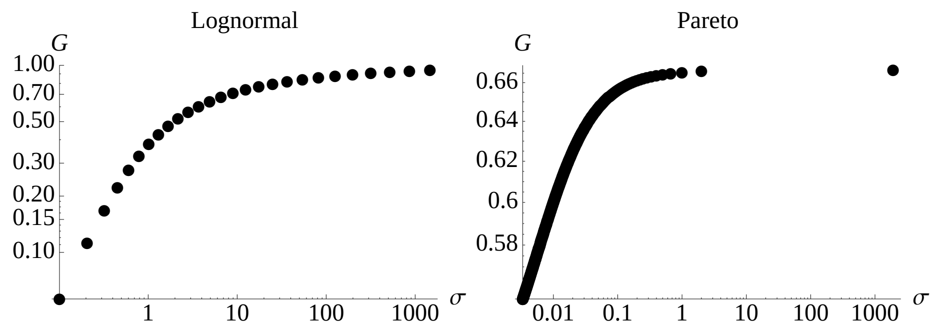


Figure 3.6: The Gini coefficients as described in equation (3.26) are calculated numerically for a lognormal distribution and a Pareto distribution. The Gini Coefficients were calculated for different parameter values and are plotted as a function of the resulting standard deviation. For the Pareto distribution (footnote 20) we used $s_0 = 0.01$ and different exponents α in the range $(2, 5]$, i.e. a finite second moment. The lower standard deviation $\sigma = 0.0033$ corresponds to $\alpha = 5$ and $\sigma = 1916.17$ corresponds to $\alpha \to 2$. For the lognormal we used $a = 0$ and different b in the range $[0.1, 2.8]$ where $b = 0.1$ corresponds to $\sigma = 0.101$ and $b = 2.757$ corresponds to $\sigma = 2000$.

The Gini coefficient (Gini, 1912) is commonly used as a measure of inequality but as we show here it is not suitable for distinguishing between highly skewed distributions when one wishes to focus on tail behavior. For a non negative size s with a CDF $F(s)$, the Gini coefficient G is given by

$$G = \frac{1}{E[s]} \int_0^\infty F(s)(1 - F(s))\mathrm{d}s, \tag{3.26}$$

where $E[s]$ is the mean (Dorfman, 1979). To illustrate the problem we compare the Gini coefficients of a Pareto distribution to those of a lognormal[19]. For a Pareto distribution with tail parameter α the $m > \alpha$ moments do not exist. This is in contrast to the lognormal distribution, for which all moments exist. Naively one would therefore expect that the Gini coefficient of the Pareto distribution (see footnote 18) to be larger than that of a lognormal since it has a heavier upper tail. This is true for a Pareto distribution with $\alpha < 2$, for which the Gini coefficient is one due to the fact that the standard deviation does not exist. However, when $\alpha < 2$, for large standard deviations the Gini coefficient of the log-normal is greater than that of the Pareto, as shown in Figure 3.6. In order to compare apples to apples in Figure 3.6 we plot the Gini coefficient as a function of the standard deviation (which is a function of the distribution parameters). For a Pareto distribution with a finite second moment ($\alpha > 2$) the lognormal has a higher coefficient.

Thus, even though the Gini coefficient is frequently used as measure for inequality, it is not a good measure when one seeks to study tail properties, particularly for comparisons of distributions with different functional representations. The reason is that the Gini coefficient is a property of the whole distribution, and depends on the shape of the body as well as the tail. Similar remarks apply to the Herfindahl index.

3.B 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

[19] The CDF of the Pareto distribution is defined as

$$F_p(s) = 1 - \left(\frac{s}{s_0}\right)^{-\alpha}, \tag{3.27}$$

where s_0 is the minimum size and α is the tail exponent. The CDF of a lognormal is given by

$$F_{ln}(s) = \frac{1}{2}\left(1 + \mathrm{Erf}\left[\frac{\log(s) - a}{\sqrt{2}b}\right]\right), \tag{3.28}$$

where a is a location parameter and b is the scale parameter.

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 λ which can depend on the fund size.

3. The rate of creation of new funds ν.

   Each new fund enters with a size ω 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 $\frac{\nu}{1+\lambda+\nu}$ 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 $\frac{\lambda}{1+\lambda}$ the fund enters.

   c. If it is not annihilated, which happens with a probability of $\frac{1}{1+\lambda}$, 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.

Chapter 4

Mutual fund trading costs

It is well established amongst practitioners that market impact induces diseconomies of scale in mutual funds. Nevertheless, market impact and its detrimental effect on mutual fund performance has yet to be fully accepted in the academic world. Here we show that market impact is not unique to financial markets but rather that for any market the laws of supply and demand dictate its existence. We argue using the CRSP holdings dataset that mutual funds trade mostly with counter-parties outside the industry and, assuming that mutual funds act as liquidity takers, their performance suffers due to market impact. We further provide an approximate functional form for mutual funds' trading costs.

4.1 Introduction

Trading costs can turn a profitable strategy into an unprofitable one as a fund gets bigger and therefore should be regarded as a driver of fund behavior. Therefore, it is important to understand their role in mutual fund dynamics. In this section we introduce and discuss some of the empirical work on trading costs as well as develop their role in the mutual fund industry.

Trading costs are comprised of both direct and indirect costs. Direct costs, which include brokerage and exchange fees, can be easily measured and controlled. Indirect costs, on the other hand, are not easily quantified. For large trades the most significant of them is the price impact, which is the effect the trade itself has on the price. The importance of each of the two cost components varies with the size of the fund. Direct costs are mainly fees paid for trading and as the fund grows it can negotiate lower fees. What this means is that the per dollar direct costs decrease with the size of the fund and can become negligible for large funds. Market impact, however, grows with the size of the trade and if the fund's position size increases as it grows then this results in an increasing per dollar costs. Thus for small funds, direct costs are of most concern while for larger funds market impact dominates. As a result, trading costs can be treated as a source of diseconomies of scale.

It is clear to practitioners and academics that paper trading is always more profitable than actual trading. Leinweber (2002) gives an explicit example of the performance difference between a fund's paper portfolio and its real portfolio. Nevertheless, the academic views on the role of trading costs in mutual funds is surprisingly polarized. Some, such as Edelen et al. (2007), argue that they are a major source of diseconomies of scale, while others argue that they are negligible at best. For instance, Chen et al. (2004) argue that they are not that important as funds can overcome increasing trading costs by splitting into independent sub funds and keep the (per dollar) trading costs constant. Another view, given in Fama and French (2010), argues that equilibrium accounting dictates that funds on average should not feel price impact. We address the equilibrium accounting view in this chapter and show empirically that it does not hold for mutual funds.

4.2 Market impact and the law of supply and demand

Market impact is usually attributed to market micro structure. However, market impact is present in any transaction as it results directly from the law of supply and demand. Stated simply: leaving all else equal, when a market participant changes her demand function this results in a change in the price.

We consider a simple case of a market for one stock with two participants. Given that participant i has a demand function $D_i(p)$ and that there are N shares, the price p^* of a single share is determined through market clearing as

$$D_1(p^*) + D_2(p^*) = N. \tag{4.1}$$

If participant 1 changes his demand function to $D_1(p) + \delta(p)$ the new price $\hat{p}$ is given by

$$D_1(\hat{p}) + \delta(\hat{p}) + D_2(\hat{p}) = N. \tag{4.2}$$

For small changes $D_1(p) \gg \delta(p)$ and a slow varying $\delta(p)$, we can approximate $\hat{p} = p^* + \Delta p$, where $p^* \gg \Delta p$ such that we can approximate (4.3) as

$$D_1(p^*) + \Delta p \frac{\partial D_2(p^*)}{\partial p} + \delta(p^*) + \Delta p \frac{\partial \delta(p^*)}{\partial p} + D_2(p^*) + \Delta p \frac{\partial D_2(p^*)}{\partial p} = N, \qquad (4.3)$$

where $\frac{\partial D_2(p^*)}{\partial p}$ is the derivative estimated at $p = p^*$. By using the market clearing condition (4.1) for p^* we get

$$\Delta p = \frac{\delta(p^*)}{\lambda}, \qquad (4.4)$$

where λ can be thought of as the liquidity and is defined as

$$\lambda = -[\frac{\partial \delta(p^*)}{\partial p} + \frac{\partial D_1(p^*)}{\partial p} + \frac{\partial D_2(p^*)}{\partial p}]. \qquad (4.5)$$

Assuming that the demand functions are decreasing function of the price, the liquidity is positive.

Thus, for a positive change in demand by participant 1, i.e. $\delta > 0$ corresponding to initiating a buy trade, the price of the asset will go up according to (4.4). Similarly, for $\delta < 0$, corresponding to selling an asset, the price of the asset will go down. This price change can be thought of as market impact. This means that for the initiator of a trade, the price changes adversely and the passive agent (agent 2), which did not change her demand, gained on the expense of the other agent. The initiating agent can be thought of as liquidity taker and the passive agent as a liquidity provider and what we have shown is that supply and demand alone dictated that liquidity providers gain at the expense of liquidity providers.

4.3 Market impact under equilibrium accounting

In the previous section we showed that market impact for a single trade results directly from the law of supply and demand. To persuade critics that market impact is truly important we must address the argument that the price impact of a single trade does not necessarily translate to an overall performance loss of a fund. Research as recent as Fama and French (2010) suggests that on aggregate (before fees) mutual funds have zero over performance with respect to the market. Therefore, one can mistakenly assume that mutual fund performance is a zero sum game where if one fund loses to impact another gains. A notion often referred to as equilibrium accounting. One can then argue that, on average, market impact should not affect performance since lose on trades is recuperated from other market participants' trading.

Equilibrium accounting can also lead to the mistaken arguement that if the market is a zero sum game then due to market impact, active investing is a negative sum game. We do not agree with such an argument. The fact that net losses equal to the total net gains does not mean that active investing is a negative sum game. We argue that instead of dividing the market into active and passive investors we argue here that the market should be divided into liquidity takers and liquidity providers. For such a distinction, equilibrium accounting dictates that on average liquidity providers gain on the expense of liquidity takers, even though both can be active traders.

To show this we investigate a simple toy model where two agents trade a single commodity. The impact can be intuitively thought of as resulting from the law of supply and demand. Stated simply: leaving all else equal, when a market participant changes her demand function this results in a change in the price. Each time a trade is initiated by a sell order the price moves down $p(t+1) = p(t) - 1$ and when it is initiated by a buy it moves up $p(t+1) = p(t) + 1$. For simplicity, the agents trade a single issue and $\Delta p = 1$. As expected from equilibrium accounting, when one participant initiates a buy the price moves up and the seller gaines from the transaction.

To investigate the long term gains and losses we take a stochastic approach: At each time step, participant 1 will choose to trade with probability ν. If participant 1 is in possession of the asset then he will chose to sell and buy otherwise. Similarly participant 2 will choose to trade with probability $1 - \nu$. The probabilities are defined such that each time step one of the participants initiates a trade. The expected price impact of buying an asset by agent 1 is then given by

$$\mathrm{E}[\Delta p|\mathrm{buy}] = \nu - (1 - \nu) = 2\nu - 1, \tag{4.6}$$

where the first term is for a buy initiated by participant 1 with probability ν and the second term is for a trade initiated by participant 2 selling with probability $1 - \nu$. Similarly, the expected price impact of buying an asset by participant 1 is given by $\mathrm{E}[\Delta p|\mathrm{sell}] = 1 - 2\nu$.

Given that the price at time t is p and that participant 2 is in possession of the asset, participant 1 will buy the asset at an expected price of

$$\mathrm{E}[p(t+1)] = p + 2\nu - 1$$

and will sell it at time $t+2$ at a price

$$\mathrm{E}[p(t+2)] = \mathrm{E}[p(t+1)] + 1 - 2\nu.$$

As we expect from equilibrium accounting, the expected price at time $t+2$ is equal to that at time t

$$\mathrm{E}[p(t+2)|p(t) = p] = p. \tag{4.7}$$

However, even though one participant's gain is another's loss, participant 1 gained Δ_1, which is given by

$$\Delta_1 = \mathrm{E}[p(t+2)] - \mathrm{E}[p(t+1)] = 1 - 2\nu \tag{4.8}$$

and the participant 2 gained

$$\Delta_2 = \mathrm{E}[p(t+1)] - p(t) = 2\nu - 1. \tag{4.9}$$

The net expected gain, $\Delta_1 + \Delta_2$, is always zero in agreement with equilibrium accounting but it does not vanish separately for each participant. The individual gains depend on the probability of a participant to initiate trades and will vanish only for the case of complete symmetry $\nu = 1/2$ where each agent is as likely to initiate a trade. For $\nu > 1/2$ participant 1 acts as a liquidity taker more frequently than as a provider and will lose on average to participant 2 who mostly acts as a liquidity provider.

4.4 Herding and equilibrium accounting of mutual fund trading

So far we have shown that one can separate the market into two types of participants, where liquidity providers on average gain at the expense of liquidity takers. The claim that mutual funds feel no impact requires that they act as both liquidity providers and liquidity takers. Equilibrium accounting suggests that this scenario can occur if they are trading solely between themselves (Fama and French, 2010). But, as we now show, they do not trade solely with each other but rather trade mostly with counterparts outside of the industry[1]. This means that if we assume that mutual funds act as liquidity takers then, on average, they lose to market impact.

To study this, we examine the directionality of trades, i.e. what fraction of the total trading volume in each asset is carried out within the industry. We define trading in a stock as one directional if all the funds that traded that stock in a given quarter were doing the same thing: either buying or selling that stock but not both. Trading in a stock is non-directional if all trades for a given quarter were carried out between mutual funds with one mutual fund selling to another.

To quantify the directionality of mutual fund trading we define the directionality parameter $\rho_j(t)$ for stock j at time t. The directionality can have values in the range $\rho_j(t) \in [0, 1]$ ranging from non-directional trading ($\rho_j(t) = 0$) to complete directionality $\rho_j(t) = 1$. We denote as $v_{ij}(t)$ the volume (in US dollars) that fund i traded in asset j at time t. The volume $v_{ij}(t)$ can be negative corresponding to the fund selling the asset or positive corresponding to the fund buying the asset. This leads to the natural definition of the directionality parameter [2]

$$\rho_j(t) = \frac{v_j^{net}(t)}{v_j^{tot}(t)}. \tag{4.10}$$

The total volume $v_j^{tot}(t)$ traded in an asset j at time t is defined as

$$v_j^{tot}(t) = \sum_i |v_{i,j}(t)|, \tag{4.11}$$

where the summation is over all funds. Similarly the net volume $v_j^{net}(t)$ traded is given by

$$v_j^{net}(t) = \left| \sum_i v_{i,j}(t) \right|. \tag{4.12}$$

To get a feel for the directionality parameter, we investigate the expected value of ρ for three hypothetical scenarios: funds trading solely with each other, funds trading solely

[1]Using the CRSP holdings data set described in Appendix 5.A we can approximate mutual fund trading by comparing the change in its reported portfolios.

[2]The directionality parameter is similar to the dollar ratio trade imbalance defined by Lakonishok et al. (1992).

with counter-parties outside the industry and funds trading with any counter-party with equal probability. If the mutual fund industry was a closed system, such that the funds were trading solely with each other, then we expect $\rho_j(t) = \sum_i v_{i,j}(t) = 0$ for all assets j. If, on the other hand, funds trade solely with counter-parties outside the industry then we expect that for all assets $|\sum_i v_{i,j}(t)| = \sum_i |v_{i,j}(t)|$, which corresponds to $\rho_j(t) = 1$. For the case were funds trade with any counter-party with equal probability, the value of $\rho_j(t)$ depends on the size of the mutual fund industry relative to the entire market. Since mutual funds hold approximately 25% of US equity[3] we expect that, on average, 75% of trades will be made with counter-parties outside the industry while 25% of the trades are with other funds. What this means is that half the trades volume will be between funds which results in $\rho_j(t) = 0.5$. We summarize this as

$$\rho_j(t) = \begin{cases} 0 & \text{Trading solely amongst funds} \\ 0.5 & \text{Uncorrelated fund trading} \\ 1 & \text{Completely correlated trading} \end{cases} \tag{4.13}$$

To investigate the actual behavior of the industry we compute the values of $\rho_j(t)$ using the empirical holdings data. For each quarter in the data set we computed ρ_j for all the assets [4] and computed the probability density $P(\rho)$ and the cumulative distribution function $P(\rho < X)$, [5] the probability of a randomly picked asset j' to have a value $\rho_{j'}(t) < X$. The results are given in Figure 4.1. From the CDF we can infer that most assets have $\rho > 0.62$, suggesting that mutual fund trading tends to be correlated and directional. Moreover, the probability density appears to have a peak near $\rho \approx 0.96$ which suggests that many assets appear to have one directional trading. What this means is that mutual fund trading is correlated such that if one fund acts in a certain way, i.e. buys or sells an asset, the others will follow suit.

To test the competing hypothesis of uncorrelated fund trades, we use a bootstrap simulation where each trade is assigned a random sign with equal probability to buy or sell. We

[3] As described in the investment company fact book available online at http://www.ici.org.

[4] To calculate $\rho_j(t)$ we used only assets with more than ten trades in that quarter.

[5] The density and cumulative distributions were estimated using a kernel smoothing technique with a normal kernel.

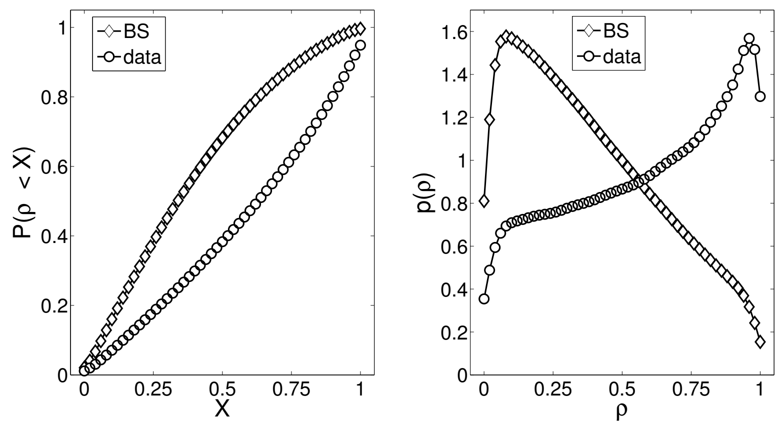


Figure 4.1: Here we show that funds mainly trade with counter-parties outside the industry by estimating the distribution of ρ as given by equation (4.10). The results ($\circ$) are compared to a bootstrap simulation of random fund trades where each trade receives a sign with equal probability ($\diamond$). **Left:** The cumulative distribution $P(\rho < X)$ for the volume traded per asset in a quarter. **Right:** The probability density $P(\rho)$ for the volume traded per asset in a quarter.

compare the results with a bootstrap estimation of a closed industry in which each trade was assigned with equal probability a random sign, corresponding to either buying or selling. The results are compared to the empirical distributions in Figure 4.1. The bootstrap estimation is strikingly different than the empirical results; the PDF has a peak around $\rho \approx 0.08$ and the CDF suggests that most assets have $\rho < 0.32$. There are two reasons why ρ is not strictly zero; the trades have different sizes and some assets have a relatively small number of trades. To test the robustness of these results with respect to the random trade signs, the estimation of the PDF and CDF was repeated 100 times and for each ρ we calculated the standard error of $p(\rho)$ and $P(\rho < X)$. The standard errors are plotted in the figure but they are smaller than the marker and can't be seen.

The results, as we have shown here, suggest that not only do mutual fund trade mainly with counter-parties outside the industry but that mutual fund strategies are correlated. It seems that funds tend to herd in the sense that if one fund buys a certain stock the other funds tend to do the same, that is, if they trade that stock it will be to buy it in agreement with past work [6].

[6]Past work has shown that mutual funds do tend to herd albeit mainly in smaller cap stocks (see Wermers

4.5 A functional form for market impact

What we have shown so far is that mutual funds are affected by market impact. We now turn to try and estimate the magnitude and components of market impact. Our goal in this section is to derive a good approximation for the functional form of mutual fund's trading costs.

4.5.1 Market impact of packaged or block trades

The functional form of the impact of a single trade has been extensively studied both theoretically and empirically. Kyle (1985) in a pioneering work argued that the per dollar impact increases linearly with the size of the trade while more recent work postulates that the impact increases as the square root of size[7]. For a recent review on price impact see Bouchaud et al. (2008). This research has enabled market practitioners to predict, within reasonable accuracy, the cost of a trade (Torre, 1997). For large institutional investors, however, the story is a bit more complicated. Institutional investors such as mutual funds trade large positions. Due to liquidity constraints large positions cannot be simply executed as if they were smaller trades, on which most empirical studies are based. These large orders are either traded in blocks or split into smaller trades, called packaged trades. The impact of these trades is harder to estimate as it will depend not only on the size of the package (or block) but also on the duration of the execution[8].

It is important that we have a sensible estimate of the functional form of market impact since our results are sensitive to its magnitude and to the way it depends on the size of the fund. In Table 4.1 we provide a short literature review that summarizes the estimated magnitudes and different functional forms for market impact. From the table, one thing is clear: market impact is not negligible. The magnitude, for the most part, is the order of tens of basis points and the functional form $\chi(v)$ appears to increase as some power of the traded volume v^{β}. Therefore, using these results as a guide, we approximate the price

---

(1999) and references therein). In more recent work Brown et al. (2009) have shown that funds tend to follow analysts' upgrades and downgrades.

[7] Torre (1997); Kempf and Korn (1999); Farmer and Lillo (2004); Gabaix et al. (2006); Hasbrouck (2009).

[8] The size dependence of blocks and packaged trades has been studied by (Holthausen et al., 1987, 1990; Loeb, 1991; Hausman et al., 1992; Keim and Madhaven, 1996; Chan and Lakonishok, 1997; Plerou et al., 2002; Leinweber, 2002; Moro et al., 2009). The impact was shown to decrease with the duration of the trade by Dufour and Engle (2000) and more recently by Almgren et al. (2005).

impact of mutual fund trades as

$$\chi(v) \approx cv^{\beta}, \tag{4.14}$$

where c is a scale constant. The scale constant c will vary between stocks with different market cap and liquidity. Regardless of the exact value of the exponent, there is an agreement that impact is a concave function of the volume. This means that the exponent takes the value $\beta \in [0, 1)$, where $\beta = 0$ corresponds to a logarithmic dependence.

4.5.2 Mutual fund trading costs

We estimate the amount of trading costs a fund incurs using the market impact models described in Section 4.5.1 as a guideline. Neglecting fixed costs the trading cost of trading a volume v in a given asset is given by the market impact

$$\chi(v) = c\, v^{1+\beta}, \tag{4.15}$$

where v is the dollar value of the trade. If a fund of size s trades in only a single asset and changes a position once a year then it trades twice a year a volume $v = s$ such that the trading cost is given by

$$s_t C_t(s_t) = \chi(s_t) = c\, s_t^{1+\beta}. \tag{4.16}$$

In a more general case a fund has a turnover ratio λ, defined as the minimum of total sales or purchases of the asset in a year divided by the average fund size. If the turnover is smaller than unity $\lambda < 1$ then the fund traded only a part of its assets, i.e. traded a volume $v = \lambda s$. If the turnover is larger than unity then we can think of it as if the fund traded $v = s$ (all its assets) λ times. Written more quantitatively, the trading cost for a fund of size s and a turnover rate of λ is given by

$$s_t C_t(s_t, \lambda) = \begin{cases} \lambda\, \chi(s_t) & \text{if} \quad \lambda \geq 1, \\ \chi(\lambda\, s_t) & \text{if} \quad \lambda < 1. \end{cases} \tag{4.17}$$

	study	Functional form $\chi(v)$	Magnitude (bps)	Trades size (USD)	Market (period)
I	Moro et al.	$v^{1/2}$	6 11	–	LSE (01-04) BME (01-04)
II	Holthausen et al.	$\beta_0 + \beta_1\, v;\, \beta_1 \approx \begin{cases} 2.03 \\ -1.38 \end{cases}$	70 buy -60 sell	–	NYSE (82-84)
III	Dufour and Engle	$\log(v)$	5	single trade	NYSE (90-91)
IV	Keim and Madhaven	v^{β}	160 buy -150 sell	–	NYSE, NASADAQ AMEX (85-92)
V	Plerou et al.	$v^{1/3}$	–	–	NYSE (94-95)
VI	Hausman et al.	$(v^{\lambda} - 1)/\lambda$	7.9 to 22.5	100 K	NYSE, AMEX (98)
VII	Torre	$v^{1/2}$	100	1 MM	NASDAQ, NYSE (97)
IIX	Chan and Lakonishok	–	50	2 MM	NASDAQ, NYSE (89-91)
IX	Almgren et al.	$v^{3/5}$	20 to 40	70 MM (IBM) 2.5 MM (DRI)	NYSE (01-03)
X	Loeb	$\approx \frac{0.12}{m^{0.54}}\, v^{0.36+0.02\log(v/m)}$	111	1 MM (for m=10B)	US small cap (70-89)
XI	Leinweber	–	40 to 70	–	Various (91,01)

Table 4.1: A summary of select empirical work on the functional form $\chi(v)$ of market impact for large trade volumes v in the form of either single large trades, block trades or packaged trades.

