Assessing Models using Monte Carlo Simulations

Sangmin Oh* Jessica Wachter†

November 12, 2017

Abstract

We establish a framework for assessing the validity of a model using Monte Carlo simulations and inferences based on sampling distributions. Using this framework, we find that geometric brownian motion underestimates the skewness in pooled monthly returns and in the cross-section of returns, but it overestimates the asymmetry in wealth creation by individual stocks. This result is robust to simulation specifications and the choice of metrics to represent wealth. Our paper also represents an often overlooked departure from the traditional way of validating asset pricing models, in which implications are derived, parameters calibrated, and point magnitudes compared to empirical data. Instead, we leverage the cross-sectional features and asymmetry present in equity returns to assess the probability that the given model can generate our realized stock market.

Keywords: Monte Carlo, Brownian Motion, Skewness, Wealth Creation
JEL Codes: G11, G12

*The Wharton School, University of Pennsylvania. Email: sangmino@wharton.upenn.edu
†The Wharton School, University of Pennsylvania. Email: jwachter@wharton.upenn.edu

1 Introduction

In this paper, we establish a framework for assessing the validity of a given model using Monte Carlo simulations and inferences based on sampling distributions. Using this framework, we find that geometric brownian motion underestimates the skewness in pooled monthly returns and in the cross-section of returns, but it overestimates the asymmetry in wealth creation by individual stocks. This result is robust to simulation specifications and the choice of metrics to represent wealth.

Our first contribution is in expanding the set of statistics used to compare the model's implications to empirical data. We introduce quantitative measures for three categories of empirical data: the distribution of pooled returns, time-series of monthly cross-sectional moments, and the distribution of wealth creation by individual firms. The use of such extensive set of statistics allows us to specifically identify areas in which a given model succeeds and fails.

Our second contribution is to provide a quantitative measure of how effectively the model captures empirical data. Albeit simple, this approach represents an often overlooked departure from the traditional way of validating asset pricing models, in which implications are derived, parameters calibrated, and point magnitudes compared to empirical data. Instead, we leverage the cross-sectional features and asymmetry present in equity returns to assess the probability that the given model can generate our realized stock market. We achieve this through hypothesis testing via the sampling distribution of statistics obtained from the simulations. Instead of simply characterizing the discrepancy as a puzzle, we can therefore specify the degree to which the model's implications are unrealistic. It then becomes possible to examine the relative performance of competing models for a given metric by comparing the magnitude of p-value from each test.

We demonstrate the usefulness of our framework using geometric brownian motion as a test case. We consider two strands of simulations - the first category considers a subset of firms that have been present throughout the period January 1970 to December 2000. The second category of simulations samples 2,440 firms from the entire CRSP universe from July 1926 to December 2016 in order to match the median number of firms in the cross-section. We show that both types of simulations assuming geometric brownian motion fail to generate most of the statistics examined in this paper. Surprisingly, they do succeed in matching the percentage of stock returns that are positive, the magnitude of which comes as a surprise to many.[1] The model also succeeds in generating a fat-tailed distribution, while more severe, of individual firm's wealth creation.

The results of our simulations assuming geometric brownian motion also yields a new empirical puzzle. The simulations heavily underestimate the skewness in both pooled distribution of returns and the monthly cross-section. Yet they overestimate the asymmetry in wealth creation by individual stocks. Together, they suggest that the asymmetric distribution of firm size does not necessitate a similar asymmetric distribution of returns. In fact, there seems to be another force at work other than the distribution of returns that gives rise to the asymmetric distribution of firm sizes and wealth creation, and any serious model should incorporate a process that successfully reconciles this discrepancy.

The remainder of the paper proceeds as follows. In section 2, we review the salient characteristics of the U.S. stock market that have been previously explored in the literature. Section 3 contains the details of simulating stock prices under the assumption of geometric brownian motion. In section 4, we present our analysis using sampling distributions from simulations, and section 5 concludes.

[1]See Bessembinder (2017)

2 Background

2.1 Distribution of Equity Returns

The mean of equity returns — and its excess over the risk-free returns — is the most widely explored aspect of stock returns. In particular, the equity premium and volatility puzzles refer to the inability of standard economic theory to generate the first and second unconditional moments of equity returns: historically, equity returns have been too high and too volatile[2]. This failure has led to many modifications to the features of the original model, including alternate assumptions on preferences[3], rare disasters [4], liquidity risk [5], and market imperfections [6]. These consumption-based approaches have primarily focused on explaining various puzzles by matching analytical implications of each model to the moments of equity returns.