I. The mean impact,for packaged trades completed within a single day with at least 10 transactions, is approximately 1.1 times the spread in the BME and 0.6 times the spread in the LSE. The dollar value of the trades is not quoted. The authors report (private communications) a mean spread of 10 bps for liquid stocks in the BME. The mean spread in the NYSE at that period was reported to be 12.9 bps Chordia et al. (2005) and 11bps Almgren et al. (2005).

II. The authors report a median block size of 0.08% of outstanding equity traded resulting in a median impact of 0.74% for buy initiated trades and -0.61% for sell initiated trades. The authors do not specify the average value of outstanding equity.

III. The authors used 144 stocks from the TORQ data set from November 1990 to January 1991. The size of the trades is not quoted.

IV. The exponent β is not quoted. The median block size as a fraction of outstanding equity traded is 0.34% for sell orders and 0.16% for buy orders.

V. The authors used the 116 most traded stocks from the TAQ data base.

VI. For $\chi(v)$ the authors used the Box-Cox transformation with λ ranging from $\lambda = 0$ (logarithm) to $\lambda \approx 0.2$. The market impact was estimated for trades of size 100,000 USD. The data set is comprised of 100 randomly chosen stocks with at least 100,000 trades.

VII. For the 1000 largest capitalization stocks for the last 5 trading days of July 1997.

IIX. The NYSE dataset consists of 166,000 packages with a median size of $\approx 2M$ and a median impact of 0.36 to 0.54%. The NASDAQ dataset consists of 68,000 packages with a median size of $\approx 2M$ and a media impact of 0.47 to 0.99%.

IX. The authors present typical costs for packages with a size 0.1% of daily volume for different trading durations.

X Parameters were fit for block trades using the COMPUSTAT small cap dataset. The volume v is in thousands of dollars and the market cap m is in millions.

XI. The author presents typical costs as a function of the volume (in shares) using the trades of a 2 B$ pension fund.

Using the definition of χ in equation (5.19) yields

$$s_t C_t(s_t, \lambda) = \begin{cases} c\,\lambda\, s_t^{1+\beta} & \text{if} \quad \lambda \geq 1, \\ c\,(\lambda s_t)^{1+\beta} & \text{if} \quad \lambda < 1. \end{cases} \tag{4.18}$$

If we assume that the traded volume is divided equally between N assets, i.e. if the total volume traded by the fund is v then it trades N assets with a volume of v/N in each. Under such conditions we write the impact as

$$s_t C_t(s_t, \lambda, N) = \begin{cases} c\,\lambda\, \frac{s_t^{1+\beta}}{N^\beta} & \text{if} \quad \lambda \geq 1, \\ c\, \frac{(\lambda s_t)^{1+\beta}}{N^\beta} & \text{if} \quad \lambda < 1. \end{cases} \tag{4.19}$$

Since most funds (about 80% of the funds in the data set) have a turnover rate below 1 we approximate the turnover rate as

$$C(s, \lambda, N) = c\,\lambda^{1+\beta} \left(\frac{s}{N}\right)^\beta . \tag{4.20}$$

4.6 Conclusions

We have shown that market impact should not be viewed as a unique phenomenon to the financial market but rather as something more general. In every market supply and demand dictates that for an initiator of a trade, the price moves adversely. Instead of dividing the market into active and passive traders we divided it into liquidity takers and liquidity providers. The liquidity provider earns, on account of market impact, at the expense of the liquidity taker.

We argue that market impact as it relates to mutual funds cannot be dismissed on the basis of equilibrium accounting. To do so we calculated the directionality of mutual fund trading. What we have shown is that on most assets mutual funds' strategies tend to be correlated; at a given quarter a stock is either mostly bought or mostly sold by mutual funds. This directionality of trading can be also viewed as an indicator for mutual fund herding.

4.A Dataset

We use the Center for Research in Security Prices (CRSP) survivor-bias-free mutual fund data set. We focus on the holdings data set containing holdings data for a subset of the mutual funds, reported at various dates for the years 2002 to 2009 (excluding the last quarter of 2009). The dataset contains data on 8487 funds. Out of this universe of funds we consider only active equity funds, which we define as funds whose portfolio consists of at least 80% of cash and stocks and have a reported yearly turnover rate larger than zero. Out of these funds we consider only funds with a real TNA of at least 1 million USD. This reduces the universe of funds we consider to 1404 funds. For these funds the trades were estimated by looking at the change in the quarterly reported portfolios.

Chapter 5

Solving the persistence paradox: a reduced form model of mutual fund behavior

Mutual funds face diminishing returns to scale as a result of convex trading costs yet there is no persistence nor a size dependence in their performance. To solve this puzzle we offer a new framework in which skillful profit maximizing fund managers compensates for decreasing performance by lowering their fees. We show that mutual fund behavior depends on size such that bigger funds charge lower fees and trade less frequently in more stocks. We present a reduced form model that is able to describe quantitatively this behavior. In addition we offer a functional form for the average before costs performance and for the trading costs of mutual funds.

5.1 Introduction

Mutual fund performance is paradoxical; while the per dollar trading costs create diseconomies of scale which should lead to persistent performance and a fund performance that decreases with the size, there is no persistence in returns and all funds yield the same performance, on average, regardless of their size[1]. In this work we resolve the no performance paradox by introducing a new framework for fund behavior in which the manager compensates for increasing costs by lowering fees in agreement with empirical data.

There are currently two popular explanations for the lack of persistence in the literature; the first explanation is a trivial no skill and no increasing costs argument while the second argument is that any over-performance is eroded by increase in costs due to investor inflows.

[1] The size independent performance has been observed in Schwarzkopf and Farmer (2010b,a) and is investigated here in Appendix 5.E. Lack of persistence has been documented by Malkiel (1995a) by Carhart (2009) and others. For a more detailed review on persistence see Berk and Green (2004) and references therein.

The first type of explanation argues that on average fund managers have no skill (Fama and French, 2010) and therefore there should be no persistence in their pre-cost performance. The lack of after-cost persistence or size dependence requires that the relative trading costs do not increase with size. One such argument against increasing relative trading costs is an equilibrium accounting 'you win some you lose some' type argument. In Chapter 4 we argue that for mutual funds the relative trading costs increase with size and therefore this type of explanation for the performance paradox cannot hold.

The second type of argument was proposed by Berk and Green (2004). In their framework skillful managers generate higher than the benchmark pre-costs performance. Investors react to over-performance by investing in the fund, which increases its size. The increase in size results in higher relative trading costs which diminishes the fund's after-cost performance. Thus, it is the performance chasing investors that are responsible for the observed lack of persistence. While theoretically appealing, we argue in Section 5.4 that, among other things, investors do not react strongly enough to past performance and conclude that the Berk and Green argument cannot explain the paradox.

Our solution to the persistence paradox is relatively simple and does not require sophistication from either the investors nor from the fund managers. Even though our solution seems similar to Berk and Green (2004) it is conceptually different. In our model, investors have no direct role. The role investors play in our model is indirect. Since funds have to compete over investors they make sure to have a performance that is as good as that of its competitors. While it is true that investors have been shown to chase past performance (Chevalier and Ellison, 1997) we show in Section 5.4 that this relation is very noisy. It is quite reasonable to assume that investors react to other things such as advertising and marketing, an assumption supported by Barber et al. (2005). Moreover, Schwarzkopf and Farmer (2010a) have shown that part of fund growth that results from investor flows shows universal features observed in firm growth and many other systems. They modelled investor decisions via a social influence process taking place over the investors social network.

This leads us to the simple model in which the managers' main agenda is to maximize the funds profit while remaining competitive. A skillful manager with before costs over-performance, her alpha, will charge fees such that the after-cost performance is equal to its competitors. What this means is that persistence is eliminated through fees. This is a very plausible behavior given that Christoffersen (2001) has shown that managers change fees dynamically in reaction to the fund's relative performance. Thus, this is a simple and very plausible solution to the persistence paradox that does not require investors to be highly sophisticated but rather that the fund managers follow a relatively simple strategy.

The manger in our model takes the size of the fund as given and seeks to maximize fees without hurting the after costs performance. To do so the manager optimizes her strategy to reduce costs by changing the number of stocks the fund trades in and the trading frequency as given by the turnover rate. We argue that a strategy with reduced costs can have a lower alpha which can result in a lower performance. What this means is that for a given fund size there is an optimal strategy that maximizes the funds' profits with respect to both alpha and costs.

To test our model we propose a reduced form model for fund behavior which we fit to empirical data. The resulting fund behavior, as defined by its expense ratio, [2] the number of different assets it trades in and the yearly turnover rate are in agreement with the empirically observed behavior. We show that on average the number of different assets a fund trades in increases with fund size, the turnover rate and the expense ratio decreases with fund size. We interpret this as a fund's reaction to increasing costs; it tries to control them through diversification and lowering of turnover rates. It compensates for whatever costs cannot be reduced by lowering its fees. Our reduced form model suggests that approximately 47% of the compensation is due to turnover rate reduction 3% due to diversification and the remaining 50% is achieved by lowering fees.

The paper is organized as follows: In Section 5.2 we discuss mutual funds trading costs and in Section 5.3 we discuss the empirical observations of mutual fund behavior as a function of size. In Section 5.4 we discuss the Berk and Green framework and its failures with respect to empirical observations. In Section 5.5 we describe our framework for mutual fund behavior and in Section 5.6 we test our model with respect to empirical data. Finally, we conclude in section 5.7.

[2]The expense ratio is defined as the fraction of the assets under management used for operating expenses, such as management and advisory fees, overhead costs, and distribution and advertising (12b-1) fees. The expense ratio does not include brokerage fees incurred by the fund.

5.2 Performance and trading costs

The after-cost performance of a fund, which is the return for an investor in the fund, can be written in the following general form

$$r = \alpha - C - f, \tag{5.1}$$

where r is defined as the excess return over the average mutual fund performance [3], α is the fund manager's skill as translated to an ability to produce excess return, C denotes the per dollar costs of the fund and f denotes the expense ratio. The excess return for an investor in a fund, given by (5.1), is a sum of two competing terms: the first is the manager's skill and the second is the costs and fees and as we soon show, all of these terms depend on the size of the fund.

As we describe in Chapter 4, trading costs are a source of diseconomies of scale in mutual funds and therefore have an important role. Trading costs are comprised of both direct and indirect costs. Direct costs, which include brokerage and exchange fees, can be easily measured and controlled. Their relative size decreases as a fund grows and has more leverage to negotiate better prices. Indirect costs, on the other hand, are not easily quantified. For large trades the most significant of them is the price impact, which is the effect the trade itself has on the price. As described in Chapter 4, impact grows with the size of the trade and if the fund's position size increases as it grows then this results in an increasing per dollar costs. Thus for small funds, direct costs are of most concern while for larger funds market impact dominates.

5.3 Empirical observation of fund behavior

One of the goals of this work is to explain the size dependence of fund behavior observed in the CRSP holdings dataset for equity funds[4]. The dataset contains the reported number of holdings N, the yearly turnover rate[5] λ and the expense ratio f for each fund for each

[3]This is similar to Berk and Green (2004) however we do not make the assumption that mutual fund average performance is equal to the benchmark.

[4]We have defined equity funds as funds whose portfolio is comprised of at least 80% cash and stocks. A more detailed description of the data set can be found in Appendix 5.A.

[5]The yearly turnover rate is reported yearly and is calculated using the minimum of the total purchases and the total sales. We use the reported turnover rates even though a better approximation would be the

quarter. To get better statistical power we aggregate over different years, where for each year we use the data reported at the end of the second quarter of that year yielding 2190 observations for which we have the three entries: N, λ and f. The average number of different stocks in the portfolio N, the average yearly turnover rate λ and the average expense ratio f are plotted as a function of fund size in Figure 5.1. What we see is that larger funds have a lower expense ratio, they hold more assets and they trade slower than smaller funds.

In addition to the plots, which enable an easy visualization of the behavior, we summarize in Table 5.1 the data according to size deciles. The number of different assets held by a fund increases with the size such that funds in the first size decile hold a median of 53 assets compared to 84 for the 10th decile. This is a weak effect and funds do not tend to increase the number of assets dramatically, in agreement with findings by Pollet and Wilson (2008). The turnover rate decreases more dramatically with a median turnover rate of 0.73 for funds in the 1st decile compared to a median turnover rate of 0.28 for funds in the largest decile. The expense ratio decreases dramatically as well: the median expense ratio for the 1st decile is 175.5 bps compared to 90 bps for funds in the 10th decile, which is more than a 50% decrease. From the table and figures the size behavior is clear; compared to smaller funds, larger funds trade in more assets, they have a lower turnover rate and they have a lower expense ratio.

We use these summary statistics to emphasize that this is a non-negligible behavior that needs to be explained. As a teaser for things to come, we added in Table 5.2 a comparison of the empirical observations to the predictions of our model described in Section 5.5. One can see that our model is in good agreement with the data and that we are able to explain these observations.

5.4 The Berk and Green framework

Berk and Green (2004) proposed, in a highly influential paper, a framework in which trading costs are an increasing convex function of fund size, fund managers are endowed with some skill, i.e. a before costs alpha and investors are bayesian updaters that use past returns

ratio of half the net trading volume, to the average fund size. To verify that the reported turnover rates are a reasonable approximation we compared them to the estimated turnover rate from the holdings data and found them to be proportional.

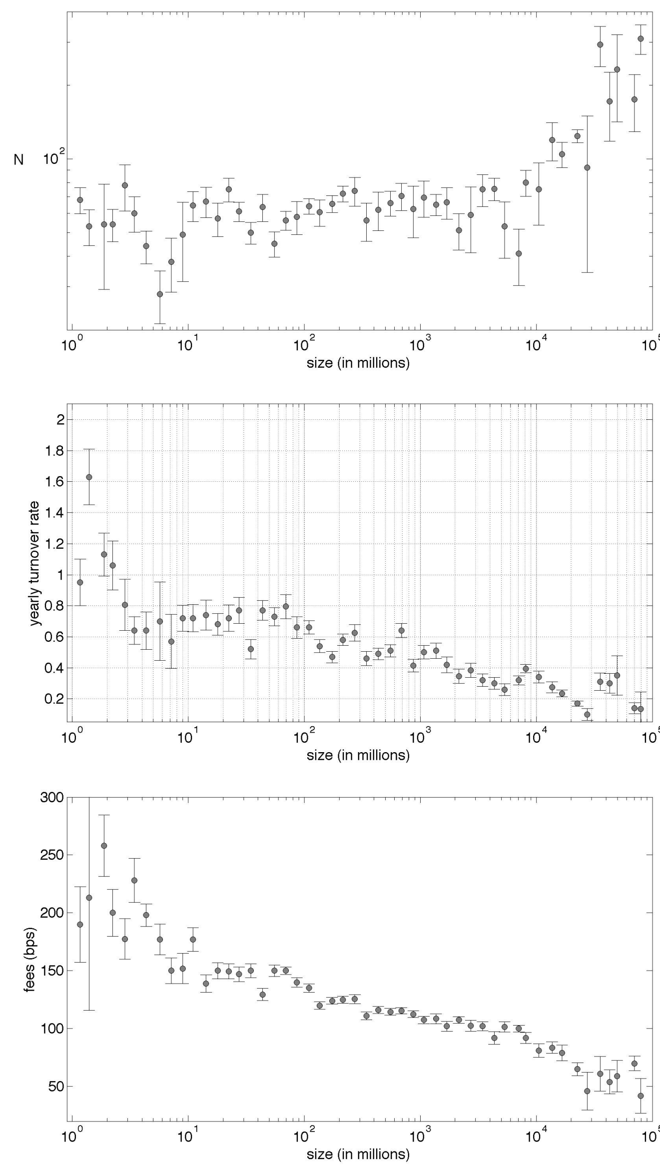


Figure 5.1: Fund behavior as a function of size (in millions of real USD). In all three panels the data was binned into 50 exponentially spaced bins. For each bin the median is quoted and the error bar corresponds to the mean absolute deviation in the bin divided by the square root of the number of observations in the bin. **Top:** The number of assets in the portfolio N. **Center:** The yearly turnover rate. **Bottom:** The yearly expense ratio.

Decile	size (in millions)	N data	N model	λ data	λ model	f (bps) data	f (bps) model
1/10	5.2	53	65	0.73	0.96	176	164
	(1.2,10.6)	(7,336)	(62,66)	(0.12,2.62)	(0.85,1.23)	(76,351)	(154,184)
2/10	19	62	67	0.73	0.76	148	146
	(12,28)	(17,387)	(66,68)	(0.11,2.85)	(0.72,0.83)	(75,260)	(142,153)
3/10	45	53	69	0.69	0.66	146	136
	(30,62)	(6,307)	(68,70)	(0.10,2.41)	(0.62,0.71)	(39,253)	(132,141)
4/10	84	57	70	0.72	0.59	140	129
	(66,113)	(11,288)	(70,71)	(0.15,2.57)	(0.56,0.62)	(67,208)	(126,132)
5/10	156	64	72	0.53	0.53	124	122
	(120,200)	(12,285)	(71,72)	(0.07,1.70)	(0.51,0.56)	(75,191)	(119,125)
6/10	261	72	73	0.61	0.48	123	117
	(208,348)	(10,338)	(72,73)	(0.10,1.75)	(0.46,0.50)	(49,197)	(113,119)
7/10	488	63	74	0.50	0.43	115	110
	(364,611)	(10,475)	(74,75)	(0.06,1.62)	(0.42,0.46)	(49,175)	(108,113)
8/10	852	70	75	0.49	0.39	114	104
	(646,1146)	(6,531)	(75,76)	(0.08,1.63)	(0.37,0.41)	(50,158)	(102,107)
9/10	1.9 [10^3]	64	77	0.44	0.34	105	97
	(1.2,3.2)[10^3]	(6,364)	(76,78)	(0.06,1.62)	(0.31,0.37)	(30,156)	(92,101)
10/10	7.4 [10^3]	84	80	0.28	0.26	90	85
	(3.6,67)[10^3]	(9,500)	(78,83)	(0.05,0.78)	(0.17,0.30)	(18,133)	(68,91)

Table 5.1: Summary statistics of mutual fund behavior and the model behavior calculated for each size decile. The 'data' values for the size (in millions of USD), the number of assets N, the yearly turnover rate λ and the expense ratio f are calculated from the CRSP holdings dataset. The 'model' values were estimated using parameter values given in Table 5.2 for fitting scheme $\mathcal{M}_a$. The median value in the corresponding size decile is quoted in each table entry. The values in parentheses represent the 5th and 95th percentile in each bin.

to estimate the fund's alpha. A fund with a positive expected before costs performance attracts inflows until its size increases to the point where trading costs offset the fund's alpha and the after cost performance is zero. Similarly, funds with negative expected before costs performance experience outflows and are driven out of business. Thus, In equilibrium their model predicts that even though all remaining funds have a positive alpha there is no persistence.

Fama and French (2010) argue that Berk and Green's framework fails because they concluded that on average they underperform the benchmark. While we agree with Fama and French that the Berk and Green framework is based on faulty assumptions, we think that other assumptions are more important. It is important to note that Berk and Green's explanation is valid only in equilibrium while Schwarzkopf and Farmer (2010a) have shown that the industry is in a non-stationary state and given that they calculated a relaxation time scale of over a century, equilibrium is an invalid assumption. Since their model can explain the lack of persistence only in equilibrium, it is clearly not the right explanation. Nevertheless, we list in this section the key assumptions that we think are either false or do not seem to agree with recent work.

An interesting point arises when one considers the relation between the skill and the size in light of the size distribution of funds, which Schwarzkopf and Farmer (2010b) have shown to be log-normal. In the Berk and Green model the equilibrium size of a fund can be mapped directly to the manager's alpha. Therefore if one assumes that the industry is in equilibrium one can back out the implied α distribution using equation (5.28). In Figure 5.2 we compare the implied alpha distribution to the distribution used by Berk and Green to empirically test their model. It is clear from the figure that they are qualitatively different. They are also quantitatively different; the implied distribution has a mean of 14% and a standard deviation of 20% while Berk and Green used a normal with a mean of 6.5% and standard deviation of 6%. This means two things; first, the parameter values and skill distribution used by Berk and Green to validate their fund flows predictions are wrong, which undermines the validity of their tests. Second the implied skill distribution has a heavy tail (it is a log-normal distribution) which does not seem plausible yet we cannot reject their model on account of this.

Second, the Berk and Green argument relies on the fact that investors are rational Bayesian updaters with infinite pockets that react swiftly to fund performance. What we

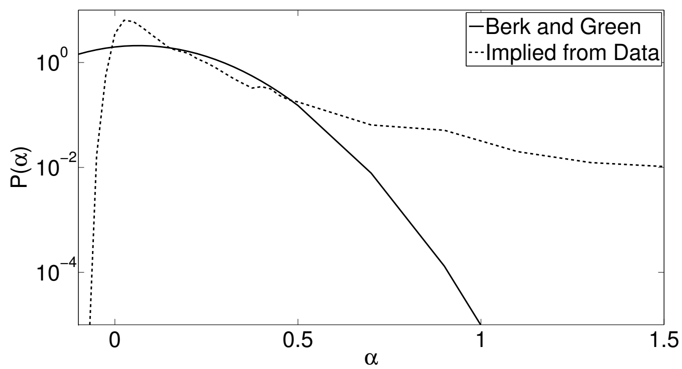


Figure 5.2: Here we show that the empirical size distribution implies a qualitatively different alpha distribution than the distribution used by Berk and Green in their empirical comparison. The implied alpha distribution was calculated from the size distribution of equity firms existing at the end of 2007 using equation (5.28) with $\beta = 1$, $c = 10^{-3}$ (for size in millions) and $f = 0.015$. The implied distribution has a mean of 14% and a standard deviation of 20%. The Berk and Green skill distribution is a normal with a mean of 6.5% and standard deviation of 6.5%. These are the parameter values used in Berk and Green (2004).

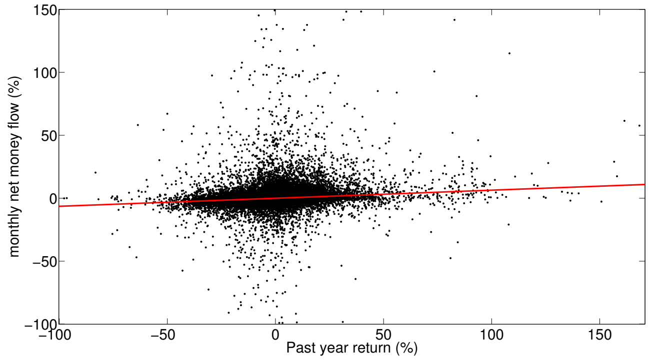


Figure 5.3: Here we show that investor performance chasing is noisy. Monthly fund flows are plotted as a function of the past yearly (idiosyncratic) returns. The data is aggregated for each for the years 1998 to 2007. The data is compared to a linear regression $\Delta_f = 0.06r - 0.014$. One can see that even though a trend exists the data is very noisy.

now show is that both of these assumptions are not valid and therefore the lack of persistence can not be explained by their framework.

According to Berk and Green investors react solely to past performance. However, when one looks at the relation between past performance and fund flows, as given in Figure 5.3, it is clear that even though a relationship exists it is very noisy. This leads us to question the importance of past performance in determining fund flows. Even so, investor money is very sticky with an average investor tenure in a fund family of 8 years [6]. Thus, if the Berk and Green explanation were true then the whole process would take place in time scales of years and one would expect to see performance persist for similar time periods.

Another reservation we have with the role investors play in their model is the complicated tasks that are required of them. There are more than 7000 funds in the US with more than 87 million Americans investing in them. The notion that these investors are capable of estimating fund performances and picking out over performing funds is utterly ridiculous and surely not the way people invest. This is strengthened by recent work by Barber et al. (2005) who suggest that investors react strongly to other things such as advertising and marketing. Moreover, Schwarzkopf et al. (2010) suggest that investor choices are strongly influenced by their social network. This leads us to conclude that their investor behavior does not hold water.

[6]Taken from the 2010 Investment Company Institute fact book.

Another key assumption made by Berk and Green is that investors have infinite pockets. They assume that investors have enough money to invest in all mutual funds regardless of their estimated skill and the corresponding size. If we use their rational, and their assumed parameter values, then to ensure efficiency in equilibrium mutual fund sizes have to be very large. If we aggregate over their hypothesized skill distribution then the resulting size of the industry is several orders of magnitude larger than it really is. What this means is that there is not enough money in the industry to ensure efficiency, i.e. that funds' expected over-performance vanishes. We discuss this in more depth in Appendix 5.D.

We conclude that in light of the arguments given above, the Berk and Green framework is qualitatively flawed and fails to describe the data in a consistent way.

5.5 A new framework for fund behavior

Here we develop a framework for mutual fund behavior that is able to explain the lack of persistence in mutual fund performance in a way that is consistent with the empirical observations of Section 5.3. Given the arguments on investor fund flows in Section 5.4 we do not include investor fund flows directly in the model but rather treat them as an exogenous process. Instead, we focus on the fund manager actions.

In this framework, the fund manager has some skill α and her main objective is to maximize profits given the current fund size.[7] The manager's constraint is that the fund has to remain competitive and by that we mean that it must perform as well as the other funds. We write the fund's after costs performance r as

$$r = \alpha - C - f, \tag{5.2}$$

where we treat r as relative to the industry average at that period. The profit of a fund of size s is given by

$$s\,f = q\,(\alpha - C - r)\,. \tag{5.3}$$

Since the manager would like to maximize her profit she would like to charge as high a fee as possible. The constraint that the fund has to remain competitive means that the return to the investor cannot drop below 0. Therefore, for a given α and C the fees the manager

[7] As the size s of a fund we refer to the dollar value of its total net assets (TNA).

charges will be the amount that ensures that $r = 0$, i.e. the after costs performance is equal to the industry average. [8] Thus the fees charged by a fund are given by

$$f = \alpha - C. \tag{5.4}$$

To further maximize the fees without hurting performance, the manager can increase the before costs performance α and/or decrease the costs C by choosing an appropriate strategy. For simplicity we define a strategy by N and λ, where N is the number of different assets held by the fund and λ is the yearly turnover rate. Since both α and C depend on all three variables s, λ and N, we write the fees as

$$f(s, \lambda, N) = \alpha(s, \lambda, N) - C(s, \lambda, N). \tag{5.5}$$

Changing N and λ in order to lower costs can have an adverse effect on α. Increasing N reduces the average position size which reduces the costs incurred by the fund due to price impact. However, if we assume that the fund trades in what it thinks are the best performing stocks, then increasing the number of stocks means that it has to include inferior stocks in the portfolio, which reduces α. [9] Similarly by decreasing λ the fund reduces costs by trading less. However, by trading less frequently the fund reacts more slowly to information and it cannot exploit inefficiencies occurring at higher frequencies. This results in a decrease in α. Therefore, for a given fund size there is an optimal number of assets $N_{opt}(s)$ and an optimal turnover rate $\lambda_{opt}(s)$ for which the fees, given by equation (5.5), are maximal, i.e.

$$\left.\frac{\partial f(s, \lambda, N)}{\partial N}\right|_{\lambda_{opt}, N_{opt}} = 0 \quad \text{and} \quad \left.\frac{\partial f(s, \lambda, N)}{\partial \lambda}\right|_{\lambda_{opt}, N_{opt}} = 0. \tag{5.6}$$

This results in an optimal fee $f_{opt}(s)$ that depends on size and is given by

$$f_{opt}(s) = \alpha(s, \lambda_{opt}(s), N_{opt}(s)) - C(s, \lambda_{opt}(s), N_{opt}(s)). \tag{5.7}$$

Thus, given a functional form for $\alpha(s, \lambda, N)$ and $C(s, \lambda, N)$, we can infer fund behavior as

[8] Christoffersen (2001) has shown empirically that managers dynamically control the amount of fees they charge through rebates on an initially high fee.

[9] Increasing N can have a positive effect on the performance when taking risk into account as it increases the diversification of the portfolio.

function of size, where the number of positions is given by $N_{opt}(s)$ the turnover rate $\lambda_{opt}(s)$ and the fee by $f_{opt}(s)$.

The functional form of skill and costs

Our cost function estimation is based on the approximate functional form of price impact described in Chapter 4 and in Appendix 5.C of this chapter. The per dollar managed trading costs for a fund of size s trading N different assets with a yearly turnover ratio λ is given by

$$C(s, \lambda, N) = c\,\lambda^{1+\beta}\left(\frac{s}{N}\right)^{\beta}, \tag{5.8}$$

where c is a constant that can be interpreted as the annual trading cost of trading a million USD. The costs increases as a power of the average position size s/N, such that as the fund gets bigger the costs increase and the after transaction performance suffers. A fund can decrease those costs by diversifying, i.e. increasing N or by trading less frequently, i.e. decreasing the turnover rate λ.

As a first approximation for the functional form, we describe α as a product of three terms

$$\alpha(s, \lambda, N) = \mathcal{Q}(s)\mathcal{L}(\lambda)\mathcal{N}(N).$$

The first term is the size dependence, the second is the turnover rate dependence and the third is the number of different assets dependence. We now approximate each one of these terms using simple economic intuition as a guide.

For the turnover rate term $\mathcal{L}(\lambda)$, we assume that $\mathcal{L}$ is an increasing function of the turnover rate $\mathcal{L}'(\lambda) > 0$. This is because we assume that a larger turnover rate corresponds to a larger volume of strategies which translates to more profit opportunities. We further assume that $\mathcal{L}$ is a concave function such that $\mathcal{L}''(\lambda) < 0$. Furthermore, if the excess return of indexing vanishes, then this leads to $\mathcal{L}(0) = 0$, which leads to the following approximation

$$\mathcal{L}(\lambda) = 1 + \log\left(1 + \frac{\lambda}{\Lambda}\right), \tag{5.9}$$

where Λ is a scale parameter.

We now turn to $\mathcal{N}(N)$, the term describing the dependence of α on the number of assets held. We expect that the over-performance will decrease with the number of assets held i.e.

$\mathcal{N}'(N) < 0$. The reasoning is quite simple. If one sorts the assets according to the expected over-performance then as we add assets starting at the best performing to the worst the expected performance of the portfolio decreases. Increasing N can have positive effects on risk, due to diversification, which we are not considering. We approximate the dependence on the number of assets as linear

$$\mathcal{N} = 1 - \frac{N}{M}, \tag{5.10}$$

where M is the maximal number of assets for which over-performance can be achieved. The linear form is a first order approximation that can be thought of as ordering according to a linear regression of performance as a function of some predictor.

For the size dependence term $\mathcal{Q}(s)$ we have somewhat less economic intuition as to the proper functional form even whether it should be an increasing or decreasing function of size. One can argue that skill should increase with size since bigger funds can afford more and better skilled employees. The skill can also decrease with size as suggested by Stein (2002) if, for instance, the hierarchical structure of the fund inhibits efficient processing of soft information which could have been turned into profit. Chen et al. (2004) have found evidence for such losses which they termed structural losses. As a first approximation we choose to approximate it as

$$\mathcal{Q}(s) = s^{\delta}. \tag{5.11}$$

Using the above approximations for $\mathcal{Q}$, $\mathcal{N}$ and $\mathcal{L}$ (equations 5.9–5.11), the fund manager's skill is given by

$$\alpha(s, \lambda, N) = \alpha_0 s^{\delta} \left(1 + \log(1 + \frac{\lambda}{\Lambda})\right) \left(1 - \frac{N}{M}\right) \tag{5.12}$$

using the functional form of cost described in equation (5.8) the managerial fee is given by

$$f(s, \lambda, N) = \alpha_0 s^{\delta} \left(1 + \log(1 + \frac{\lambda}{\Lambda})\right) \left(1 - \frac{N}{M}\right) - c\lambda^{1+\beta} \left(\frac{s}{N},\right)^{\beta}. \tag{5.13}$$

which has six free parameters out of which four parameters describe alpha: α_0, δ, Λ and M, and two parameters describe the cost: c and β.

Profit maximization

As we now show, the maximization will result in non-trivial size dependence of the fees, the number of assets and the turnover rate. Plugging in the expressions for C and α in

Model	α_0 (bps)	β	c (bps)	δ $[10^{-3}]$	Λ $[10^{-2}]$	M	MAR
$\mathcal{M}_a$	60 (58,62)	0.15 (0.14,0.16)	67 (55,73)	-31 (-33,-28)	53 (50,59)	1919 (1819,2167)	0.38
$\mathcal{M}_b$	66 (57,71)	[0.5] –	71 (51,84)	-35 (-41,-27)	59 (50,68)	821 (721,986)	0.43

Table 5.2: A summary of the parameter values resulting from the best fit of the model to the data. The values in parenthesis represent the 95% confidence interval for each parameter obtained using the bootstrap technique described in Appendix 5.B. The fit was conducted by minimizing the median absolute (relative) residual (MAR). A parameter that was held constant has its value in square brackets.
$\mathcal{M}_a$. The MAR fit of our model described in Section 5.5.
$\mathcal{M}_b$: similar to $\mathcal{M}_a$ but holding $\beta = 0.5$ fixed.

equations (5.8) and (5.12) into equation (5.3), we get that the fund's profit is given by

$$s\, f(s,\lambda,N) = \alpha_0 s^{1+\delta}\left(1+\log(1+\frac{\lambda}{\Lambda})\right)\left(1-\frac{N}{M}\right) - c\lambda^{1+\beta}\frac{s^{1+\beta}}{N^{\beta}}. \tag{5.14}$$

The first order conditions, described in equation (5.6), for maximizing the profit (5.14) results in the following set of equations

$$\begin{aligned} \lambda_{opt}^{\beta}(\Lambda+\lambda_{opt}) &= \frac{\alpha_0}{c(1+\beta)}\,\frac{N_{opt}^{\beta}(1-N_{opt}/M)}{s^{\beta-\delta}} \\ N_{opt}^{1+\beta} &= \frac{\beta c M}{\alpha_0}\,\frac{\lambda_{opt}^{1+\beta}\, s^{\beta-\delta}}{1+\log\left(1+\lambda_{opt}/\Lambda\right)}. \end{aligned} \tag{5.15}$$

There is no closed form solution to these equations but for a given size they can be easily solved numerically. The optimal fees can then be calculated by using equation (5.7). Thus, we have a reduced form model that describes the behavior of the fund as given by $\lambda_{opt}(s)$, $N_{opt}(s)$ and $f_{opt}(s)$. In the next section we compare the predictions of this model to the empirical data and show that the model fits the data extremely well.

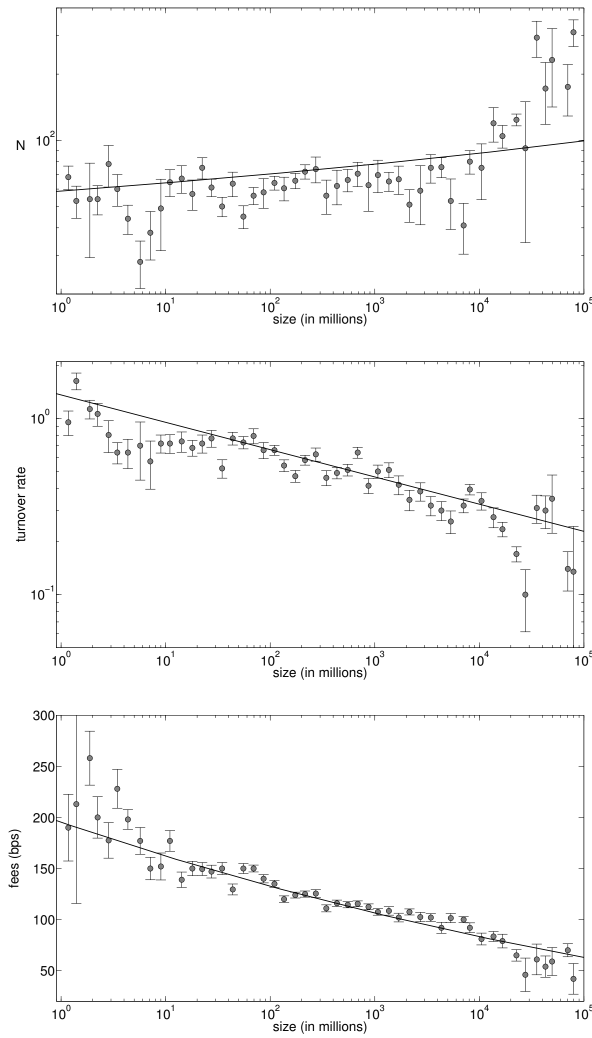


Figure 5.4: The fund behavior as a function of size is compared to the model predictions. The data (◦) is compared to the best fit (solid line), with parameter values given in Table 5.2 for model $\mathcal{M}_a$. In all three panels the data was binned into 50 exponentially spaced bins. For each bin the median is quoted and the error bar corresponds to the mean absolute deviation in the bin divided by the square root of the number of observations in the bin. **Top:** The number of assets in the portfolio (◦) is plotted as a function of the fund size (in millions). The data is compared to the MAR solution for N_{opt}. **Center:** The turnover rate (◦) is plotted as a function of the fund size (in millions). The data is compared to the MAR solution for λ_{opt}. **Bottom:** The expense ratio (◦) is plotted as a function of the fund size (in millions). The data is compared to the MAR solution for $f(\lambda_{opt}, N_{opt}, q)$.

5.6 Empirical verification of the model

In this section we show that the observed fund behavior can be described by our model; by fitting $N_{opt}(s)$, $\lambda_{opt}(s)$ and $f(\lambda_{opt}, N_{opt}, s)$ to the data according to equations (5.14) and (5.15). The fitting procedure is carried out such that we find the parameter values that minimized the median absolute residual (MAR) [10]. The result of the fitting procedure is an estimation of the parameters α_0, β, c, δ, Λ and M. It is important to emphasize that we are fitting with respect to all three observables N, λ and f (6270 data points in total). The fitting procedure was carried out twice where in the second time we hold $\beta = 0.5$ fixed. The different fits are denoted as schemes $\mathcal{M}_a$ and $\mathcal{M}_b$ respectively. The resulting parameter values for, α_0, β, c, δ, Λ and M are described in Table 5.2.

Scheme $\mathcal{M}_a$ corresponds to the most general fit where all 6 parameters are allowed to vary. This fits the data very well as can be seen in Figure 5.4 where we compare the best fits to the binned data. To try and gauge the capability of the model to describe the data we look not only on the resulting goodness of fit but also on the parameter values given their economic meaning. For instance, the number of assets above which no over-performance is possible is $M = 1919$ which is about a quarter of the approximately 8000 stocks traded in the NYSE [11]. This is a relatively large number but is still reasonable.

An important consequence of this fitting procedure is that it results in an estimation of the cost function. The fitted value for the cost function scale parameter is $c \approx 66$ (bps), a value which is close to the estimates reviewed in Chapter 4. The exponent of the cost function was found to be $\beta \approx 0.15$, a value that is close to some of the estimates given in Table 4.1 in Chapter 4. This gives us an estimation of the costs incurred by a fund of a given size, with N assets, and turnover rate λ.

There is an increasing amount work, both empirical and theoretical, that argues for the square root law, i.e. $\beta = 0.5$, a value which is significantly different than our estimation. To this end we carried out the fitting procedure again while holding $\beta = 0.5$ fixed, which we denote sceme $\mathcal{M}_b$. As can be expected holding one parameter fixed results in an inferior

[10]We chose to use a median absolute residual (MAR) optimizing scheme because of the non-normal nature of the residuals. The residuals distribution is heavier tailed than a normal and skewed such that most of the large fluctuations tend to be for funds with large N and large λ. The fitting procedure is described in more detail in Appendix 5.B.

[11]Taken form the NYSE website www.nyse.com.

fit when considering the value of the MAR (as can be seen in Table 5.2). However, the resulting parameter estimation of $M \approx 821$ has a greater appeal on an intuitive level and therefore might be an argument for considering scheme $\mathcal{M}_b$ even though it offers an inferior fit. We expand on this point later on.

In both fitting schemes $\mathcal{M}_a$ and $\mathcal{M}_b$ our parameter estimations resulted in a weak size dependence of skill with $\delta \approx -0.03$. This value is negative on a statistically significant level. What this means is that holding all else equal funds have a decreasing ability to create pre-cost excess returns. For instance a large fund with net assets of 100 billion USD can produce 30% less excess returns compared to a similar fund that manages solely 1 million USD. This is in line with the results of Chen et al. (2004) who call this organizational loss.

To conclude, we have shown that our proposed model is capable of explaining the observed mutual fund behavior. Moreover, the resulting parameter values, summarized in Table 5.2, are sensible with respect to their economic content. In the following section we examine the implications of the results by estimating performance loss and trading costs incurred by funds of different sizes.

5.6.1 Estimating performance loss and costs

In this section we examine the implied costs and performance loss as a function of fund size. To do so we use the fitted parameters given in Table 5.2 to estimate the cost and alpha, the pre-transaction cost performance, as given by equations (5.8) and (5.12) for each of the funds in the data base. The results for each size decile are summarized in Table 5.3. What these results show is that larger funds suffer a performance loss relative to smaller funds, i.e. alpha is smaller for larger funds than for smaller ones.

For both fitting schemes $\mathcal{M}_a$ and $\mathcal{M}_b$ the alpha estimation exhibit similar a pattern. For $\mathcal{M}_a$ the estimated performance is 207 bps for the smallest decile and decreases to 162 bps for the largest decile. We conclude that there is a clear downward trend even though the values vary widely in each bin. We report similar results for $\mathcal{M}_b$, where funds in the smallest size decile have a median alpha of 212 bps which decreases to 162 for funds in the largest decile. Thus, due to the increase in size, the lowering of the turnover rate and diversification, the funds have a diminished ability to produce excess return. We denote this decrease in alpha as performance loss.

When examining the estimated costs the two fitting schemes differ qualitatively. For

$\mathcal{M}_a$ the estimated costs on the other hand do not seem to exhibit a trend and remain approximately 40 bps annually. What this means is that under this model funds manage to hold their (relative) costs fixed by diversifying and lowering the turnover rates. As we have shown this comes at the expense of performance.

For scheme $\mathcal{M}_b$ the results are quite different. Since we hold $\beta = 0.5$ fixed the estimated costs increase much more than the estimated costs using scheme $\mathcal{M}_a$ for which we fit $\beta = 0.15$. We estimate that funds in the smallest size decile have an annual cost of about 12 bps that increase dramatically to an annual cost of about 125 bps for the largest decile. Thus, two different fitting schemes $\mathcal{M}_a$ and $\mathcal{M}_b$ result fundamentally different expected behavior. While under $\mathcal{M}_a$ funds manage to control their cost they are not able to do so for $\mathcal{M}_b$ where the relative costs increase dramatically.

5.6.2 Maximal fund size

In this section we investigate how the profitability of a fund changes with the size of the fund. To study this expected behavior over a large range of fund sizes, we solve the profit maximization problem numerically using the fit parameters summarized in Table 5.2.

The resulting number of assets N and the turnover rate λ are presented graphically in Figures 5.5 and 5.6. The expected optimal behavior is non trivial; while the optimal turnover rate decreases monotonically towards zero, the optimal number of assets held N_{opt} increases initially but for large fund sizes it decreases and eventually goes to zero. What this means is that past a certain size a fund cannot sustain the costs while remaining an active investor[12]. As can be seen in Figures 5.7 and 5.8, the fees a fund charges decrease as it grows but the absolute profit increases. Thus, a profit maximizing manager would prefer to grow up to s_{max} in which point the fund profitability cannot be sustained.

The maximal size of a fund differs dramatically between fitting scheme $\mathcal{M}_a$ and $\mathcal{M}_b$; in fitting scheme $\mathcal{M}_a$ the maximal size of the fund is $s_{max} \approx 2000$ trillion US dollars. At that size the maximal fees are 30 bps yielding a yearly profit of 6 trillion USD. This size is so large that what it means is that funds can essentially grow indefinitely. However, at this size the fund is trading in very few assets and such a strategy may not be feasible because of the market cap of the stocks even if the turnover rate is very low. The maximal number

[12]Similar arguments for an optimal fund size have been discussed by Farmer (2002), where he has shown that for value strategies an optimal size s_{max} exists above which a strategy is less profitable.

decile	size (in millions)	model $\mathcal{M}_a$ cost (bps)	model $\mathcal{M}_a$ alpha (bps)	model $\mathcal{M}_b$ cost (bps)	model $\mathcal{M}_b$ alpha (bps)
1/10	5.28	31	207	12	212
	(1.2,10.6)	(4,131)	(118,286)	(1,97)	(120,295)
2/10	19	38	210	21	210
	(12,28)	(4,196)	(105,289))	(1,222)	(90,302)
3/10	45	41	209	35	211
	(30,62)	(5,179)	(119,279)	(2,262)	(111,290)
4/10	84	47	210	51	215
	(66,113)	(8,201)	(131,280)	(5,314)	(120,289)
5/10	156	37	193	42	198
	(120,200)	(4,138)	(104,259)	(4,251)	(95,266)
6/10	261	45	198	57	199
	(208,348)	(7,166)	(106,259)	(4,470)	(90,274)
7/10	488	39	189	69	191
	(364,611)	(3,151)	(91,253)	(2,440)	(69,260)
8/10	852	44	186	87	188
	(646,1146)	(5,178)	(100,257)	(4,850)	(58,267)
9/10	1.9 $[10^3]$	41	182	107	181
	(1.2,3.2) $[10^3]$	(4,183)	(91,254)	(5,997)	(78,263)
10/10	7.4	34	162	125	162
	(3.6,66.7) $[10^3]$	(4,108)	(80,215)	(6,700)	(49,223)

Table 5.3: A summary of the estimated skill and costs of funds using fitting schemes $\mathcal{M}_a$ and $\mathcal{M}_b$. The cost represents the estimated early trading cost incurred by the fund given by (5.8). alpha represents the estimates of the (pre-cost) yearly excess performance given by (5.12). The results were obtained by estimating the skill and costs using equations (5.12) and (5.8) for each of the funds in the dataset using the best fit parameter values given in Table 5.2. For each of the models we bin the results according to size deciles. For each decile the entry in the table represents the median quantity and the parentheses represent the 5% and 95% quantiles for the quantities in that decile.

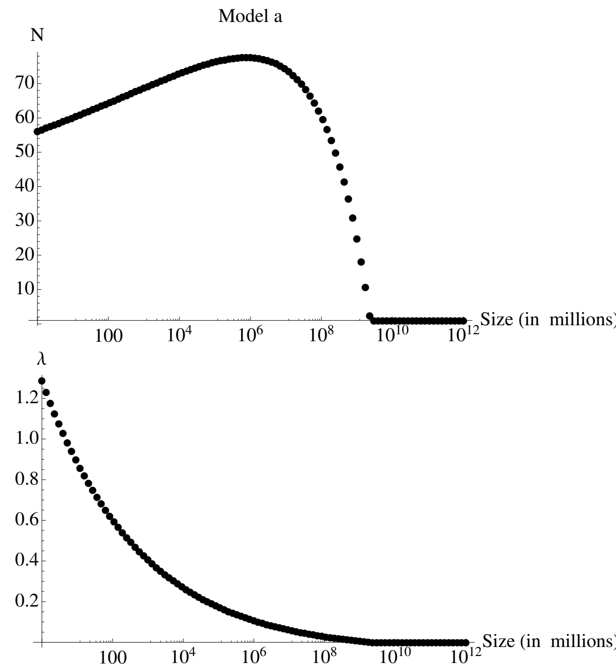


Figure 5.5: Here we show the optimal fund behavior as a function of size as resulting from fitting scheme $\mathcal{M}_a$. The optimal number of assets N_{opt} and the optimal yearly turnover rate λ_{opt} as a function of fund size (in millions of USD). The values were calculated by by maximizing $sf_{opt}(s)$ numerically, for a given size, using equation (5.14) with the fit parameters for fitting scheme $\mathcal{M}_a$ summarized in Table 5.2.