Many studies also seek to explain why different securities earn vastly different returns on average. The Capital Asset Pricing Model (CAPM) has been the pioneering explanation for the cross-sectional differences. The poor empirical performance of the CAPM [7], however, has led to a set of new unconditional multi-factor models.[8]. The validity of these models have been primarily assessed by examining the R-squared and significance of the intercept in a regression framework. Yet the presence of extensive data mining and the lack of motivating theory have rendered many of such models subject to doubt.

Another set of papers looks at the cross-sectional *dispersion* or *volatility* which

[2]See Mehra and Prescott (1985)

[3]See Abel (1990), Bansal and Yaron (2004), Campbell and Cochrane (1999), Epstein and Zin (1991)

[4]See Barro (2006), Gabaix (2012), and Wachter (2013)

[5]See Alvarez and Jermann (2001)

[6]See Aiyagari and Gertler (1991), Constantinides, Donaldson, and Mehra (1995), Heaton and Lucas (1996)

[7]See Fama and French (2004)

[8]See Carhart (1997), Ang, Hodrick, Xing, and Zhang (2006), Daniel and Titman (2006)

captures the distribution of individual stock returns around the market return. Recent literature has been paying attention to its role in forecasting market returns[9], implications for asset managers[10], and pricing of the cross-section of stock returns[11].

There is also an increasing interest regarding the skewness in asset returns, which stems from the observation that unconditional returns distribution cannot be adequately characterized by mean and variance alone [12]. Kraus and Litzenberger (1976) first extend the CAPM to incorporate the effect of skewness on valuation, illustrating that prior empirical findings interpreted as anomalies were due to the omission of a higher moment variable. Since then, scholars have extensively assessed both individual stocks return skewness and the co-skewness of stock returns with the market.[13]

2.2 Distribution of Value Creation

The cumulative value or wealth created by individual firms over an extended period of time also poses an interesting empirical observation. With no clear consensus, researchers have explored different metrics to capture its essence, including firm size, cumulative return, and aggregate wealth creation.

One clear measure of firm's lifetime growth is its growth in size or market capitalization. The skewness in firm sizes - small number of large firms and large number of small firms - has been robust over time, immune to new firm entries and bankruptcies as well as mergers and acquisitions[14]. Recently, Gabaix (2016) has effectively used the observed skewness in firm sizes to examine how standard economic theories fit with the empirical data.

[9]See Garcia et al. (2014), Goyal and Santa-Clara (2003)
[10]See Gorman, Sapra, and Weigand (2010)
[11]See Verousis and Voukelatos (2015)
[12]See Harvey and Siddique (2000)
[13]See Mitton and Vorkink (2007), Conrad, Dittmar, and Ghysels (2013), and Amaya, Christoffersen, Jacobs, and Vasquez (2015).
[14]See Axtell (2011)

Another intuitive measure of firm's value creation is its cumulative return over a period of time, and similar skewness observed in firm sizes is also exhibited. For example, Savor and Wilson (2013) find that over 60% of the cumulative annual excess return is earned on just 13% of the trading days when important macroeconomic news is scheduled. Also, Bessembinder (2017) finds that lifetime holding period returns are dominated by a very small number of firms.

However, the use of a cumulative holding period return as a measure of aggregate wealth creation is not entirely accurate. The cumulative return calculation assumes that equity investors reinvest dividends but make no intermediate transactions after the initial purchase of shares. Bessembinder (2017) illustrates one way of circumventing this limitation by creating a separate measure of dollar wealth creation of each firm. Bessembinder quantitatively measures the investor's final wealth in excess of the wealth the investor would have attained had she invested entirely in the risk-free asset. Using this metric, he finds that the entire wealth creation in the U.S. stock market is attributable to a mere four percent of listed stocks.

3 Simulation with Geometric Brownian Motion

3.1 Motivation

One goal of this paper is to examine the implications of geometric brownian motion in light of empirical evidence. Instead of deriving analytically the expressions for the market premium, skewness, and aggregate value creation, we adopt an approach based on Monte Carlo simulations. Specifically, we simulate the stock prices and compute a set of pre-determined statistics. Repeating the simulations N times yields a sampling distribution with size N, and we can use the resulting sampling distribution to make inferences regarding the real stock market data.

This approach yields two major benefits. First, it enables a more robust quantitative analysis. When analytical expressions are unattainable, the benefits of Monte Carlo are obvious. Even when such analytical derivations are possible, inference using the sampling distribution allows us to quantify how likely — or unlikely — the current stock market can arise from the assumptions of our simulation. Second, this approach allows us to explore a wider variety of scenarios than what historical data can provide. Stochastic stock price growth inherently implies that the observed stock market represents only one realization; Monte Carlo allows us to overcome this limitation and leverage the power of large numbers.