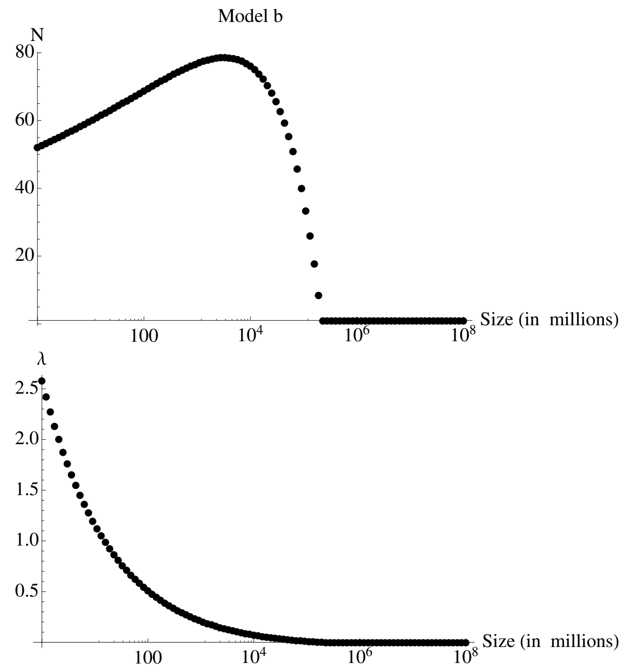


Figure 5.6: Here we show the optimal fund behavior as a function of size as resulting from fitting scheme $\mathcal{M}_b$. The optimal number of assets N_{opt} and the optimal yearly turnover rate λ_{opt} as a function of fund size (in millions of USD). The values were calculated by by maximizing $sf_{opt}(s)$ numerically, for a given size, using equation (5.14) with the fit parameters for fitting scheme $\mathcal{M}_b$ summarized in Table 5.2.

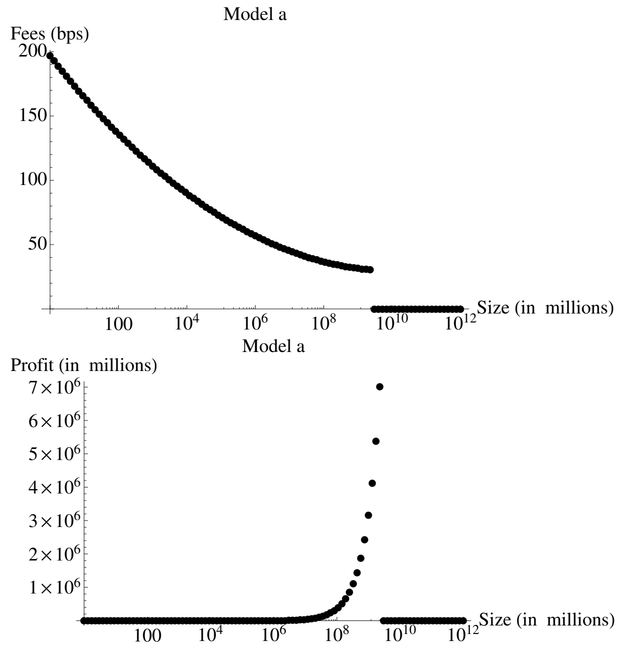


Figure 5.7: Here we show the maximal profitability of a fund as a function of size as resulting from fitting scheme $\mathcal{M}_a$. The maximal (yearly) fees $f_{opt}(s)$ and the corresponding yearly profit (in millions of USD) as a function of fund size (in millions of USD). The values were calculated by by maximizing $sf_{opt}(s)$ numerically, for a given size, using equation (5.14) with the fit parameters for fitting scheme $\mathcal{M}_a$ summarized in Table 5.2.

of assets is reached at a fund size of approximately 1 trillion USD after which the number of assets decreases to zero passing 50 assets at 200 trillion. Realistically due to market cap limitation and diversification needs, active fund size cannot exceed 1 trillion USD. Still this is an order of magnitude larger than the largest funds existing today. When using fitting scheme $\mathcal{M}_b$ the trading costs are much higher due to the $\beta = 0.5$ constraint. What this means is that the maximal size of funds is smaller on average. As can be seen in Figure 5.8, the maximal size of a fund, under this fitting scheme, is $s_{max} \approx 200$ billion USD, which is on the order of magnitude of the largest funds existing today.

As an example, the size of the largest fund in our sample is approximately 100 billion USD. This fund has an average expense ratio of 70 bps, an average turnover rate of 0.23 and trades approximately 200 assets. Except for the number of assets which is a bit high, our model under fitting scheme $\mathcal{M}_a$ is able to predict fund behavior pretty well. A summary comparison of the full data set is given in Table 5.1.

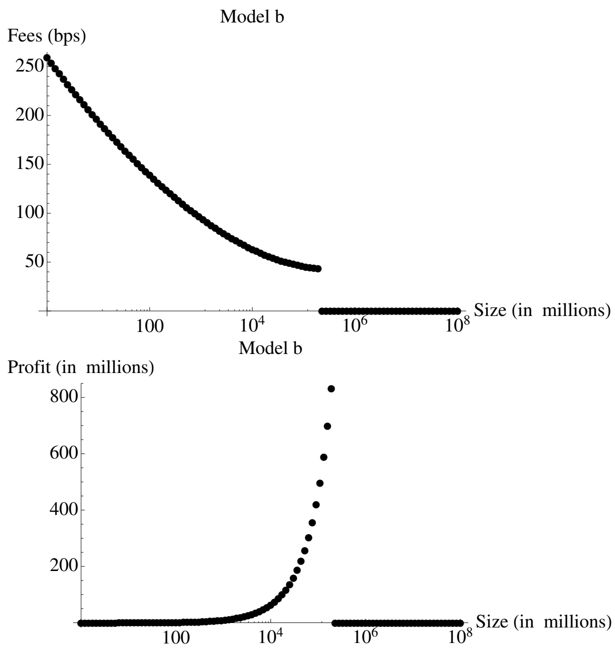


Figure 5.8: Here we show the maximal profitability of a fund as a function of size as resulting from fitting scheme $\mathcal{M}_b$. The maximal (yearly) fees $f_{opt}(s)$ and the corresponding yearly profit (in millions of USD) as a function of fund size (in millions of USD). The values were calculated by by maximizing $sf_{opt}(s)$ numerically, for a given size, using equation (5.14) with the fit parameters for fitting scheme $\mathcal{M}_b$ summarized in Table 5.2.

We can conclude that our model predicts that the maximal size of an actively traded fund is larger than what we observe today. Since the profit increases as the fund gets bigger, fund managers have an incentive to grow as much as they can. When they approach the maximal size, they can either limit the size of a fund, or turn into an passively managed fund for which no over-performance is expected.

5.7 Conclusions

Mutual fund performance displays a paradoxical lack of persistence given the diminishing returns to scale induced by trading costs. Moreover, we have shown that mutual fund behavior, as given by the fees they charge, the number of positions and the turnover rate, displays a dependence on size. To solve the persistence paradox while explaining mutual fund behavior we have proposed a framework for mutual fund behavior. We argue that it is not investor behavior but rather managerial behavior that is responsible for ensuring the lack of persistence. Fund managers optimize their behavior in a way that ensures that all funds have the same expected after cost returns. An important point is that our model allows for the industry to be out of equilibrium and agrees with the diffusion type dynamics of the random exit and entry process proposed in Schwarzkopf and Farmer (2010a).

A point worth iterating is that in our model diseconomies of scale are caused by both alpha and the costs; as funds grow their costs increase and their performance (alpha) decreases. A small 5.2 million USD fund, represented by the median values for the first size decile, has an estimated alpha of 207 bps, it incurs a cost of 32 bps and charges an investor 175 bps of fees. The after cost annual performance of such a fund is approximately -16 bps which is well within the noise and can be treated as zero. If the fund were to keep the same number of stocks, the same turnover rate and charge the same amount of fees as it grew to be a large 7.4 billion USD fund, represented by the tenth size decile, then the costs would increase by about 200% to 96 bps and alpha would decrease by approximately 20% to 166 bps. This would mean that the fund would have an after cost performance of -105 bps, which is a significant underperformance. Thus, if funds were to keep their behavior unchanged as they grew they would become uncompetitive.

What we have shown in this work is that fund managers compensate for increasing costs and decreasing alpha by lowering their fees (while still increasing profit). In order to maximize their profit fund managers try to keep the fee decrease to a minimum by controlling costs through turnover rate reduction and diversification. The performance, alpha, is adversely affected by turnover rate reduction and diversification and acts as a constraint on the manager's profit maximization problem.

Using the best fit parameters for fitting scheme $\mathcal{M}_a$ and the summary statistics in Table 5.3, we estimate the contribution of λ, N, and f to the effort of the fund to remain competitive as it increases in size. If the fund was to compensate by solely reducing its fees is would need to reduce them by 20 bps more than it does, an amount that represents a significant profit loss. If on the other hand the fund was just to increase the number of assets from 53 to 84 our model predicts the costs would decrease by 7 bps and while decreasing alpha by 3 bps. Thus such weak diversification by itself reduces the under performance by a mere 4 bps. The decreasing of the turnover rate from 0.73 to 0.28 has a larger effect. The costs decrease by 64 bps, which is approximately the costs incurred by the small fund and the alpha decreases by 42 bps. Thus, lowering the turnover rate reduces the underperformance by 22 bps. None of the effects is enough to compensate and the fund manager adapts by changing both the number of assets and the turnover rate simultaneously such that the profits (fees) are maximized. Our results (and observations) suggest that diversification plays a relatively small role in the manager's strategy; [13] it accounts for just 9% of the cost reduction while contributing 3% to the alpha loss. The role of turnover rate decrease is much more dramatic; it accounts for 91% of the decrease in costs while contributing 47% to the alpha loss. Note, that 50% of the alpha loss is solely due to size.

5.A Data sets

We use the Center for Research in Security Prices (CRSP) survivor-bias-free mutual fund data set. We focus on the holdings data set containing holdings data for a subset of the mutual funds, reported at various dates for the years 2002 to 2009 (excluding the last quarter of 2009). The dataset contains data on 8487 funds.

For each fund in the holdings dataset we match the yearly turnover rate, the expense ratio and total net assets under management (TNA) corresponding to the reported date.

[13] The observation that funds only weakly increase their number of positions to accommodate growth was made previously by Edelen et al. (2007).

The TNA, reported monthly, was corrected for inflation with respect to June 2009's consumer price index[14]. The yearly turnover rate, reported yearly, is defined as the minimum of aggregated sales or aggregated purchases of securities, divided by the average 12-month Total Net Assets of the fund. The expense ratio is reported yearly as of the most recently completed fiscal year. The expense ratio is the ratio of total investment that shareholders pay for the funds operating expenses, which include 12b-1 fees and includes waivers and reimbursements.

Out of this universe of funds we consider only active equity funds, which we define as funds whose portfolio consists of at least 80% of cash and stocks and have a reported yearly turnover rate larger than zero. Out of these funds we consider only funds with a real TNA of at least 1 million USD. This reduces the universe of funds we consider to 1404 funds with a total of 2190 observations corresponding to holdings reported at the end of the second quarter of each year. For each observation we have the number of assets, the turnover rate and the expense ratio resulting in 6570 data points.

5.B Fitting procedure

Our model relates the size of the fund s to the expense ratio f, the number of assets N and the turnover rate λ through equation (5.15) which we rewrite here

$$\begin{aligned} \lambda_{opt}^{\beta}(\Lambda + \lambda_{opt}) &= \frac{\alpha_0}{c(1+\beta)} \frac{N_{opt}^{\beta}(1 - N_{opt}/M)}{s^{\beta-\delta}} \\ N_{opt}^{1+\beta} &= \frac{\beta c M}{\alpha_0} \frac{\lambda_{opt}^{1+\beta} s^{\beta-\delta}}{1 + \log(1 + \lambda_{opt}/\Lambda)}. \end{aligned} \tag{5.16}$$

Since there is no closed form solution (that we were able to get) for λ_{opt} and N_{opt} equation (5.16) has to be solved numerically. Given λ_{opt} and N_{opt} the predicted expense ratio f_{opt} is given by

$$f_{opt}(\lambda_{opt}, N_{opt}, q) = \alpha_0 s^{1+\delta} \left(1 + \log(1 + \frac{\lambda_{opt}}{\Lambda})\right) \left(1 - \frac{N_{opt}}{M}\right) - c\lambda_{opt}^{1+\beta} \frac{s^{1+\beta}}{N_{opt}^{\beta}}. \tag{5.17}$$

[14] The consumer price index (CPI) is reported online on the US Bureau of Labor Statistics website www.bls.gov/cpi/.

Thus, for a fund of size s we have a prediction for $\hat{\lambda}$, $\hat{N}$ and $\hat{f}$ for that fund that depends on the six parameters α_0, β, c , δ, M and γ. Our goal in the fitting procedure is to estimate the values of the six parameters such that our estimators best describe the reported λ_i, N_i and f_i for fund i with a reported size s_i.

The fit is carried out for the 2190 observations in the aggregated dataset in the following manner: For each fund, using the reported s_i we solve for the estimators $\hat{\lambda}_i$, $\hat{N}_i$ and $\hat{f}_i$ and calculate the relative residuals vector res_i

$$\mathrm{res}_i = \left[\frac{\hat{\lambda}_i - \lambda_i}{\lambda_i}, \frac{\hat{N}_i - N_i}{N_i}, \frac{\hat{f}_i - f_i}{f_i}\right]. \tag{5.18}$$

We then follow to build the aggregated residual vector $\mathrm{res} = [\mathrm{res}_1, \mathrm{res}_2, \mathrm{res}_3 \ldots]$, which has $3 \times 2190 = 6570$ entries. We use relative residuals since N, λ and f have different scales. Finally, the parameters are estimated by minimizing the median absolute residual (MAR), which is a robust statistic. We use a robust statistic because of the non-normal nature of the residuals, which are better described by a heavy tailed skewed distribution. The heavy tailed nature of the residuals is due to due to the heavy tailed distribution of the number of assets and of the turnover rates.

The confidence bounds for the parameter estimation were calculated using a bootstrapping technique. The fitting procedure was carried out 100 times for a random subset of the observations. The subset is constructed by incorporating each point with a probability that ensures that on average the subset contains 500 points. The confidence intervals are then calculated using the quantiles of the resulting parameter estimations.

5.C Mutual fund trading costs' functional form

We estimate the amount of trading costs a fund incurs using the market impact models described in Chapter 4 as a guideline. Neglecting fixed costs the trading cost of trading a volume v in a given asset is given by the market impact

$$\chi(v) = c\, v^{1+\beta}, \tag{5.19}$$

where v is the dollar value of the trade. If a fund of size s trades in only a single asset and changes a position once a year then it trades twice a year a volume $v = s$ such that the trading costs are given by

$$s_t C_t(s_t) = \chi(q_t) = c\, s_t^{1+\beta}. \tag{5.20}$$

In a more general case a fund has a turnover ratio λ, defined as the minimum of total sales or purchases of the asset in a year divided by the average fund size. If the turnover is smaller than unity $\lambda < 1$ then the fund traded only a part of its assets, i.e. traded a volume $v = \lambda s$. If the turnover is larger than unity then we can think of it as if the fund traded $v = s$ (all its assets) λ times. Written more quantitatively, the trading cost for a fund of size s and a turnover rate of λ is given by

$$s_t C_t(s_t, \lambda) = \begin{cases} \lambda\, \chi(s_t) & \text{if} \quad \lambda \geq 1, \\ \chi(\lambda\, s_t) & \text{if} \quad \lambda < 1. \end{cases} \tag{5.21}$$

Using the definition of χ in equation (5.19) yields

$$q_t C_t(s_t, \lambda) = \begin{cases} c\, \lambda\, s_t^{1+\beta} & \text{if} \quad \lambda \geq 1, \\ c\, (\lambda s_t)^{1+\beta} & \text{if} \quad \lambda < 1. \end{cases} \tag{5.22}$$

If we assume that the traded volume is divided equally between N assets, i.e. if the total volume traded by the fund is v then it trades N assets with a volume of v/N in each. Under such conditions we write the impact as

$$s_t C_t(s_t, \lambda, N) = \begin{cases} c\, \lambda\, \frac{s_t^{1+\beta}}{N^\beta} & \text{if} \quad \lambda \geq 1, \\ c\, \frac{(\lambda s_t)^{1+\beta}}{N^\beta} & \text{if} \quad \lambda < 1. \end{cases} \tag{5.23}$$

For most funds (about 80% of the funds in the data set) the turnover rate is below 1. Thus we approximate the turnover rate as

$$C(s, \lambda, N) = c\, \lambda^{1+\beta} \left(\frac{s}{N}\right)^\beta. \tag{5.24}$$

5.D Implications of an infinite pockets assumption

Under the Berk and Green (2004) model the optimal size of a fund s^* (without indexing) is given by a demand that the marginal benefit of investing in a fund is equal to the expected

$\mu = \mathrm{E}[\phi]$	$\sigma = \sqrt{\mathrm{Var}[\phi]}$	$\bar{\phi}$
6.5%	6%	3%

Table 5.4: Parameter values for the prior distribution as used in Berk and Green (2004). $\bar{\phi}$ is the mean prior under which a fund's size is too small for it to extract the necessary rents to survive.

excess return it is believed to be able to deliver

$$\frac{\partial}{\partial s}\left[sC(s)\right] = \phi, \tag{5.25}$$

where ϕ is the expected excess return. For a cost function of the form

$$C(s) = c\ s^{\beta} \tag{5.26}$$

the optimal size of a fund s given by

$$s^* = \left(\frac{\phi}{(1+\beta)c}\right)^{1/\beta}. \tag{5.27}$$

By taking indexing into account the size of a fund s is given by

$$s(\phi) = \frac{s^*\phi - s^*C(s^*)}{f} = \frac{\beta}{c^{1/\beta}(1+\beta)^{(1+\beta)/\beta}}\frac{\phi^{(1+\beta)/\beta}}{f}. \tag{5.28}$$

Using the assumption that the priors are normally distributed $\phi \sim N(\mu, \sigma)$ with a mean of μ and a standard deviation σ the average size of a fund is given by

$$\mathrm{E}[s] = \frac{\int_{\bar{\phi}}^{\infty} s(\phi)\exp\left[-\frac{(\phi-\mu)^2}{2\sigma^2}\right]}{\int_{\bar{\phi}}^{\infty}\exp\left[-\frac{(\phi-\mu)^2}{2\sigma^2}\right]}. \tag{5.29}$$

The mean prior under which a fund's size is too small for it to extract the necessary rents is given by $\bar{\phi}$.

To check the plausibility of the resulting sizes we solve equation (5.29) for the parameter values used by Berk and Green (2004) with different values of β and c. The results are summarized in Figure 5.9. The values of c can be interpreted as the yearly performance loss for a fund of a size of a million USD, e.g. $c = 10^-3$ corresponds to a loss of 10 basis

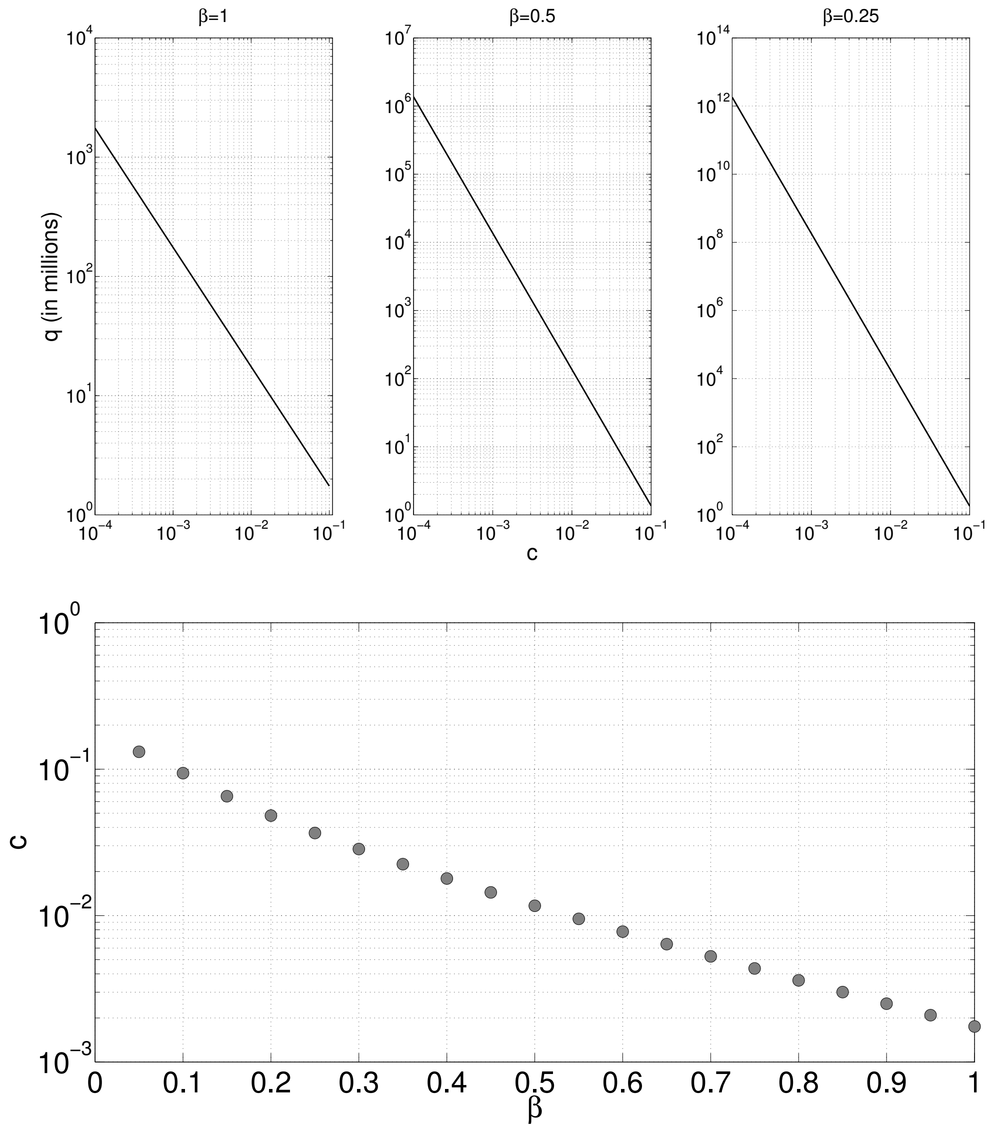


Figure 5.9: **Top**: the average fund size given by (5.29) for different parameter values of the cost function (5.26) and prior distribution parameters described in Table 5.D. **Bottom**: For each value of β in equation (5.26) we calculate the corresponding c such that the average size given by equation (5.29) is 100 million USD.

points a year. What we see is that for realistic values of c and $\beta = 1$ the fund size is larger than the observed mean of ≈ 100 million USD but not by much. However, recent studies suggest that β is far smaller than one and is in the range of $\beta \approx 0.5$ for which the average sizes become unreasonably large. To stress the connection between the resulting average size of a fund and the parameters of the cost function, we plot the value of c as a function of β that ensures that the average size is 100 million USD. One can see that indeed the resulting values of c are unreasonable.

If we take diversification and turnover rate into account, as described in Appendix 5.C, the for the cost function given by equation (5.8) the optimal size s^* is given by

$$s^*(N, \lambda, \alpha) = \begin{cases} \frac{N}{\lambda^{1/\beta}} s^* & \text{if} \quad \lambda \geq 1, \\ \frac{N}{\lambda^{(1+\beta)/\beta}} s^* & \text{if} \quad \lambda < 1. \end{cases}, \tag{5.30}$$

where s^* is given by equation (5.27) and similarly the size would be given by

$$s(N, \lambda, \alpha) = \begin{cases} \frac{N}{\lambda^{1/\beta}} s & \text{if} \quad \lambda \geq 1, \\ \frac{N}{\lambda^{(1+\beta)/\beta}} s & \text{if} \quad \lambda < 1. \end{cases}, \tag{5.31}$$

where s is given in equation (5.28). Therefore if the average fund holds $N \approx 100$ assets and a turnover of $\lambda \approx 1$ then the average size $\mathrm{E}[s]$ would be 100 times more than what is calculated in Figure 5.9 using equation (5.29). This makes the average size of a fund unreasonably large and the size of the whole industry several orders of magnitude larger than it really is. The infinite investor pockets assumption seems to be unreasonable.

We conclude that if this was indeed the sole mechanism compensating for fund over-performance then for any reasonable set of β and c parameters the resulting fund sizes are extremely large compared to what we observe. What this means is that the Berk and Green model would predict an industry with funds of an average size of more than a billion dollars. So either fund over-performance is not saturated and we should observe persistence, or that this is not the entire story.

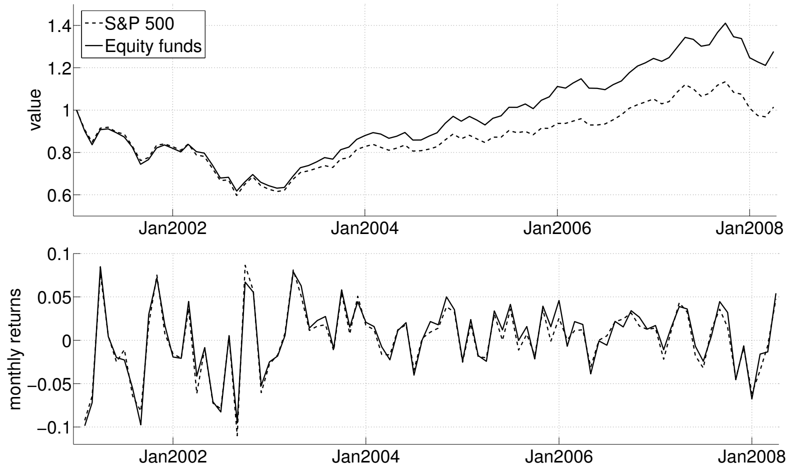


Figure 5.10: The aggregated value-weighted (VW) performance of the mutual funds in the data set are compared to the performance of the S&P 500 index. **Top:** the value of 1 USD invested, at the end of 2000, in the VW aggregated funds compared to 1 USD invested in the S&P 500 index. **Bottom:** The monthly VW return of the mutual funds compared to the monthly return on investing in the S&P 500.

5.E Mutual fund performance

5.E.1 Mutual fund skill and over-performance

One of our goals in this paper is to explain the observed lack of size dependence in mutual fund performance. Our assumption about the competitive nature of the market, i.e. that funds must remain competitive, leads to a natural definition of performance; the return on investment for a mutual fund investor. This return is defined as the change in net asset value (NAV) of the mutual fund and is reported monthly in the data set[15]. Similarly, we define skill (α) as the excess returns above a benchmark, the S&P 500 index for instance. This definition of skill differs from the approach used by some researchers, such as Fama and French (2010). They investigate skill by decomposing the returns using three factor (Fama and French, 1993) and four factor (Carhart, 1997) models. It is hard to believe that most investors use such metrics to compare fund performances and it is even harder to believe that investors can actually reproduce the market performance defined by these models. It is much more reasonable to assume that an investor compares the fund's performance to

[15] A more detailed description of the data set can be found in Appendix 5.A.

what he can easily access which is the market performance as given by the S&P 500.

Under our simplistic definitions the aggregate skill is deduced by comparing the aggregate fund performance to the benchmark. To this end, we compute the value-weighted (VW) monthly returns for equity funds in the period 2002 to March 2008. In Figure 5.10 the VW aggregate monthly performance of the industry as a function of time is compared to the S&P 500 index returns[16]. To visualize the comparison we use the aggregated returns to compare the value of 1 USD invested in the aggregated VW fund industry compared to the S&P 500 index. The results seem to suggest that α, as we defined it, is positive. Indeed there are periods were the mutual fund industry underperforms the market but there are also periods of significant over-performance. In the period at hand we are better of investing in the VW funds. This is in contradiction with Fama and French (2010) who conclude that the after costs performance is below the benchmark. This investigation of α is by no means rigorous, but what is important to take from this is that fund performance is not necessarily below the benchmark and funds need to actively trade to remain competitive.

5.E.2 Performance VS. size

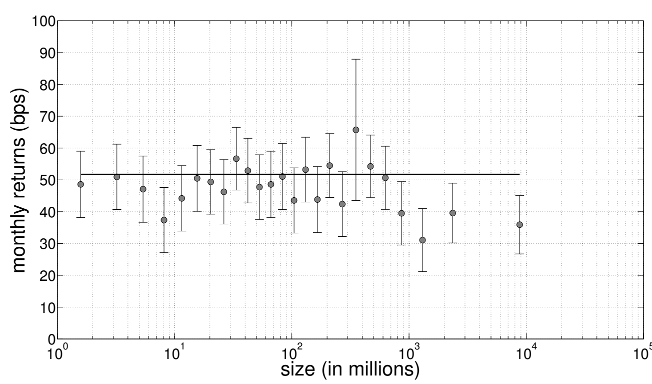


Figure 5.11: The monthly (log) returns as a function of fund size. The aggregated monthly data is binned into 25 equal occupation bins according to size. In each bin the mean size and mean return are quoted. The error bars represent the 2 sigma confidence interval for the mean in each bin. The binned data is compared to a constant 57 bps return.

[16]The S&P 500 monthly returns were calculated using the monthly index values available online at www.finance.yahoo.com.

We are interested in testing whether fund performance depends on size or stated differently, do larger funds have, on average, different performance than smaller funds. Our assumption, based on previous work (Schwarzkopf and Farmer, 2010b,a), is that performance does not depend on size. We cannot choose a better performing fund, on average, just by looking at its size. To test this we used cross-sectional linear regressions. As a first test we regressed the entire aggregated data of monthly returns $r_{it} = a_0 + a_1 \log_{10}(s_{it})$, where r_{it} is the monthly return of fund i at time t and s_{it} is its size (in millions). The resulting parameter values (in bps) are $a_1 = -0.96\,(-1.98, 0.05)$ and $a_0 = 51.7\,(46.7, 56.7)$, where the values in parenthesis represent the 95% confidence bounds. This result means that a_1 is not statistically significantly different than zero and the hypothesis of size independent performance cannot be rejected. To visualize this, we plot the binned aggregated data as a function of size in Figure 5.11.

Figure 5.10 suggests that the overall performance, relative to the benchmark, changes from month to month. To check if this is also the case for the size dependence we regressed the cross-sectional data for each month separately. The results are summarized in Table 5.5. The results are interesting by their own merits. In some months the performance seems to increase with size, in some months it decreases while in others it seems not to depend on size. Furthermore, when we aggregate the monthly data for each year, the results mainly suggest that performance is indeed independent of size in agreement with market efficiency.

	2001	2002	2003	2004	2005	2006	2007	2008	2009
Jan.	-36.9	-13.3	1.10	-1.07	-3.45	-4.78	-0.80	8.03	–
	(-49.1,-24.8)	(-20.5,-6.1)	(-2.09,4.28)	(-3.76,1.62)	(-7.79,0.90)	(-8.19,-1.38)	(-3.26,1.66)	(2.86,13.20)	–
Feb.	-7.06	-3.09	9.24	-7.08	8.87	-2.64	1.61	-6.15	–
	(-15.0,0.9)	(-7.57,1.39)	(5.33,13.16)	(-11.0,-3.2)	(6.24,11.49)	(-6.12,0.85)	(-1.31,4.53)	(-10.4,-1.9)	–
Mar.	21.0	-17.0	6.17	11.4	3.19	-0.14	5.26	3.62	–
	(12.7,29.2)	(-24.1,-9.8)	(1.13,11.21)	(5.51,17.20)	(-1.16,7.54)	(-4.83,4.54)	(2.57,7.95)	(-0.51,7.76)	–
Apr.	1.71	0.71	3.39	5.19	-0.28	8.36	-0.69	–	–
	(-3.50,6.91)	(-5.08,6.51)	(-1.54,8.33)	(1.79,8.60)	(-4.83,4.28)	(4.32,12.40)	(-3.65,2.27)	–	–
May.	1.20	-14.1	-7.65	2.05	-4.51	2.67	0.97	–	43.9
	(-4.35,6.75)	(-20.1,-8.1)	(-10.6,-4.7)	(-0.77,4.86)	(-8.19,-0.82)	(0.15,5.18)	(-2.95,4.90)	–	(25.3,62.4)
Jun.	6.98	0.97	-9.57	-0.77	-3.05	-2.93	6.53	119	–
	(1.23,12.73)	(-5.26,7.19)	(-14.0,-5.1)	(-5.90,4.36)	(-6.50,0.41)	(-8.72,2.86)	(1.60,11.46)	(-150,388)	–
Jul.	-16.8	5.12	-5.42	-1.35	-0.98	1.76	-2.46	–	2.87
	(-23.2,-10.5)	(0.81,9.44)	(-10.2,-0.6)	(-5.71,3.01)	(-4.89,2.92)	(-1.00,4.52)	(-6.09,1.18)	–	(-10.0,15.8)
Aug.	-1.64	-6.58	-9.28	-8.80	-8.96	-0.69	5.76	-41.7	-70.8
	(-10.4,7.1)	(-12.9,-0.3)	(-13.4,-5.1)	(-12.9,-4.8)	(-13.8,-4.1)	(-4.05,2.67)	(0.19,11.34)	(-719,636)	(-163,21)
Sep.	1.98	17.8	-3.88	-2.49	6.97	-1.93	4.85	–	-0.50
	(-5.52,9.49)	(9.84,25.78)	(-8.30,0.53)	(-5.43,0.46)	(3.70,10.23)	(-4.64,0.77)	(-0.70,10.39)	–	(-10.2,9.2)
Oct.	9.69	3.96	1.49	-5.82	-0.19	1.33	5.80	–	–
	(3.14,16.24)	(-2.63,10.55)	(-1.89,4.87)	(-9.44,-2.20)	(-3.45,3.06)	(-1.83,4.49)	(1.66,9.94)	–	–
Nov.	-6.89	-7.26	-2.22	1.15	-4.63	0.94	16.5	78.8	–
	(-11.7,-2.1)	(-14.9,0.4)	(-7.27,2.83)	(-2.51,4.80)	(-9.25,-0.01)	(-2.29,4.16)	(-16.5,49.5)	(-31.9,189.4)	–
Dec.	-2.29	-1.12	-0.57	-2.40	-11.1	1.59	5.24	–	–
	(-8.27,3.69)	(-4.29,2.05)	(-4.33,3.18)	(-5.96,1.16)	(-17.1,-5.1)	(-1.13,4.32)	(-0.94,11.41)	–	–
Yearly	-7.00	-6.80	-1.84	-1.85	-0.89	0.45	2.80	0.88	-1.29
	(-10.7,-3.3)	(-10.1,-3.5)	(-3.82,0.14)	(-3.74,0.03)	(-2.81,1.02)	(-1.08,1.97)	(-0.72,6.31)	(-3.30,5.06)	(-10.5,7.9)

Table 5.5: The cross-sectional regression of monthly returns versus the logarithm of the fund size (in millions). The entries represent the coefficient $a_1(t)$ for a regression $r(t) = a_0(t) + a_1(t)\log_{10}(s(t))$, where t represents the given corresponding month, r is the monthly return and s is the fund size (in millions). The values in parentheses represent the 95% confidence bounds. The bottom row represents regression coefficient for the aggregated yearly data. Missing entries correspond to months where there were less then 10 observations. A linear regression of the entire aggregating yields $a_1 = -0.96\,(-1.98, 0.05)$ and $a_0 = 51.7\,(46.7, 56.7)$ in bps.

Chapter 6

The cause of universality in growth fluctuations

Phenomena as diverse as breeding bird populations, the size of U.S. firms, money invested in mutual funds, the GDP of individual countries and the scientific output of universities all show unusual but remarkably similar growth fluctuations. The fluctuations display characteristic features, including double exponential scaling in the body of the distribution and power law scaling of the standard deviation as a function of size. To explain this we propose a remarkably simple additive replication model: At each step each individual is replaced by a new number of individuals drawn from the same replication distribution. If the replication distribution is sufficiently heavy tailed then the growth fluctuations are Levy distributed. We analyze the data from bird populations, firms, and mutual funds and show that our predictions match the data well, in several respects: Our theory results in a much better collapse of the individual distributions onto a single curve and also correctly predicts the scaling of the standard deviation with size. To illustrate how this can emerge from a collective microscopic dynamics we propose a model based on stochastic influence dynamics over a scale-free contact network and show that it produces results similar to those observed. We also extend the model to deal with correlations between individual elements. Our main conclusion is that the universality of growth fluctuations is driven by the additivity of growth processes and the action of the generalized central limit theorem.

6.1 Introduction

Recent research has revealed surprising properties in the fluctuations in the size of entities such as breeding bird populations along given migration routes, U.S. firm size, money invested in mutual funds, GDP, scientific output of universities, and many other phenomena[1].

[1]The anomalous growth fluctuations were observed for bird population (Keitt and Stanley, 1998), U.S. firm size (Stanley et al., 1996; Amaral et al., 1997; Bottazzi and Secchi, 2003a; Matia et al., 2004; Bottazzi

This is illustrated in Figures 6.1 and 6.2. The first unusual property is in the logarithmic annual growth rates g_t, defined as $g_t = \log(N_{t+1}/N_t)$, where N_t is the size in year t. As seen in the top panel of Figure 6.1, all of the data sets show a similar double exponential scaling in the body of the distribution, indicating heavy tails. The second surprising feature is the power law scaling of the standard deviation σ with size, as illustrated in Figure 6.2. In each case the standard deviation scales as $\sigma \sim N^{-\beta}$ with $\beta \approx 0.3$.

These results are viewed as interesting because they suggest a non-trivial collective phenomena with universal properties. If the individual elements fluctuate independently, then (with a caveat we will state shortly) the standard deviation of the growth rates scales as a function of size with an exponent $\beta = 1/2$, whereas if the individual elements of the population move in tandem the standard deviation scales with $\beta = 0$, i.e. it is independent of size. The fact that we instead observe a power law with an intermediate exponent $0 < \beta < 1/2$ suggests that the individual elements neither change independently nor in tandem. Instead it suggests some form of nontrivial long-range coupling. Why should phenomena as diverse as breeding bird populations and firm size show such similar behavior? There is a substantial body of previous work attempting to explain individual phenomena, such as firm size or GDP [2]. However none of these theories has the generality to explain how this behavior could occur so widely.

The caveat in the above reasoning is the assumption that the fluctuations of the individual elements are well-behaved, in the sense that they are not too heavy-tailed. As we show in a moment, if the growth fluctuations of the individual elements are sufficiently heavy-tailed then the fluctuations of the population are also heavy tailed, even if there are no collective dynamics. Under the simple additive replication model that we propose the fluctuations in size are Levy distributed in the large N limit. This predicts a scaling exponent $0 < \beta < 1/2$ and the shape parameter of the Levy distribution predicts the value

and Secchi, 2005; Axtell et al., 2006), money invested in mutual funds (Schwarzkopf and Farmer, 2008), GDP (Canning et al., 1998; Lee et al., 1998; Castaldi and Dosi, 2009), scientific output of universities (Matia et al., 2005), and many other phenomena (Plerou et al., 1999a; Keitt et al., 2002; Bottazzi et al., 2007; Podobnik et al., 2008; Rozenfeld et al., 2008).

[2] R. (1931); De Fabritiis et al. (2003); Fu et al. (2005); Riccaboni et al. (2008); Simon and Bonini (1958); Ijiri and Simon (1975); Amaral et al. (1997b); Buldyrev et al. (1997); Amaral et al. (1998); Bottazzi (2001); Sutton (2001); Wyart and Bouchaud (2003); Bottazzi and Secchi (2003b); Gabaix (2009); Schweiger et al. (2007).

year	α	κ	c	μ
NABB	1.40	0.81	0.156	-0.037
Mutual funds	1.48	0.3	0.111	-0.015
Firms	1.53	0.80	0.16	-0.05

Table 6.1: The parameter values for fitting the data with a Levy distribution.

of β. We show here that this model provides an excellent fit to the data.

In the first part of this paper we develop the additive replication model and show that it gives a good fit to the data. Our analysis in the first part is predicated on the existence of a heavy-tailed replication distribution. In the second part of the paper we present one possible explanation for the heavy-tailed replication distribution in terms of stochastic influence dynamics on a scale-free contact network, and argue that such an explanation could apply to any of the diverse settings in which these scaling phenomena have occurred. This influence dynamics is an example of "nontrivial" collective dynamics. Thus, the process that generates the heavy tails in individual fluctuations may come from nontrivial collective dynamics even though the replication model does not depend on this.

6.2 The additive replication model

We assume an additive replication process: At each time step each individual element is replaced by k new elements drawn at random from a *replication distribution* $p(k)$, where $0 \le k < \infty$. An individual element could be a bird, a sale by a given firm, or the holdings of a given investor in a mutual fund. By definition the number of elements N_{t+1} on the next time step is

$$N_{t+1} = \sum_{j=1}^{N_t} k_{jt}, \tag{6.1}$$

where k_{jt} is the number of new elements replacing element j at time t. The growth G_t is given by

$$G_t = \frac{N_{t+1} - N_t}{N_t} = \frac{\sum_{j=1}^{N_t} k_{jt}}{N_t} - 1. \tag{6.2}$$

The simplest version of our model assumes that draws from the replication distribution $p(k)$ are independent; we later relax this assumption to allow for correlations.

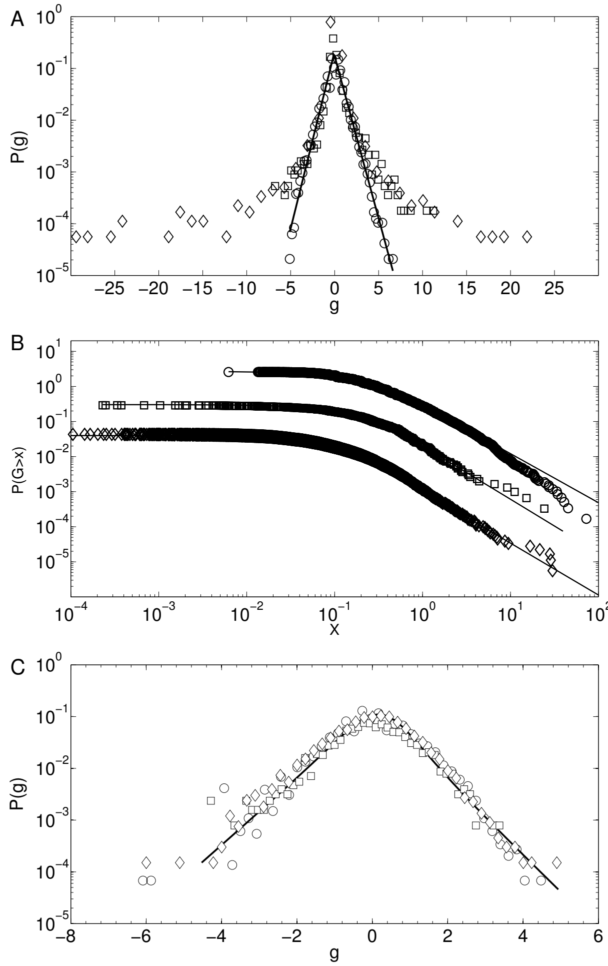


Figure 6.1: An illustration of how our theory reveals the underlying regularity in the distribution of growth fluctuations of highly diverse phenomena. The three data sets studied here are North American Breeding Birds (○), US firm sales (□) and US equity mutual funds (◇). The data is the same in all three panels, the only change is the presentation. **A**: *The traditional view.* Histograms of the logarithmic growth rates are plotted on semi-log scale, normalized such that the mean vanishes $\mathrm{E}[g] = 0$ and the variance is unity $\mathrm{Var}[g] = 1$. The collapse is good for the body of the distribution, revealing double-exponential scaling, but poor in the tails, where the three data sets look quite different. **B**: *Comparison to a Levy distribution.* The cumulative distribution $P(G > X)$ of relative growth rates for the three data sets are compared to fits to the Levy distributions predicted by our theory (solid curves) and plotted on double logarithmic scale (for positive X only). See Table 6.4 for parameter values. **C**: *Superior collapse onto a single curve when the data is scaled as predicted by our theory.* The empirical values of the relative growth G (rather than the logarithmic growth rate g) are normalized so they all have a scale parameter approximately one, as described in the text. In order to compare to the top panel, we plot the logarithmic growth g and compare to a Levy distribution (solid curve). This gives a better collapse of the data which works in the tails as well as the body.

Why might such a model be justified? First note that additivity of the elements is automatic, since by definition the size is the sum of the number of elements. The assumption that each element replicates itself in the next year amounts to a persistence assumption, i.e. that the number of elements in one year is linearly related to the number in the previous year, with each element influencing the next year independently of the others. We also assume uniformity by letting all elements have the same replication distribution $p(k)$. For the case of firms, for example, each sale in year t can be viewed as replicating itself in year $t+1$. This is plausible if the typical customer remains faithful to the same firm, normally continuing to buy the product from the same company, but occasionally changing to buy more or less of the product. For migrating birds this is plausible if the number of birds taking a given route in a given year is related to the number taking it last year, either because of the survival probability of individual birds or flocks of birds, or because individual birds influence other birds to take a given migration route.

6.3 Predictions of the model

Given that the size N_t at time t is known and the drawings from $p(k)$ are independent, the growth rate G_t is a sum of N_t I.I.D. random variables. Under the generalized central limit theorem (Zolotarev, 1986; Resnick, 2007), in the large N_t limit the growth $\mathcal{P}_G$ converges to a Levy skew alpha-stable distribution

$$\mathcal{P}_G(G_t|N_t) = N_t^{-\frac{1-\alpha}{\alpha}} L_\alpha^\kappa(G_t\, N_t^{-\frac{1-\alpha}{\alpha}}; c, \mu). \tag{6.3}$$

$0 < \alpha \leq 2$ is the shape parameter, $-1 \leq \kappa \leq 1$ is the asymmetry parameter, μ is the shift parameter and c is a scale parameter.

The normal distribution is a special case corresponding to $\alpha = 2$. This occurs if the second moment of $p(k)$ is finite. However, if the second moment diverges according to extreme value theory, under conditions that are usually satisfied, it is possible to write $p(k) \sim k^{-\gamma}$ for large k [3]. When $1 < \gamma < 3$ the Levy distribution has heavy tails that asymptotically scale as a power law with $P(G > x) \sim x^{-\alpha}$, where $\alpha = \gamma - 1$.

The additive replication process theory predicts power law behavior for $\sigma(N)$ and pre-

---

[3] Under extreme value theory there are distributions for which there is no convergent behavior; the power law assumes convergence.

Degree Distribution	β
$p(k) \sim k^{-\gamma}$ with $1 < \gamma < 3$	$\frac{\gamma-2}{\gamma-1}$
$p(k) \sim k^{-\gamma}$ with $3 < \gamma$	$\frac{1}{2}$
Thinner tail than a power law	$\frac{1}{2}$

Table 6.2: The dependence of the relative growth rate fluctuation scaling exponent β on the underlying replication distributions $p(k)$.

dicts its scaling exponent based on the growth distribution. If $\gamma > 3$ the growth rate distribution converges to a normal with $\beta = 1/2$ [4]. However, when $\alpha = \gamma + 1 < 2$, using standard results in extreme value theory (Zolotarev, 1986; Resnick, 2007) the standard deviation scales as a power law with size, $\sigma_G \sim N_t^{-\beta}$, where

$$\beta = (\gamma - 2)/(\gamma - 1). \tag{6.4}$$

The dependence of β on the underlying replication distributions $p(k)$ is summarized in Table 6.2.

6.4 Testing the predictions

To test the prediction that the data is Levy-distributed, in the central panel of Figure 6.1 we compare each of our three data sets to Levy distributions. The three data sets are (1) the number of birds of a given species observed along a given migration route, (2) the size of a firm as represented by its sales, and (3) the size of a U.S. mutual fund. The data shown in the middle panel of Figure 6.1 are exactly the same as in the upper panel, except that we plot the growth fluctuations G rather than their logarithmic counterpart g, we plot a cumulative distribution rather than a histogram, and we graph the data on double logarithmic scale. The fits are all good.

Because we are lucky enough that the shape parameter α and the asymmetry parameter κ are similar in all three data sets, we can collapse them onto a single curve. This is done by transforming all the data sets to the same scale in G by dividing by an empirically computed scale factor equal to the 0.75 quantile minus the 0.25 quantile (we do it this way rather

[4] For $\gamma = 3$ and $\gamma = 2$ there are logarithmic corrections to the results.

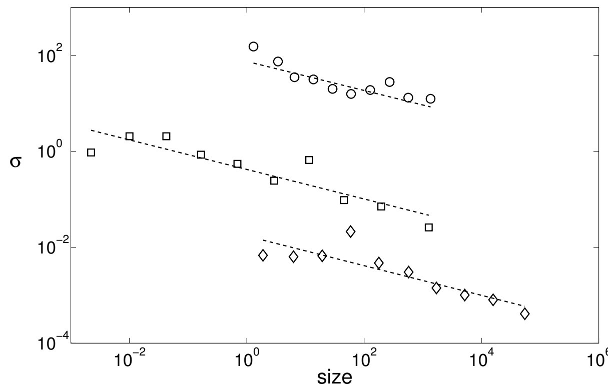


Figure 6.2: Illustration of the non-trivial scaling of the standard deviation σ as a function of size N. The straight lines on double logarithmic scale indicate power law scaling. Same symbols as in Figure 6.1. The standard deviation is computed by binning the data into bins of exponentially increasing size and computing the sample standard deviation in each bin. For clarity the breeding bird population is shifted by a factor of 10 and the mutual fund data set by a factor of 10^{-1}. The empirical data are compared to lines of slopes -0.303, -0.308 and -0.309 respectively.

than dividing by the standard deviation because the standard deviation does not exist). It is important that this normalization is done in terms of G, in contrast to the standard method which normalizes the logarithmic growth g. The standard method, illustrated in the top panel, produces a collapse for the body of the distribution, but there is no collapse for the tails – mutual funds have very heavy tails while the breeding birds closely follow the exponential even for large values of g. In contrast, the collapse using G as suggested by our theory, illustrated in the bottom panel, works for both the body and the tails.

To test the prediction of the power law scaling of the standard deviation with size we estimated γ from the data shown in Figure 6.1 and β from the data in Figure 6.2. We then make a prediction $\hat{\beta}$ for each data set using Eq. 6.4 and the estimated value of γ for each data set. The results given in Table 6.1 are in good statistical agreement in every case. (See Materials and Methods.)

year	β	$\hat{\beta}$	γ
NABB	0.30 ± 0.07	0.29 ± 0.03	2.40 ± 0.06
Mutual funds	0.29 ± 0.03	0.32 ± 0.04	2.48 ± 0.08
Firms	0.31 ± 0.07	0.35 ± 0.03	2.53 ± 0.07

Table 6.3: A demonstration that the Levy distribution makes a good prediction of the scaling of the standard deviation as a function of size. The measured value of γ based on the center panel of Figure 6.1 is used to make a prediction, $\hat{\beta}$, of the exponent of the scaling of the standard deviation. This is in good statistical agreement with β, the measured value. NABB stands for North American Breeding Birds.

6.5 Why is the replication distribution heavy-tailed?

Part of the original motivation for the interest in the non-normal properties and power law scalings of the growth fluctuations is the possibility that they illustrate an interesting collective growth phenomenon with universal applicability ranging from biology to economics. Our explanation so far seems to suggest the opposite: In our additive replication model each element acts independently of the others. As long as the replicating distribution is heavy tailed the scaling properties illustrated in Figures 6.1 and 6.2 will be observed, even without any collective interactions.

There is a subtle point here, however. Our discussion so far leaves open the question of why the replication distribution might be heavy-tailed. Based on the limited data that is currently available there are many possible explanations – it is not possible to choose one over another. One can postulate mechanisms that involve no collective behavior at all, for example, if individual birds had huge variations in the number of surviving offspring. (This might be plausible for mosquitos but does not seem plausible for birds). One can also postulate mechanisms that involve collective behavior, as we do in the next section.

6.6 The contact network explanation for heavy tails

In this section we present a plausible explanation for power law tails of $p(k)$ in terms of random influence on a scale-free contact network. This example nicely illustrates how the heavy tails of the individual replication distribution $p(k)$ can be caused by a collective phenomenon.

Assume a contact network where each node represents individuals[5]. They are connected by an edge if they influence each other. For simplicity assume that influence is bi-directional and equal, i.e. that the edges are undirected and unweighted. Let individual i be connected to d_i other individuals, where $d_i \in \{1, \ldots M\}$ is the degree of the node. The degree distribution $\mathcal{D}(d)$ is the probability that a randomly selected node has degree d.

Let each individual belong to one of Γ groups. For example, belonging to group $a \in \{1 \ldots \Gamma\}$ can represent a consumer owning a product of firm a, an investor with money in mutual fund a, or a bird of a given species taking migration route a. The groups are the same as the populations discussed earlier, i.e. N_t^a is the size of group a at time t. The dynamics are epidemiological in the sense that an individual will stay in her group unless her contacts influence her to switch. The switching is stochastic: An individual in group a with a contact in group b will switch to group b with a rate ρ_{ab}. Furthermore, the switching rate is linearly proportional to the number of contacts in that group, i.e. if an individual belonging to group a has n contacts in group b, she will switch with a rate $n\rho_{ab}$. As an example, the individual in the center of the graph in Fig. 6.6 has a degree $d = 8$ and belongs to group a. She will switch to group b with a rate $4\rho_{ab}$, to group c with a rate $2\rho_{ac}$ and to group d with a rate ρ_{ad}.

For example consider firm sales. If a given consumer likes the product of a given firm, she might influence her friends to buy more, and if she doesn't like it, she might influence them to buy less. Thus each sale in a given year influences the sales in the following year. A similar explanation applies to mutual funds, under the assumption that each investor influences her friends, or it applies to birds, under the assumption that each bird influences other birds that it comes into contact with[6].

We now show how the contact network gives rise to an additive replication model. To calculate N_{t+1}^a consider each of the N_t^a individuals in group a one at a time. Individual j in group a replicates if she remains in the group, and/or if one or more of her contacts that belong to other groups join group a. She fails to replicate if she leaves the group and also fails to influence anyone else to join. Let the resulting number of individuals that replace

[5] For a review on complex networks and dynamical processes on networks see Newman (2003) and Dorogovtsev et al. (2008).

[6] It has recently been shown that influence in flocking pigeons is hierarchical (Kurvers et al., 2009; Nagy et al., 2010).

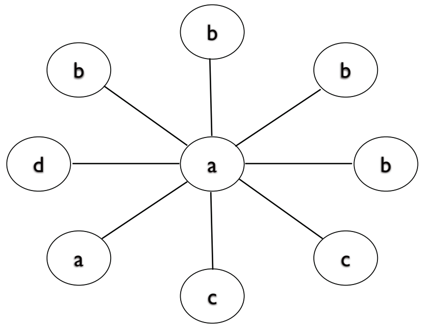


Figure 6.3: Here we show an example of a simple network. Each node represents an individual and each edge represents a contact between them. The labels represent the group the individual belongs to.

individual j be k_{jt}. This implies

$$N_{t+1}^{a} = \sum_{j \in \text{Group a}}^{N_t^a} k_{jt}, \tag{6.5}$$

which is identical to Eq. 6.1 except for the group label (which was previously implicit).

The replication factor k_{jt} is a random number with values in the range $k_i \in [0, d_j]$. Given the stochastic nature of the influence process we approximate[7] k_{jt} as a Poisson random variable with mean $E[k_{jt}] = (1-\theta_a)d_j$, where θ_a is the probability that a randomly selected contact belongs to group a. This means that the replication factor k_{jt} is proportional to the degree, i.e. $k_{jt} \sim d_j$, and that the replication distribution is proportional to the degree distribution,

$$p(k) \approx (1-\theta_a)\mathcal{D}((1-\theta_a)d). \tag{6.6}$$

Thus the influence dynamics of the contact network are an additive replication process with the individual replication distribution proportional to the degree distribution of the network. If the network is scale free, i.e. if for large k the degree distribution is a power law with $\gamma < 3$, then the growth fluctuations will be Levy distributed. It is beyond the scope of this paper to explain why the contact groups in the various settings that have been studied might be scale free, but there is at this point a large literature demonstrating that

[7]This approximation is valid for random networks, which have a local tree-like structure Dorogovtsev et al. (2008).

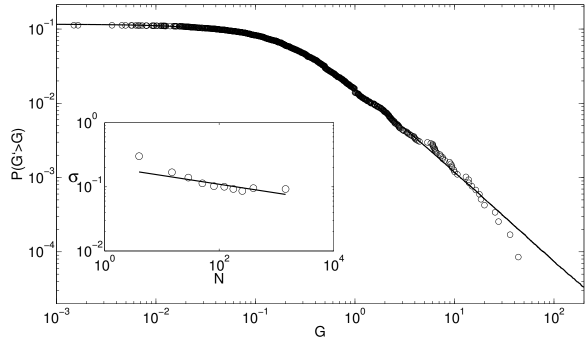


Figure 6.4: A demonstration that influence dynamics on a scale-free contact network give rise to the Levy behavior predicted by the additive replication model. The influence model was simulated for 10^3 groups on a network of 10^6 nodes, an average degree $\langle d \rangle = 10$ and a power law degree distribution $\mathcal{D}(d) \sim d^{-\gamma}$ with $\gamma = 2.2$. The cumulative growth rate distribution $P(G' > G)$ is in good agreement with the predicted Levy distribution (6.3) **Inset:** the fluctuations are compared to a line of slope $\beta = -0.1667$, illustrating the expected power law scaling.

such behavior is common (Albert and Barabasi, 2002; Newman, 2003).

A numerical simulation verifies these results[8]. We simulated a network of 10^6 nodes with a power law tailed degree distribution $\mathcal{D}(d) \sim d^{-\gamma}$ with $\gamma = 2.2$ and average degree $\langle d \rangle = 10$. The dynamics were simulated for 10^3 groups with a homogeneous switching rate $\rho_{ab} = \rho$. As expected the growth rates have a Levy distribution $P(G) \sim G^{-\gamma}$ as shown in Figure 6.6. The fitted parameter values are $\alpha = 1.2$, $\kappa = 0.25$, $c = 0.09$ and $\mu = -0.17$. The fitted value of the fluctuation scaling $\beta = 0.14 \pm 0.03$, shown in the inset of Figure 6.6, is in agreement with the predicted value of $\beta = (\gamma - 2)/(\gamma - 1) = 1/6$.

[8]The average number of individuals and the average growth rate of a group can be approximated using a mean field approach. The mean field growth rates are given by $\partial N_a/\partial t = \langle d \rangle M \theta_a (1 - \theta_a) \sum_{b=1}^{\Gamma} (\rho_{ab} - \rho_{ba})$Pastor-Satorras and Vespignani (2001). and $\theta_a = \langle d \rangle^{-1} \sum_{d'} d' f_a^{d'} \mathcal{D}(d')$, where f_a^d is the fraction of individual elements with degree d that belong to group a. We know of no analytic method to compute the growth fluctuations.

6.7 Correlations

So far we have assumed that the growth process for individual elements is uncorrelated, i.e. that the draws from $p(k)$ are I.I.D. Sufficiently strong correlations can change the results substantially. There can be correlations among the individual elements or correlations in time. For example, suppose some groups are intrinsically more or less popular than others. For example, the popularity of a city might depend on its economy and living conditions.

This can be modeled by assuming that the replication of individual j in group i is given by a random variable $\hat{k}^i_{jt}$ which is the sum of a random variable that depends on the individual and one that is common for the group, i.e.

$$\hat{k}^i_{jt} = k_{jt} + \zeta_{it}, \tag{6.7}$$

where k_{jt} is the uncorrelated individual replication factor, which we have defined and used previously, and ζ_{it} corresponds to the attractiveness of group i at time t, an I.I.D random variable and is drawn from a distribution p_ζ.

It is important to note that while we are adding correlations between elements of the same group, we are not adding correlation between elements of different groups. The replication factors $\hat{k}^i_{jt}$ of the different elements are identically distributed but are now correlated with a correlation factor given by

$$\rho = \frac{Cov(\hat{k}^i_{jt}, \hat{k}^i_{j't})}{\sigma^2_{\hat{k}}} = \frac{\sigma^2_\zeta}{\sigma^2_k + \sigma^2_\zeta}, \tag{6.8}$$

where $\sigma^2_\zeta = Var[\zeta]$ and $\sigma_k = Var[k]$. The correlation coefficient depends on the ratio of the variances and, assuming σ_ζ is non diverging, will vanish for distributions $p(k)$ with diverging second moments. It is important to note that since ζ_{it} is an I.I.D random variable drawn separately for each group, i.e. the common factor is uncorrelated among different groups, the replication factors of elements belonging to different groups are uncorrelated $Cov(\hat{k}^i_{jt}, \hat{k}^{i'}_{jt}) = 0$.

In Appendices 6.B.5 and 6.B.6 we investigate empirically US population growth and GDP growth. We show that the empirical results agree with the implications of adding group correlations to the model, which we now describe.

The resulting growth rate distribution $P_{\hat{G}}$

Given that there are N_t elements at time t we write

$$N^i_{t+1} = \sum_{j=1}^{N^i_t} \hat{k}^i_{jt} = \sum_{j=1}^{N^i_t} [k_{jt} + \zeta_{it}] \tag{6.9}$$

and the growth rate is given by

$$\hat{G}^i_t = \zeta_{it} + G^i_t, \tag{6.10}$$

where G^i_t is the growth rate without correlations as discussed in the main text. The growth rate distribution $P_{\hat{G}}$ is given by

$$P_{\hat{G}}(\hat{G}_t) = [P_G * P_\zeta](\hat{G}_t), \tag{6.11}$$

which is the convolution of the distribution P_G and P_ζ and the tails of the resulting distribution $P_{\hat{G}}$ are determined by the distribution with the heavier tail of the two. As the size of the group increases, P_G will converge to a stable levy distribution. The speed of convergence and the parameters of the distribution depend on $p(k)$. If P_ζ has heavier tail then P_G, then the tails of $P_{\hat{G}}$ are determined by P_ζ. However, the dominance of P_ζ over P_G might occur only at large group sizes depending on the convergence of P_G.

As an example for the behavior of $P_{\hat{G}}$, consider an individual replication distribution $p(k) \sim k^{-5}$ and a common replication factor distribution $P(\zeta) \sim \exp(-\zeta)$. Since the replication distribution $p(k)$ has a finite second moment, we expect P_G to converge, as the size of the group increases, to a normal distribution. However, this convergence is slow and for small group sizes, i.e. a small number of random summands comprising G, the distribution has not yet converged to a normal and will obey $P_G \sim G^{-5}$. The question is how will the distribution $P_{\hat{G}}$ behave? For small group sizes the distribution $P_G \sim G^{-5}$ dominates over P_ζ and we can expect the growth rate distribution to be $P_{\hat{G}} \sim \hat{G}^{-5}$. However, for large enough group sizes P_G will have converged to a normal and will be dominated by P_ζ such that $P_{\hat{G}} \sim \exp(-\hat{G})$. Thus, as the group size increases, we observe a change in $P_{\hat{G}}$ such that for small groups the distribution obeys $P_{\hat{G}} \sim P_G$ while for large groups we will observe $P_{\hat{G}} \sim P_\zeta$. The group size for which this transition occurs depends on the individual replication distribution $p(k)$ and the common replication factor distribution P_ζ.

The resulting fluctuation scaling of $\boldsymbol{\sigma_{\hat{G}}}$

We now show that for small sizes the individual fluctuations k_{jt} dominate, so that there is a power law scaling of σ, but for larger sizes the group fluctuations ζ_{it} dominate, and σ becomes constant (i.e. $\beta = 0$).

Since the individual replication factor k_{ijt} is independent of the common factor ζ_{it}, the relative growth rate variance $\sigma_{\hat{G}}$ is given by

$$\sigma^2_{\hat{G}} = \sigma^2_{\zeta} + \sigma^2_{G}, \tag{6.12}$$

where σ_G depends on size according to the fluctuation scaling $\sigma_G = \sigma_1 N^{-\beta}$ with β as given in Table 6.2. The fluctuation scaling as a function of the group size is given by

$$\sigma^2_{\hat{G}} = \sigma^2_{\zeta} + \sigma^2_1 N^{-2\beta}. \tag{6.13}$$

For $\sigma_\zeta < \sigma_1$, the second term in the RHS of equation (6.13) dominates for small N while for large enough N the first term will dominate. We define the transition size N^* as the group size for which the volatility of both terms on the RHS are equal such that

$$N^* = \left(\frac{\sigma_1}{\sigma_\zeta}\right)^{\frac{2}{\beta}}$$

and the volatility $\sigma_{\hat{G}}$ obeys

$$\begin{aligned} N \ll N^* \quad \sigma_{\hat{G}} &= \sigma_1 N^{-\beta} \\ N \gg N^* \quad \sigma_{\hat{G}} &= \sigma_\zeta. \end{aligned} \tag{6.14}$$

This corresponds to a fluctuation scaling exponent $\hat{\beta}$ that depends on size such that

$$\begin{aligned} N \ll N^* \quad \hat{\beta} &= \beta \\ N \gg N^* \quad \hat{\beta} &= 0. \end{aligned} \tag{6.15}$$

For $\sigma_\zeta \geq \sigma_1$ there is no transition size and the volatility can be approximated as

$$\sigma_{\hat{G}} \approx \sigma_\zeta.$$

and as a result the fluctuations are independent of size, i.e. $\hat{\beta} = 0$, in accordance with Gibrat's law. However, the size of the group for which this transition takes place can be very large such that effectively we can have $\hat{\beta} = \beta$ for the entire observation range.

6.8 Finite size effects

So far we have also assumed in our analysis that the number of elements is infinite, i.e. that there is no upper limit on the replication factors. For finite systems the growth of one group is at the expense of another. This can induce correlations which affect both the growth rate distribution and the fluctuation scaling. Nevertheless, as our simulation shows, under appropriate circumstances the theory can still describe finite systems to a very good approximation.

We examine here a setting for growth in which several groups compete over a fixed number of elements N_{total}. In this setting elements that join one group must do so at the expense of another group resulting in a decrease in the group size. At each time step the number of elements in group a changes according to the replication process such that

$$N_{t+1}^a = \sum_{j=1}^{N_t^a} k_{jt}, \tag{6.16}$$

where k_{jt} is the number of new elements replacing element j at time t. The net change in the number of elements in group a is defined as

$$\Delta_t^a = N_{t+1}^a - N_t^a$$

drawn from the distribution $P_\Delta(|\Delta_t|\,|N_t)$. This distribution is equivalent to the distribution of N_{t+1}, a distribution of a sum of N_t random variables, described in the main text. For more than a single group, under the constraint on the number of elements, a negative net size change in a group correspond to elements leaving the group to join other groups. We generalize the distribution of the net size change in group a as $\mathcal{P}_\Delta(|\Delta_t^a|\,|N_t^a)$ as

$$\begin{aligned}\mathcal{P}_\Delta(\Delta_{it}|N_t) &= P_\Delta(|\Delta_{it}|\,|N_t)\Theta(\Delta_{it})\Theta(N_{total}-N_{it}-\Delta_{it}) \\ &+ \left[\sum_{\{\Delta_t^b\}}{}' \sum_{\{N_j^b\}}{}' P_\Delta(|\Delta_t^b|\,|N_t^b)\right]\Theta(-\Delta_t^a)\Theta(N_t^a-|\Delta_t^a|),\end{aligned} \tag{6.17}$$

where Δ_t^b is the change in the number of elements in group b, N_t^b is the occupation of that group and Θ is the unit step function. The summations are over all configurations where $\Delta_t^b \geq 0$ and $N_t^b \geq 0$ such that $\sum_{b\neq a} \Delta_t^b = \Delta_t^a$ and $\sum_{b\neq a} N_t^b = N_{total} - N_{it}$. The first term in (6.17) corresponds to elements joining group a and that number is bound from above by the number of elements in the other groups $N_{total} - N_t^a$. The second term corresponds to elements leaving group a and is written in terms of elements joining the other groups.

Since the relative size change is given by $G_t = \Delta_t/N_t$ we can write the relative size change distribution as $P_G(G|N) = N\mathcal{P}_\Delta(NG|N)$ yielding

$$\begin{aligned}P_G(G_t^a|N_t^a) &= N_t^a P_\Delta(|N_t^a G_t^a|\,|N_t^a)\Theta(G_t^a)\Theta(\frac{N_{total}-N_t^a}{N_t^a}-G_t^a) \\ &+ \left[\sum_{\{G_t^b\}}{}' \sum_{\{N_t^b\}}{}' N_t^b P_\Delta(|N_t^b G_t^b|\,|N_t^b)\right]\Theta(-G_t^a)\Theta(1-|G_t^a|)\end{aligned} \tag{6.18}$$

where the G_t^b are constrained such that $G_t^b \geq -1$ and $\sum_b N_t^b G_t^b = N_t^a G_t^a$. The second term corresponding to elements leaving group a is a sum of elements joining the different groups and as such is a sum of random variables and will depend on the distribution of the number of elements in a group.

The distribution can be written explicitly for $\Gamma \gg 1$ groups with an equal number of elements $N = N_{total}/\Gamma \gg 1$ for which

$$\begin{aligned}P_G(G|N) &= \frac{1}{N^{1/\alpha-1}}L_\alpha^\kappa(\frac{G}{N^{1/\alpha-1}};c,\mu)\Theta(G)\Theta(\frac{N_{total}-N}{N}-G) \\ &+\frac{1}{\Gamma^{1/\alpha}N^{1/\alpha-1}}L_\alpha^\kappa(\frac{G}{\Gamma^{1/\alpha}N^{1/\alpha-1}};c,\mu)\Theta(-G)\Theta(1-|G|).\end{aligned} \tag{6.19}$$

In general, the growth distribution under a constraint on the number of elements (6.18) can be different than that of the non constrained case described in Section 6.2.

6.9 Discussion

The explanation that we offer here is widely applicable and very robust. The idea that a larger entity can be decomposed into a sum of smaller elements, and that the smaller elements can be modeled as if they replicate, is quite generic. As discussed in the previous section this can be broken if the growth of the elements is too correlated. Our explanation for the heavy tailed growth rate distributions and fluctuation scaling requires that the replication distribution $p(k)$ is heavy tailed. The key thing we have shown is that when this occurs, the generalized central limit theorem dictates that the growth distribution P_G will be Levy, which in turn dictates the power law size dependence of the standard deviation, $\sigma(N)$.

The previous models which are closest to ours are the model of firm size of Wyart and Bouchaud (2003) and the model of GDP due to Gabaix (2009). Both of these models assume that the size distribution $P(N_t)$ has power law tails and that firms grow via multiplicative fluctuations. They each suggested (without any testing) that additivity might lead to Levy distributions for their specific phenomena (GDP or firm size). This is in contrast to our model, which requires neither the assumption of power tails for size nor multiplicative growth. This is a critical point because the size of mutual funds does not obey a power law distribution (Schwarzkopf and Farmer, 2010b), which rules out both the the Wyart and Bouchaud and Gabaix models as general explanations. We are apparently the first to realize that these diverse phenomena all obey Levy distributions, and that this explains the power law scaling of $\sigma(N)$.

There are many possible explanations that could generate a heavy tailed replication distribution $p(k)$. Here we proposed an influence process on a scale free contact network as a possible example. This mechanism is quite general and relies on the assumption that an individual element's actions are affected by those of its contacts. Scale free networks are surprisingly ubiquitous and the existence of social, information and biological networks with power law tails with $2 < \gamma < 3$ is well documented (Albert and Barabasi, 2002; Newman, 2003), and suggests that the assumption that the degree distribution $\mathcal{D}(d)$ and hence the replication distribution $p(k)$ are heavy-tailed is plausible.

The influence model shows that the question of whether the interesting scaling properties of these systems should be regarded as "interesting collective dynamics" can be subtle. On one hand the additive replication model suppresses this – any possibility for collective action is swept into the individual replication process. On the other hand, the influence model shows that the heavy tails may nonetheless come from a collective interaction. More detailed data is needed to make this distinction.

Our model shows that, whenever its assumptions are satisfied, one should expect universal behavior as dictated by the central limit theorem: The growth fluctuations should be Levy distributed (with the normal distribution as a special case). Our model does not suggest that the tail parameter should be universal, though of course this could be possible for other reasons. Based on our model there is no reason to expect that the value of the exponent α (or equivalently γ or β) will not depend on factors that vary from example to example. Thus the growth process is universal in one sense but not in another.

6.A The logarithmic growth rate distribution

The logarithmic size change $g = \log(N_{t+1}/N_t)$, as opposed to the relative size change G, is not explicitly modeled in our theory. However, it is used by most researchers for its additivity property, which makes calculations easier and in order to compare this work with other models we describe here shortly the distribution of log changes $P_g(g)$ and its size scaling.

Given that we know the growth rate distribution P_G, the distribution of log size changes can be achieved through a change of variables $G = \exp(g) - 1$ which yields[9]

$$P_g(g|N) = \mathrm{e}^g P_G(\mathrm{e}^g - 1|N). \tag{6.20}$$

For distribution P_G that decay slow enough as G approaches-1, i.e. in the limit $g \to -\infty$, the lower tail of the distribution can be approximated as

$$P_g(g \ll 0) \sim \mathrm{e}^{-|g|}. \tag{6.21}$$

That is, the lower tail is an exponential independent of the form of the replication distri-

[9]Under a change of variables the distribution functions transform such that $P_g(g)\mathrm{d}g = P_G(G)\mathrm{d}G$ holds. Given that we know $P_G(G)$ and a change of variables $G = f(g)$ the distribution $P_g(g)$ is then given by $P_g(g) = f'(g)P_G(f(g))$.

bution $p(k)$.

The upper tail, as opposed to the lower tail, depends on the replication distribution $p(k)$ through P_G and can be approximated as

$$P_g(g \gg 0) \sim \mathrm{e}^g P_G(\mathrm{e}^g).$$

As was discussed previously, for a power law replication process $p(k)$ with $\gamma < 3$ the resulting distribution P_G converges to a Levy stable distribution. The Levy stable distributions can be shown Nolan (2009) to have an asymptotic power law decay

$$L^{\beta}_{\gamma-1}(x; c, \mu) \sim x^{-\gamma} \tag{6.22}$$

in the $x \to \infty$ limit. This suggests that the resulting logarithmic growth rate distribution P_g will have an exponentially decaying upper tail

$$P_g(g \gg 0) \sim e^{-(\gamma-1)g}. \tag{6.23}$$

For distributions with a well defined mean, i.e. with $\gamma > 2$, the upper tail (6.23) always decays faster then the lower tail (6.21), which means that the logarithmic growth rate distribution P_g is always asymmetric with more weight in the lower tail. This asymmetry in the logarithmic growth distribution was observed in many systems.

6.B Additional empirical evidence

In this section we describe in more detail the empirical evidence for five data sets. Three of these data sets have, as it turns out, growth rate distributions with diverging second moments for which the group correlations are negligible; the growth rate of breeding birds of North America in Section 6.B.2, the growth, due to investors, of mutual funds for the years 1997 to 2007 in Section 6.B.3 and the growth of firms with respect to sales in Section 6.B.4. We offer here additional results to those discussed in the main text.

On the other hand, US population growth and the GDP of countries, have, as it turns out, a growth rate distribution with a finite second moment. We argue here that the empirical observations for these systems are consistent with the model with group correlations.

For such systems the model predicts two things: first, the growth rate distribution for small N is different than the distribution for large N since P_G converges to a normal. Moreover, if P_ζ is heavier tailed than a normal, then as N increases the distribution will converge to P_ζ. Second, we predict a transition of the scaling exponent β from 1/2 to 0 as the size increases. This is indeed what we observe for GDP and cities.

6.B.1 Empirical fitting procedures

The empirical investigation of the three data sets was conducted as follows: first, the fluctuation scaling exponent β is estimated from the data. Then, we normalize the growth rate distribution followed by an estimation of the tail exponent γ. Lastly we will compare the measured β and γ to the expected relationship from our model. To estimate the fluctuation scaling exponent β the relative growth rate distribution $G = N_{t+1}/N_t - 1$ was binned into 20 exponentially spaced bins according to size N_t. For each bin i, the variance of the growth rates σ_i^2 was empirically estimated. Then the logarithm of the measured variances were regressed on the logarithm of the average size $\bar{N}_i$

$$\log(\sigma) = \beta \log(N) + \sigma_1 \tag{6.24}$$

such that the slope is the ordinary least squares (OLS) estimator of β. Using the fluctuation scaling exponent β we predict the tail exponent

$$\hat{\gamma} = \frac{2-\beta}{1-\beta},$$

which we will compare to the measured maximum likelihood estimator (MLE) of the tail exponent. To estimate the tail exponent, we normalize the growth rate G such that it has zero mean and unit variance. For distributions with a power law tail, the MLE power law exponent was estimated using the technique described in Clauset et al. (2009) for fitting power law tails. The method used uses the following modified Kolmogorov-Smirnoff statistic

$$KS = \max_{x > x_{min}} \frac{|s(x) - p(x)|}{\sqrt{p(x)[1-p(x)]}},$$

where s is the empirical cumulative distribution and p is the hypothesized cumulative distribution.

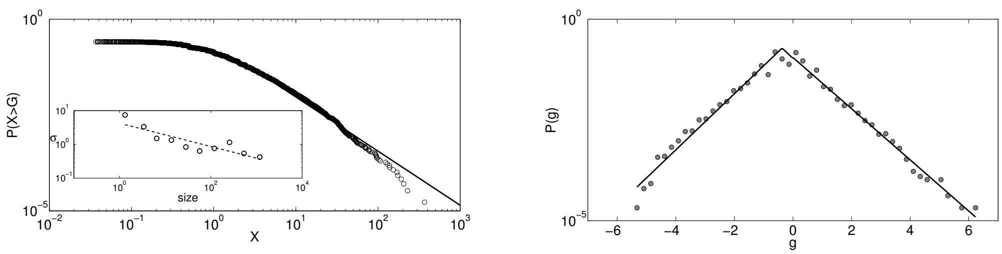


Figure 6.5: Left panel: The cumulative distribution for the relative growth rates $P(G > X)$ for the year 2007 is plotted on a log-log scale. The distributions is compared to a linear line with a lope corresponding to the upper tail exponent from the MLE fit for γ given in Table 6.B.2. Inset: The relative growth rate fluctuations σ for the year 2007 as a function of the number of birds, measured as yearly sales, is compared to a line with a slope corresponding to β given in Table 6.B.2.

Right panel: The relative growth rates density $P_g(g)$ for the year 2007, as resulting form binning the data, is plotted on a semi-log scale. The distributions is compared to a laplacian distribution.

6.B.2 Breeding birds of America

North american breeding birds dataset

We use the the North American breeding bird survey, which contains 42 yearly observations for over 600 species along more than 3,000 observation routes. For each route the number of birds from each species is quoted for each year in the period 1966-2007. For each year in the data set, from 1966 to 2007, we computed the yearly growth with respect to each species in each route. The data set can be found online at ftp://ftpext.usgs.gov/pub/er/md/laurel/BBS/DataFiles/.

Empirical results

For each year in the data set, from 1966 to 2007, we computed the yearly growth with respect to each specie in each route and compared the data to the model. For each year we computed the Maximum Likelihood Estimators (MLE) for the power law exponents of the upper tail. The results are summarized in Table 6.B.2 and the growth rate distribution for the year 2007 is plotted in Figure 6.5. Remarkably, for most years, the MLE of the upper law tail exponent γ is in agreement with the estimations from the growth fluctuations $\hat{\gamma}$

year	β	$\hat{\gamma}$	γ
1966	0.39 ± 0.10	2.64 ± 0.27	2.37 ± 0.11
1967	0.36 ± 0.10	2.56 ± 0.25	2.53 ± 0.13
1968	0.44 ± 0.15	2.77 ± 0.46	2.39 ± 0.09
1969	0.30 ± 0.15	2.43 ± 0.30	2.45 ± 0.12
1970	0.42 ± 0.11	2.72 ± 0.31	2.29 ± 0.08
1971	0.31 ± 0.15	2.44 ± 0.32	2.38 ± 0.11
1972	0.30 ± 0.11	2.44 ± 0.23	2.61 ± 0.21
1973	0.32 ± 0.13	2.47 ± 0.27	2.53 ± 0.13
1974	0.26 ± 0.15	2.35 ± 0.28	2.61 ± 0.17
1975	0.33 ± 0.05	2.48 ± 0.12	2.64 ± 0.10
1976	0.33 ± 0.10	2.49 ± 0.21	2.46 ± 0.10
1977	0.29 ± 0.11	2.41 ± 0.22	2.65 ± 0.14
1978	0.44 ± 0.12	2.78 ± 0.37	2.42 ± 0.12
1979	0.50 ± 0.62	3.00 ± 2.48	2.63 ± 0.14
1980	0.41 ± 0.14	2.70 ± 0.41	2.48 ± 0.10
1981	0.30 ± 0.09	2.42 ± 0.19	2.43 ± 0.08
1982	0.40 ± 0.12	2.66 ± 0.32	2.85 ± 0.27
1983	0.43 ± 0.13	2.74 ± 0.40	2.58 ± 0.15
1984	0.34 ± 0.14	2.52 ± 0.33	2.31 ± 0.07
1985	0.33 ± 0.18	2.49 ± 0.40	2.36 ± 0.10
1986	0.34 ± 0.10	2.51 ± 0.24	2.46 ± 0.12
1987	0.35 ± 0.19	2.55 ± 0.45	2.26 ± 0.05
1988	0.32 ± 0.13	2.47 ± 0.28	2.60 ± 0.13
1989	0.33 ± 0.06	2.49 ± 0.12	2.33 ± 0.06
1990	0.32 ± 0.12	2.47 ± 0.26	2.76 ± 0.18
1991	0.30 ± 0.15	2.43 ± 0.30	2.73 ± 0.17
1992	0.35 ± 0.08	2.54 ± 0.19	2.52 ± 0.08
1993	0.35 ± 0.14	2.55 ± 0.33	2.87 ± 0.16
1994	0.34 ± 0.11	2.53 ± 0.26	2.53 ± 0.11
1995	0.31 ± 0.11	2.46 ± 0.24	2.42 ± 0.06
1996	0.48 ± 0.26	2.91 ± 0.96	2.71 ± 0.12
1997	0.40 ± 0.26	2.66 ± 0.71	2.35 ± 0.05
1998	0.30 ± 0.10	2.43 ± 0.20	2.61 ± 0.10
1999	0.34 ± 0.10	2.51 ± 0.24	3.06 ± 0.25
2000	0.22 ± 0.18	2.29 ± 0.30	2.51 ± 0.09
2001	0.32 ± 0.16	2.48 ± 0.36	2.69 ± 0.10
2002	0.39 ± 0.11	2.65 ± 0.29	2.70 ± 0.11
2003	0.42 ± 0.22	2.71 ± 0.65	2.71 ± 0.14
2004	0.30 ± 0.14	2.42 ± 0.28	2.55 ± 0.08
2005	0.27 ± 0.17	2.37 ± 0.33	2.95 ± 0.17
2006	0.33 ± 0.08	2.49 ± 0.18	2.54 ± 0.09
2007	0.30 ± 0.14	2.43 ± 0.29	2.41 ± 0.06
mean	0.35 ± 0.02	2.54 ± 0.04	2.55 ± 0.01

Table 6.4: The parameter values for the proposed growth model as measured for from the North American breeding bird data. The parameters of the model were measured for the aggregated monthly data during each year and the errors are the 95% confidence interval. The last row corresponds to the average value over the different years. The parameters are as follows:

β - The OLS estimator for the exponent of the size dependence of the fluctuations in the relative growth rate.

$\hat{\gamma}$ - The predicted power law exponent as can be inferred from β. To be compared with γ.

γ - MLE for the power law exponent of the upper tail of $P(G)$.

. Moreover, for each of these values the mean over the different years was calculated and all three are in agreement. Our model seems to describe well the relationship between the growth rate distribution and the growth rate fluctuations observed observed the North American breeding birds.

6.B.3 Mutual Funds

US equity mutual fund dataset

We use the Center for Research in Security Prices (CRSP) mutual fund database, restricted to equity mutual funds existing in the years 1997 to 2007. An equity fund is one with at least 80% of its portfolio in stocks. As the size of the Mutual fund we use the total net assets value (TNA) in real US dollars as reported monthly. Growth in the mutual fund industry, measured by change in TNA, is comprised of two sources: growth due to the funds performance and growth due to flux of money from investors, i.e. mutual funds can grow in size if their assets increase in value or due to new money coming in from investors. We define the relative growth in the size of a fund at time t as

$$G_{TNA}(t) = \frac{TNA_{t+1}}{TNA_t} - 1$$

and decompose it as follows;

$$G_{TNA}(t) = r_t + G_t, \tag{6.25}$$

where r_t is the fund's return, quoted monthly in the database, and G_t is the growth due to investors. For our purposes here we only consider G_t, the growth due to investors.

Empirical results

For each year we computed the Maximum Likelihood Estimators (MLE) for the power law exponents of the upper tail. The results are summarized in Table 6.5 and the fits for some of the years are given in Figure 6.6. For all years, the upper law tail exponent is in the range $\gamma \in (2, 3)$ and except for the years 1999, 2002 and 2005 the measured value γ is in agreement with the estimations from the growth fluctuations $\hat{\gamma}$. Moreover, for each of these values the mean over the different years was calculated and all three are in agreement. Our model seems to describe well the relationship between the growth rate distribution and the

year	β	$\hat{\gamma}$	γ
1997	0.34 ± 0.11	2.52 ± 0.25	2.36 ± 0.12
1998	0.38 ± 0.09	2.62 ± 0.24	2.67 ± 0.17
1999	0.49 ± 0.08	2.98 ± 0.30	2.53 ± 0.09
2000	0.35 ± 0.04	2.55 ± 0.09	2.58 ± 0.08
2001	0.38 ± 0.06	2.62 ± 0.17	2.85 ± 0.14
2002	0.44 ± 0.08	2.80 ± 0.26	2.50 ± 0.08
2003	0.32 ± 0.05	2.47 ± 0.10	2.68 ± 0.11
2004	0.37 ± 0.07	2.60 ± 0.17	2.63 ± 0.09
2005	0.42 ± 0.10	2.71 ± 0.29	2.35 ± 0.07
2006	0.34 ± 0.06	2.52 ± 0.15	2.56 ± 0.08
2007	0.29 ± 0.08	2.42 ± 0.16	2.53 ± 0.08
mean	0.38 ± 0.02	2.62 ± 0.05	2.57 ± 0.03

Table 6.5: The parameter values for the proposed growth model as measured for monthly size growth (in US dollars) of equity mutual funds using the CRSP database. The parameters of the model were measured for the aggregated monthly data during each year and the errors are the 95% confidence interval. The last row corresponds to the average value over the different years. The parameters are as follows:

β - The OLS estimator for the exponent of the size dependence of the fluctuations in the relative growth rate.

$\hat{\gamma}$ - The predicted power law exponent as can be inferred from β. To be compared with γ.

γ - MLE for the power law exponent of the upper tail of $P(G)$.

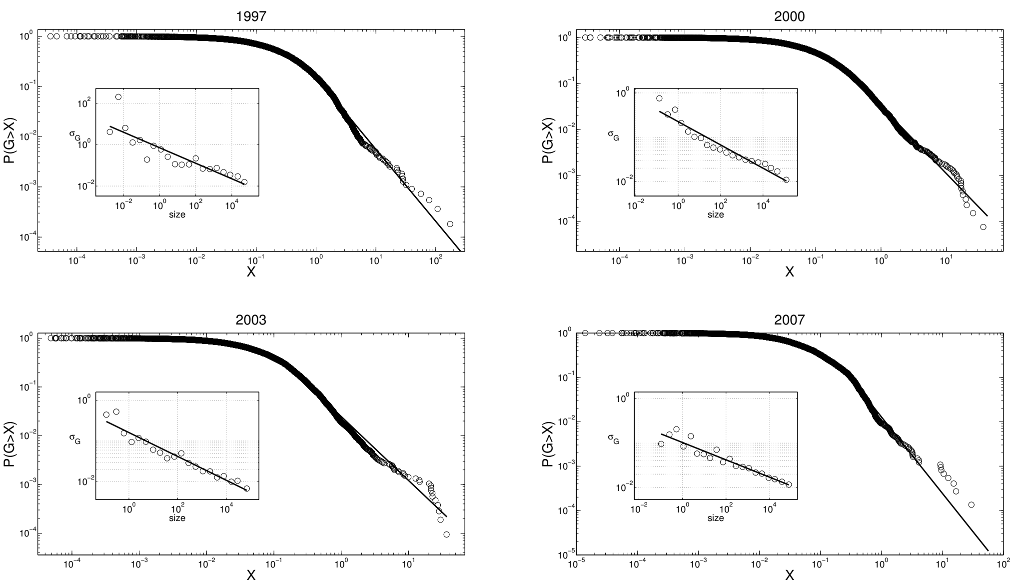


Figure 6.6: The cumulative distribution for the relative growth rates $P(G > X)$ are plotted on a log-log scale for the years 1997, 2000, 2003 and 2007. The upper tail is compared to a line on a log-log plot with a slope corresponding to the MLE fits for γ given in Table 6.5. Insets: The relative growth rate fluctuations σ as a function of the size of the mutual fund are compared to lines with slopes corresponding to β_G given in Table 6.5.

β	$\hat{\gamma}$	γ
0.31 ± 0.07	2.45 ± 0.15	2.53 ± 0.07

Table 6.6: The parameter values for the proposed growth model as measured for the yearly growth (in US dollars) of firm sales using the COMPUSTAT database. The errors are the 95% confidence interval. The parameters are as follows:
β - The OLS estimator for the exponent of the size dependence of the fluctuations in the relative growth rate.
$\hat{\gamma}$ - The predicted power law exponent as can be inferred from β. To be compared with γ.
γ - MLE for the power law exponent of the upper tail of $P(G)$.

growth rate fluctuations observed in the mutual fund industry.

Another interesting and non trivial observation is that the tail is heavy in the sense that the second moment does not exist especially when one considers the fact that this growth is solely due to investors. It seems that in mutual funds the individual replication dominates over the common replication and growth is mostly affected by the ability of each new client to generate more clients than the mere attractiveness of the fund.

6.B.4 Firm Growth

US public firms dataset

We use the 2008 COMPUSTAT dataset containing information on all US public firms. As the size of a firm we use the dollar amount of sales. Growth is given by the 3 year growth in sales.

Empirical results

The OLS estimator for β, the resulting tail exponent prediction $\hat{\gamma}$ and the MLE tail exponent γ are summarized in Table 6.6. Our model seems to describe well the relationship between the growth rate distribution and the growth rate fluctuations observed for the growth of public firms.

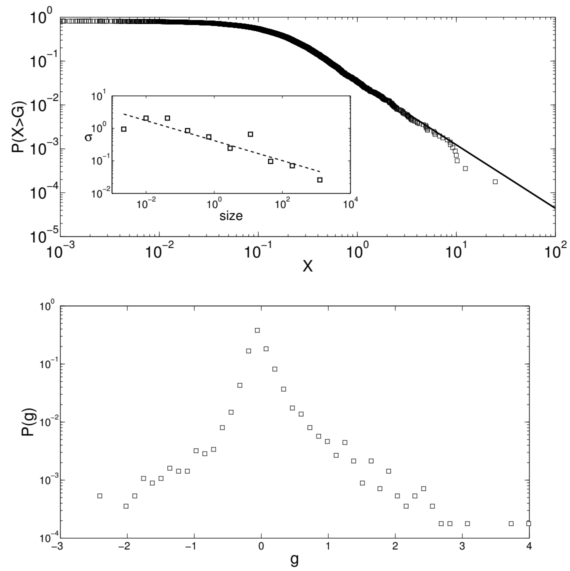
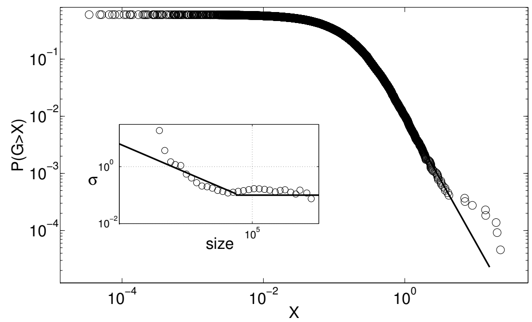


Figure 6.7: Left panel: The cumulative distribution for the relative growth rates $P(G > X)$ is plotted on a log-log scale. The upper tail is compared to a line on a log-log plot with a slope corresponding to the MLE fits for γ given in Table 6.6. Inset: The relative growth rate fluctuations σ as a function of the size of the company, measured as yearly sales, is compared to a line with a slope corresponding to β given in Table 6.6.

Right panel: The logarithmic growth rate distribution for the relative growth rates $P_g(g)$, as resulting from binning the data, is plotted on a semi-log scale.



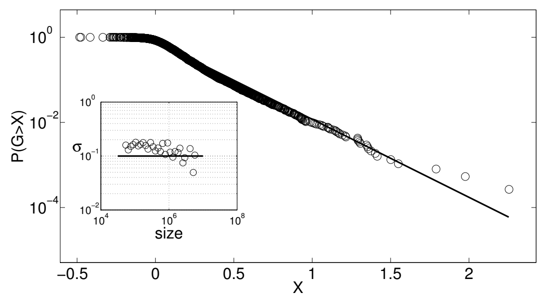


Figure 6.8: Here we show that the observed growth rate for US counties agrees with our model with the addition of group correlations. **Top:** the cumulative growth rate distribution $P(G > X)$ calculated for the US Census decadal data on all counties (in the US) populations in the years 1920 to 1990. The upper tail of the distribution is compared to linear slope of -2.2 which corresponds to a power law upper tail with an MLE exponent of $\gamma = 3.2$. **Top inset:** the fluctuations in the relative growth rate σ as a function of population. **Bottom:** as in the top panel but for counties with a reported population of at least 50000 residents.

6.B.5 US population growth

US population dataset

We test our model on population growth using two data sets: US county population census counts [10] for the years 1920-90 and US core based statistical areas (CBSA) population for the years 2000 and 2007 [11].

[10]The US county population counts contains data for every decade in the period 1900 to 1990 and can be found online on the US Census Bureau website.

[11]The CBSA population for January 1st 2007 is an estimation by the United States Census Bureau. The CBSA population for April 1st 2000 is based on the United States Census 2000

Empirical results

We begin by investigating the empirical relative growth rate distribution P_G for counties using the US census county data set, which contains data on counties of all sizes including less populated ones. In the top panel of Figure 6.8 the cumulative distribution $P(G > X)$ for all counties is plotted on a double logarithmic scale. The distribution is estimated as having a power law tail with an MLE of $\gamma = 3.2$, which corresponds to a finite second moment. Our theory, with group correlations, predicts that the scaling exponent β will transition from 1/2 at small N to 0 at large N. This is indeed what we see in the inset of the top panel of Figure 6.8. Moreover, it seems that the transition size is approximately $N^* \approx 50000$.

Another prediction of our model, given that the P_G has a finite second moment, is that for $N > N^*$ the distribution should either converge to a normal or converge to P_ζ (only if P_ζ is heavier tailed than a Gaussian). To check this we plot the growth rate distribution for all counties with a population larger than N^*. The results are plotted in the bottom panel of Figure 6.8. What we see as that the growth rate distribution is exponential, which corresponds to a straight line on a semi-logarithmic scale. Since the exponential is heavier tailed than a normal, we conclude that what we are observing is P_ζ.

To verify the observations for counties, we now examine another data set containing data on the population of core based statistical areas (CBSA) which have a population of at least 10,000. These are large populations with most falling at $N > N^*$ so we expect to observe the same phenomena as we did for large counties; exponential growth rate distribution and size independent fluctuations $\beta = 0$. This is indeed the case as can be seen in Figure 6.9.

To conclude, it seems that for urban growth we can say the following: the individual replication factor is heavy tailed yet has a well defined second moment $p(k) \sim k^{-3.2}$. The common replication factor is drawn from a exponential distribution $P_\zeta(\zeta_t) \sim \exp(-c\zeta_t)$ where c is some constant. This means that for small urban areas the growth is determined mostly buy the actions of the individuals and the incentives they manage to create. For larger urban areas on the other hand, the growth is dominated by the common factor that is the attractiveness of the urban area.

It is important to note that the nature of the scalings for cities is controversial and strongly depends on how a city is defined – our results are in agreement with those who claim

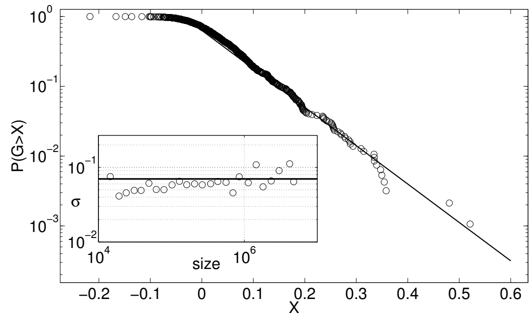


Figure 6.9: Here we show that the observed growth rate for US core based statistical areas (CBSA) agrees with our model with the addition of group correlations. The relative growth rate cumulative distribution $P(G > X)$ calculated from the US CBSA growth between the years 2000 to 2007. The CDF is plotted on a log-linear scale and is compared to a straight line which corresponds to an exponential decaying tail for the CDF. **Inset:** the fluctuations in the relative growth rate σ as a function of population.

the scaling is not very good (Eeckhout, 2004). Rather than using the census definitions, Rozenfeld et al. (2008) use a clustering algorithm for defining cities and then the fluctuation scaling (without the group correlations) seems to hold.

6.B.6 GDP Growth

GDP dataset

We investigate the growth of the gross domestic product (GDP) of countries using the World Bank estimates[12]. The data set contains GDP data for 209 countries and territories for the years 1960 to 2007 quoted in current USD. To avoid biases and to maximize the number of observations, we chose to work only with countries that had data available for the entire time range. This resulted in 96 countries with 48 observations each for a total of 4608 observations.

Empirical results

We investigate the annual GDP growth rate distribution by aggregating the data for the various countries over all the years in the data set. By doing so we are using the approxima-

[12] We used publicly available data that can be found on the world bank website *http://www.worldbank.org*.

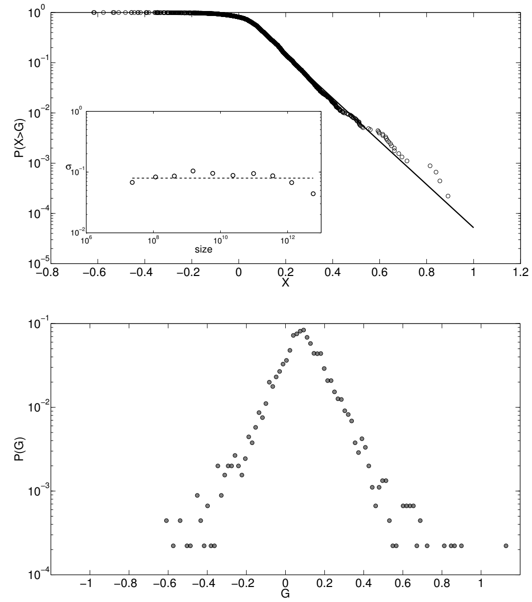


Figure 6.10: Here we show that the observed growth rate for GDP agrees with our model with the addition of group correlations. **Top:** the annual GDP growth rate CDF (plotted on a semi logarithmic scale) is compared to a straight line which corresponds to an exponential distribution. **Top inset:** the annual GDP growth rate fluctuation σ is plotted as a function of GDP on a double logarithmic scale. The data was binned into 10 exponentially spaced bins. The observations are compared to a straight line of slope 0, which corresponds to fluctuations independent of size. **Bottom:** the annual GDP growth rate distribution $P_G(G)$ as measured by binning the data into 100 equally spaced bins.

tion that the observation for different years are independent. The aggregated growth rate distribution $P_G(G)$ plotted in Figure 6.10 seems to be a Laplacian distribution which has a finite second moment. Moreover, the growth rate fluctuations have been found to be size independent with $\beta \approx 0$ [13]. Under our model, with the group correlations, this reflects that the common replication factor distribution P_ζ dominates the sum of individual replication factors. This in agreement with a case were the individual replication factor k is drawn from some distribution $p(k)$ with a finite second moment and the common replication factor ζ is drawn from a exponential distribution $P_\zeta(\zeta_t) \sim \exp(-c\zeta_t)$ where c is some constant. This result is similar to what we have observed for city population growth.

[13] The regression parameters resulted in parameter estimations of $\beta = 0.02 \pm 0.04$. Thus, a size independent fluctuation null hypothesis can not be rejected.

Bibliography

Reka Albert and Albert-Laszlo Barabasi. Statistical mechanics of complex networks. *Reviews of Modern Physics*, 74:47, 2002.

Robert Almgren, Chee Thum, Emmanuel Hauptmann, and Hong Li. Equity market impact. *Risk magazine*, July 2005.

L. A. N. Amaral, S. V. Buldyrev, S. Havlin, P. Maass, M. A. Salinger, H. E. Stanley, and M. H. R. Stanley. Scaling behavior in economics: The problem of quantifying company growth. *Physica A*, 244:1–24, Oct 1997a.

L. A. N. Amaral, S. V. Buldyrev, S. Havlin, P. Maass, M. A. Salinger, H. E. Stanley, and M. H. R. Stanley. Scaling behavior in economics: The problem of quantifying company growth. *Physica A*, 244:1–24, Oct 1997b.

Luís A. Nunes Amaral, Sergey V. Buldyrev, Shlomo Havlin, Michael A. Salinger, and H. Eugene Stanley. Power law scaling for a system of interacting units with complex internal structure. *Phys. Rev. Lett.*, 80(7):1385–1388, Feb 1998. doi: 10.1103/PhysRevLett.80.1385.

Lus A. Nunes Amaral, Sergey V. Buldyrev, Shlomo Havlin, Heiko Leschhorn, Michael A. Salinger, H. Eugene Stanley, and Michael H. R. Stanley. Scaling behavior in economics: I. empirical results for company growth. *J. Phys. I France*, 7:621–633, 1997.

Robert Axtell, Rich Perline, and Daniel Teitelbaum. Volatility and asymmetry of small firm growth rates over increasing time frames. The Office of Advocacy Small Business Working Papers 06rarpdt, U.S. Small Business Administration, Office of Advocacy, 2006.

Robert L. Axtell. Zipf distribution of U.S. firm sizes. *Science*, 293(5536):1818–1820, Sep 2001. doi: 10.1126/science.1062081.

Louis Bachelier. Théorie de la spéculation. *Annales Scientifiques de l'Ecole Normale Supérieure*, 17(3):21–86, 1900.

Brad M. Barber, Terrance Odean, and Lu Zheng. Out of sight, out of mind: The effects of expenses on mutual fund flows. *The Journal of Business*, 78(6):2095–2119, 2005.

Jonathan B. Berk and Richard C. Green. Mutual fund flows and performance in rational markets. *Journal of Political Economy*, 112(6):1269–1295, 2004. doi: 10.1086/424739.

Nicolas P. B. Bollen and Jeffrey A. Busse. Short-Term Persistence in Mutual Fund Performance. *Rev. Financ. Stud.*, 18(2):569–597, 2005. doi: 10.1093/rfs/hhi007.

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

Giulio Bottazzi. Firm diversification and the law of proportionate effect. LEM Papers Series 2001/01, Laboratory of Economics and Management (LEM), Sant'Anna School of Advanced Studies, Pisa, Italy, December 2001.

Giulio Bottazzi and Angelo Secchi. Common properties and sectoral specificities in the dynamics of U.S. manufacturing companies. *Review of Industrial Organization*, 23(3-4), Dec 2003a.

Giulio Bottazzi and Angelo Secchi. A stochastic model of firm growth. *Physica A: Statistical Mechanics and its Applications*, 324(1-2):213–219, June 2003b.

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

Giulio Bottazzi, Elena Cefis, Giovanni Dosi, and Angelo Secchi. Invariances and diversities in the evolution of manufacturing industries. *Small Business Economics*, 29(1):137–159, 2007.

Jean-Philippe Bouchaud. Economics needs a scientific revolution. *Nature*, 455(1181):7217, October 2008.

Jean-Philippe Bouchaud, Yuval Gefen, Marc Potters, and Matthieu Wyart. Fluctuations and Response in Financial Markets: The Subtle Nature of 'Random' Price Changes. *Quantitative Finance*, 4(176), 2004.

Jean-Philippe Bouchaud, J. Doyne Farmer, and Fabrizio Lillo. *"How markets slowly digest changes in supply and demand" In Handbook of Financial Markets: Dynamics and Evolution, eds. Thorston Hens and Klaus Schenk-Hoppe*. Elsevier: academic press, 2008.

Nerissa C. Brown, Kelsey D. Wei, and Russell R. Wermers. Analyst Recommendations, Mutual Fund Herding, and Overreaction in Stock Prices. *SSRN eLibrary*, 2009.

Mark Buchanan. Waiting for the maths. *Nature Physics*, 5(11):776, November 2009.

Sergey V. Buldyrev, Lus A. Nunes Amaral, Shlomo Havlin, Heiko Leschhorn, Michael A. Salinger, H. Eugene Stanley, and Michael H. R. Stanley. Scaling behavior in economics: Modeling of company growth. *Journal de Physique I France*, 7:635–650, 1997.

Jeffrey A. Busse. Another look at mutual fund tournaments. *The Journal of Financial and Quantitative Analysis*, 36(1):53–73, Mar 2001. ISSN 0022-1090.

David Canning, Luís A. Nunes Amaral, Youngki Lee, Martin Meyer, and H. Eugene Stanley. Scaling the volatility of gdp growth rates. *Economics Letters*, 60(3):335–341, 1998.

Mark M. Carhart. On persistence in mutual fund performance. *The Journal of Finance*, 52(1):57–82, 1997. ISSN 00221082.

Mark M. Carhart. On persistence in mutual fund performance. *The Journal of Finance*, 52(1):57–82, March 2009.

Carolina Castaldi and Giovanni Dosi. The patterns of output growth of firms and countries: Scale invariances and scale specificities. *Empirical Economics*, 37(3):475–495, 2009.

Louis K. C. Chan and Josef Lakonishok. Institutional equity trading costs: Nyse versus nasdaq. *The Journal of Finance*, 52(2):713–735, 1997.

Joseph Chen, Harrison Hong, Ming Huang, and Jeffrey D. Kubik. Does fund size erode mutual fund performance? the role of liquidity and organization. *American Economic Review*, 94(5):1276–1302, 2004.

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

Chevalier, Judith and Ellison, Glenn. Risk taking by mutual funds as a response to incentives. *The Journal of Political Economy*, 105(6):1167–1200, Dec 1997. ISSN 0022-3808.

Tarun Chordia, Asani Sarkar, and Avanidhar Subrahmanyam. An empirical analysis of stock and bond market liquidity. *The Review of Financial Studies*, 18(1):85–129, 2005.

Susan E. K. Christoffersen. Why do money fund managers voluntarily waive their fees? *The Journal of Finance*, 56(3):1117–1140, 2001.

Aaron Clauset, Cosma Rohilla Shalizi, and M. E. J. Newman. Power-law distributions in empirical data. *SIAM Reviews*, 2007.

Aaron Clauset, Cosma R. Shalizi, and M. E. J. Newman. Power-law distributions in empirical data. *SIAM Review*, 51(4):661, Feb 2009.

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

G. De Fabritiis, F. Pammolli, and M. Riccaboni. On size and growth of business firms. *PHYSICA A*, 324:38, 2003.

Robert Dorfman. A formula for the gini coefficient. *The Review of Economics and Statistics*, 61(1):146–149, 1979.

S. N. Dorogovtsev, A. V. Goltsev, and J. F. F. Mendes. Critical phenomena in complex networks. *Reviews of Modern Physics*, 80(4):1275, 2008.

Giovanni Dosi. Statistical regularities in the evolution of industries. a guide through some evidence and challenges for the theory. LEM Papers Series 2005/17, Laboratory of Economics and Management (LEM), Sant'Anna School of Advanced Studies, Pisa, Italy, July 2005.

Alfonso Dufour and Robert F. Engle. Time and the price impact of a trade. *Journal of Finance*, 55(6):2467–2498, December 2000.

Roger M. Edelen, Richard B. Evans, and Gregory B. Kadlec. Scale Effects in Mutual Fund Performance: The Role of Trading Costs. *SSRN eLibrary*, 2007.

Jan Eeckhout. Gibrat's law for (all) cities. *American Economic Review*, 94(5):1429–1451, December 2004.

J. P. Embrechts, C. Kluppelberg, and T. Mikosch. *Modeling Extremal Events for Insurance and Finance*. Springer-Verlag, Berlin, 1997.

Eugene F. Fama and Kenneth R. French. Common risk factors in the returns on stocks and bonds. *Journal of Financial Economics*, 33(1):3–56, February 1993.

Eugene F. Fama and Kenneth R. French. Luck versus Skill in the Cross Section of Mutual Fund Alpha Estimates. *The Journal of Finance (forthcoming)*, 2010.

J. Doyne Farmer. Market force, ecology and evolution. *Industrial and Corporate Change*, 11(5):895–953, November 2002.

J. Doyne Farmer and Duncan Foley. The economy needs agent based modeling. *Nature*, 460(7256):685, August 2009.

J. Doyne Farmer and Fabrizio Lillo. On the origin of power law tails in price fluctuations. *Quantitative Finance*, 4, 2004.

D. Fu, F. Pammolli, S. V. Buldyrev, M. Riccaboni, K. Matia, K. Yamasaki, and H. E. Stanley. The growth of business firms: Theoretical framework and empirical evidence. *Proc. Natl. Acad. Sci.*, 102:18801–18806, 2005.

Xavier Gabaix. The granular origins of aggregate fluctuations. Working Paper 15286, National Bureau of Economic Research, August 2009. URL `http://www.nber.org/papers/w15286`.

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

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

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

Crispin W. Gardiner. *Handbook of Stochastic Methods.* Springer, 3rd ed. edition, April 2004. ISBN 3540208828.

R. Gibrat. Les inegalites economiques, 1931.

Lszl Gillemot, J. Doyne Farmer, and Fabrizio Lillo. There's more to volatility than volume. *Quantitative Finance*, 6(5):371–384, 2006.

Corrado Gini. Variabilitá e mutabilitá. Technical report, Bologna: Tipografia di Paolo Cuppini, 1912.

Parameswaran Gopikrishnan, Vasiliki Plerou, Xavier Gabaix, and Stanley. H. Eugene. Statistical properties of share volume traded in financial markets. *Phys. Rev. E*, 62(4): R4493–R4496, Oct 2000. doi: 10.1103/PhysRevE.62.R4493.

Guercio, Diane Del and Tkac, Paula A. The determinants of the flow of funds of managed portfolios: Mutual funds vs. pension funds. *The Journal of Financial and Quantitative Analysis*, 37(4):523–557, Dec 2002. ISSN 0022-1090.

Joel Hasbrouck. Trading costs and returns for u.s. equities: Estimating effective costs from daily data. *The Journal of Finance*, 64(3), June 2009.

Jerry A. Hausman, Andrew W. Lo, and A. Craig MacKinlay. An ordered probit analysis of transaction stock prices. *Journal of Financial Economics*, 31(3):319–379, June 1992.

Robert W. Holthausen, Richard W. Leftwich, and David Mayers. The effect of large block transactions on security prices: A cross-sectional analysis. *Journal of Financial Economics*, 19(2):237–267, December 1987.

Robert W. Holthausen, Richard W. Leftwich, and David Mayers. Large-block transactions, the speed of response, and temporary and permanent stock-price effects. *Journal of Financial Economics*, 26:71–95, 1990.

Y. Ijiri and H. Simon. Some distributions associated with bose-einstein statistics. *Proc. Nat. Acad. Sci.*, page 1654, 1975.

Y. Ijiri and H. A. Simon. *Skew Distribution and the Size of Business Firms.* North Holland, 1977.

Donald B Keim and Ananth Madhaven. The upstairs market for large-block transactions: Analysis and measurement of price effects. *Review of Financial Studies*, 9(1):1–36, 1996.

Timothy H. Keitt and Eugene H. Stanley. Dynamics of north american breeding bird populations. *Nature*, 393:257–260, 1998.

Timothy H. Keitt, Luis A. N. Amaral, Sergey V. Buldyrev, and Eugene H. Stanley. Scaling in the growth of geographically subdivided populations: invariant patterns from a continent-wide biological survey. *The Royal Society B*, 357(1421):627–633, May 2002.

Alexander Kempf and Olaf Korn. Market depth and order size. *Journal of Financial Markets*, 2(1):29 – 48, 1999.

Ralf H.J.M. Kurvers, Babette Eijkelenkamp, Kees van Oers, Bart van Lith, Sipke E. van Wieren, Ronald C. Ydenberg, and Herbert H.T. Prins. Personality differences explain leadership in barnacle geese. *Animal Behaviour*, 78(2):447 – 453, 2009.

Albert S. Kyle. Continuous auctions and insider trading. *Econometrica*, 53(6):1315–1335, 1985.

Josef Lakonishok, Andrei Shleifer, and Robert W. Vishny. The impact of institutional trading on stock prices. *Journal of Financial Economics*, 32(1):23 – 43, 1992.

Youngki Lee, Luís A. Nunes Amaral, David Canning, Martin Meyer, and H. Eugene Stanley. Universal features in the growth dynamics of complex organizations. *Phys. Rev. Lett.*, 81 (15):3275–3278, Oct 1998.

David J. Leinweber. Using information from trading in trading and portfolio management: Ten years later. Social Science Working Paper 1135, California Institute of Technology, May 2002.

M. Levy, S. Solomon, and G. Ram. Dynamical explanation for the emergence of power law in a stock market model. *International Journal of Modern Physics C*, 7:65–72, 1996. doi: 10.1142/S0129183196000077.

Fabrizio Lillo and J. Doyne Farmer. The long memory of the efficient market. *Studies in Nonlinear Dynamics & Econometrics*, 8(3):1, 2004.

Thomas F. Loeb. Is there a gift from small-stock investing? *Financial Analysts Journal*, 47(1):39–44, Jan. - Feb. 1991.

F. M. Longin. The asymptotic distribution of extreme stock market returns. *Journal of Business*, 69:383–408, 1996.

Thomas Lux. The stable paretian hypothesis and the frequency of large returns: an examination of major german stocks. *Applied Financial Economics*, 6(6):463–475, 1996.

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

Burton G. Malkiel. Returns from investing in equity mutual funds 1971 to 1991. *The Journal of Finance*, 50(2):549–572, Jun 1995b. ISSN 0022-1082.

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

Kaushik Matia, Dongfeng Fu, Sergey V. Buldyrev, Fabio Pammolli, Massimo Riccaboni, and H. Eugene Stanley. Statistical properties of business firms structure and growth. *EUROPHYS LETT*, 67:498, 2004.

Kaushik Matia, Luis A. Nunes Amaral, Marc Luwel, Henk F. Moed, and H. Eugene Stanley. Scaling phenomena in the growth dynamics of scientific output: Research articles. *J. Am. Soc. Inf. Sci. Technol.*, 56(9):893–902, 2005. ISSN 1532-2882.

Michael Mitzenmacher. A brief history of generative models for power law and lognormal distributions. *Internet Mathematics*, 1(2):226–251, 2003.

Elliott W. Montroll and Michael F. Shlesinger. On 1/f noise and other distributions with long tails. *Proceedings of the National Academy of Sciences of the United States of America*, 79(10):3380–3383, 1982.

Esteban Moro, Javier Vicente, Luis G. Moyano, Austin N. Gerig, J. D. Farmer, Gabriella Vaglica, Fabrizio Lillo, and Rosario N. Mantegna. Market impact and trading profile of large trading orders in stock markets. *Physical Review E.*, 80(6), 2009.

Mate Nagy, Zsuzsa Akos, Dora Biro, and Tamas Vicsek. Hierarchical group dynamics in pigeon flocks. *Nature*, 464:890–893, 2010.

M. E. J. Newman. The structure and function of complex networks. *SIAM Review*, 45:167, 2003.

J. P. Nolan. *Stable Distributions - Models for Heavy Tailed Data.* Birkhäuser, Boston, 2009. In progress, Chapter 1 online at academic2.american.edu/∼jpnolan.

Romualdo Pastor-Satorras and Alessandro Vespignani. Epidemic spreading in scale-free networks. *Phys. Rev. Lett.*, 86(14):3200–3203, Apr 2001. doi: 10.1103/PhysRevLett.86.3200.

Vasiliki Plerou, Luis A. Nunes Amaral, Parameswaran Gopikrishnan, Martin Meyer, and H. Eugene Stanley. Ivory tower universities and competitive business firms. *Nature*, 400: 433, 1999a.

Vasiliki Plerou, Parameswaran Gopikrishnan, Luís A. Nunes Amaral, Martin Meyer, and H. Eugene Stanley. Scaling of the distribution of price fluctuations of individual companies. *Phys. Rev. E*, 60(6):6519–6529, Dec 1999b.

Vasiliki Plerou, Parameswaran Gopikrishnan, Xavier Gabaix, and H. Eugene Stanley. Quantifying stock-price response to demand fluctuations. *Phys. Rev. E*, 66(2):027104, Aug 2002.

Boris Podobnik, Davor Horvatic, Fabio Pammolli, Fengzhong Wang, H. Eugene Stanley, and I. Grosse. Size-dependent standard deviation for growth rates: Empirical results and theoretical modeling. *Physical Review E (Statistical, Nonlinear, and Soft Matter Physics)*, 77(5):056102, 2008. doi: 10.1103/PhysRevE.77.056102.

Joshua M. Pollet and Mungo Wilson. How does size affect mutual fund behavior? *The Journal of Finance*, 63(6):2941–2969, 2008.

Gibrat R. *Les inégalités économiques.* Librairie du Recueil Sirey, 1931.

William J. Reed. The pareto, zipf and other power laws. *Economics Letters*, 74(1):15–19, December 2001.

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

Sidney I. Resnick. *Heavy-tail phenomena: probabilistic and statistical modeling.* Springer, 2007.

M. Riccaboni, F. Pammoli, S. V. Buldyrev, L. Pontace, and H. E. Stanley. The size variance relationship of business firm growth rates. *Proc. Natl. Acad. Sci.*, 105:19595–19600, 2008.

B. M. Roehner. Fifteen years of econophysics: worries, hopes and prospects. *ArXiv e-prints*, April 2010.

Hernan D. Rozenfeld, Diego Rybski, Jose S. Andrade, Michael Batty, H. Eugene Stanley, and Hernan A. Makse. Laws of population growth. *Proceedings of the National Academy of Sciences*, 105(48):18702–18707, 2008.

Yonathan Schwarzkopf and J. D. Farmer. What Drives Mutual Fund Asset Concentration? Working paper series, SSRN eLibrary, 2010a.

Yonathan Schwarzkopf and J. Doyne Farmer. Time evolution of the mutual fund size distribution. SFI Working Paper Series 08-08-31, Santa Fe Institute, 2008.

Yonathan Schwarzkopf and J. Doyne Farmer. Empirical study of the tails of mutual fund size. *Phys. Rev. E*, 81(6):066113, Jun 2010b. doi: 10.1103/PhysRevE.81.066113.

Yonathan Schwarzkopf, Robert Axtell, and J. D. Farmer. The Cause of Universality in Growth Fluctuations. Working paper series, SSRN eLibrary, 2010.

A. O. Schweiger, S. V. Buldyrev, and H. E. Stanley. A transactional theory of fluctuations in company size. Quantitative Finance physics/0703023, arXiv.org, March 2007.

William F. Sharpe. Capital asset prices: A theory of market equilibrium under conditions of risk. *The Journal of Finance*, 19(3):425–442, September 1964.

A. Christian Silva and Victor M. Yakovenko. Temporal evolution of the "thermal" and "superthermal" income classes in the usa during 19832001. *EPL (Europhysics Letters)*, 69(2):304, 2005.

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

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

Simon, H. A. and Bonini, Charles P. The size distribution of business firms. *The American Economic Review*, 48(4):607–617, Sep 1958. ISSN 0002-8282.

Sirri, Erik R. and Tufano, Peter. Costly search and mutual fund flows. *The Journal of Finance*, 53(5):1589–1622, Oct 1998. ISSN 0022-1082.

S. Solomon and P. Richmond. Power laws of wealth, market order volumes and market returns. *Physica A Statistical Mechanics and its Applications*, 299(1–2):188–197, October 2001.

H. E. Stanley, V. Afanasyev, L. A. N. Amaral, S. V. Buldyrev, A. L. Goldberger, S. Havlin, H. Leschhorn, P. Maass, R. N. Mantegna, C. K. Peng, P. A. Prince, M. A. Salinger, M. H. R. Stanley, and G. M. Viswanathan. Anomalous fluctuations in the dynamics of complex systems: from dna and physiology to econophysics. *Physica A: Statistical and Theoretical Physics*, 224(1-2):302 – 321, 1996. ISSN 0378-4371. Dynamics of Complex Systems.

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

M. H. R. Stanley, L. A. N. Amaral, S. V. Buldyrev, S. Havlin, H. Leschhorn, P. Maass, M. A. Salinger, and H. E. Stanley. Scaling behaviour in the growth of companies. *Nature*, 379:804–806, February 1996. doi: 10.1038/379804a0.

Jeremy C. Stein. Information production and capital allocation: Decentralized versus hierarchical firms. *The Journal of Finance*, 57(5):1891–1921, 2002.

Livio Stracca. Delegated portfolio management: A survey of the theoretical literature. *Journal of Economic Surveys*, 20(5):823–848, December 2006.

John Sutton. Gibrat's legacy. *Journal of Economic Literature*, 35(1):40–59, Mar 1997. ISSN 0022-0515.

John Sutton. The variance of firm growth rates: The scaling puzzle. STICERD - Economics of Industry Papers 27, Suntory and Toyota International Centres for Economics and Related Disciplines, LSE, Sep 2001.

N. Torre. Barra market impact model handbook. Technical report, BARRA Inc., Berkeley, 1997.

Duncan J. Watts and Peter S. Dodds. Influentials, networks and public opinion formation. *Journal of consumer research*, December 2007.

Russ Wermers. Mutual fund herding and the impact on stock prices. *The Journal of Finance*, 54(2):581–622, 1999.

Matthieu Wyart and Jean-Philippe Bouchaud. Statistical models for company growth. *Physica A: Statistical Mechanics and its Applications*, 326(1-2):241 – 255, 2003.

V. M. Zolotarev. *One-Dimensional Stable Distributions (Translations of Mathematical Monographs - Vol 65).* American Mathematical Society, 1986.