Simulating stock prices necessitates an assumption regarding the time-series behavior of stock prices. We start with the simplest and most widely used model: geometric brownian motion. Its biggest merit is its non-negative value and the independence of expected returns from the value of the process. One can also imagine incorporating time-changing volatility or exposures to disaster risk, but here we focus on brownian motion with constant volatility and continuous price processes.

We consider two types of simulations. For the first type, we examine all stocks whose returns are available in CRSP throughout the period from January 1970 to December 2000. We restrict our universe to stocks with less than 5 days of daily returns missing, since missing data renders the calculation of lifetime wealth creation inaccurate. Applying the restriction yields the final universe which consists of 431 stocks that have been in existence from 1970 to 2000.

In the second type of simulations, we take a slightly different approach. We consider all stocks with at least 60 months of returns data available in CRSP from July 1926 to December 2016, yielding a universe of 16,087 firms. We also set the number of stocks in our simulation to match them median number of firms at the start of each month throughout CRSP history: 2,440. Therefore, each simulation starts by drawing 2,440 firms from the universe of 16,087 firms, whose price movements are simulated from July 1926 to December 2016.

3.2 Parameter Estimation

Geometric brownian motion starts with the following stochastic differential equation:

$$\frac{dS_i(t)}{S_i(t)} = \mu_i dt + \sigma_i dW(t) \tag{1}$$

where $S_i(t)$ is the price of a security i at time t, μ_i the drift parameter, σ_i the volatility parameter, and $W_i(t)$ the value of a Wiener process at time t. Its major implication is that log returns are normally distributed:

$$\log R_{i,t} \sim N\left(\left(\mu_i - \frac{1}{2}\sigma_i^2\right) t, \sigma_i^2 t\right) \tag{2}$$

which shows that μ_i and σ_i need to be estimated to conduct simulations of stock price. We present three different methods of estimating μ_i (see Appendix B). The first method invokes the CAPM for log returns and yields the following expression for μ_i:

$$\mu_i = (1 - \beta_i)E\left[\log R_f\right] + \beta_i \log E[R_M] \tag{3}$$

The second and third methods are based on direct estimates from returns in our sample period. They are obtained from equations (4) and (5) respectively:

$$\mu_i = E\left[\log R_i\right] + \frac{1}{2}\sigma_i^2 \tag{4}$$

$$\mu_i = \log[E[R_i]] \tag{5}$$

The market risk premium is estimated directly by taking the mean of returns. For the volatility parameter σ, we base our estimate on historical log returns:

$$\hat{\sigma_i} = \sqrt{\frac{1}{n-1}\sum_{t=1}^{n}(r_{i,t} - \bar{r_i})^2} \tag{6}$$

where $r_{i,t}$ is the log return at time t and $\bar{r_i}$ is the average of the n returns. Similarly, we estimate β_i for each firm using daily returns from the following regression:

$$\log R_{i,t} = \alpha_i + \beta_i \log R_{M,t} + \epsilon_{i}, \tag{7}$$

3.3 Simulation Overview

We describe the process in which stock prices are simulated. We consider two types of simulations: for the first type, we examine all stocks that had been present in CRSP throughout the period from January 1970 to December 2000 with less than 5 days of daily returns missing; for the second type, we consider all stocks with at least 60 months of returns in CRSP from July 1926 to December 2016, yielding a universe of 16,087 firms. For convenience, we will refer to each as Type 1 and Type 2 simulations.

In implementing both simulations, we wish to differentiate a market-wide shock from an idiosyncratic shock to each firm. Having estimated the σ_i parameter for firm i, we can decompose it into a systematic component and an idiosyncratic component:

$$\sigma_i^2 = \beta_i^2 \sigma_M^2 + \sigma_\epsilon^2 \tag{8}$$

where σ_M denotes the market return volatility and σ_ϵ denotes idiosyncratic volatility. Therefore, the stock price at time t following geometric brownian motion can be expressed as:

$$S_i(t) = S_i(0) \cdot \exp\left(\left(\mu_i - \frac{1}{2}\sigma_i^2\right)t + \beta_i \sigma_M \epsilon_M + \sigma_\epsilon \epsilon_i\right) \tag{9}$$

where ϵ_M and ϵ_i represent the market shock and idiosyncratic shock for firm i.

Type 1 simulation consists of daily stock price simulations for all 431 stocks between January 1970 to December 2000. We conduct the simulation 300 times, and each simulation has the same set of 431 stocks. For type 2 simulations, our universe of eligible stocks has 16,087 firms whereas we would like to restrict the number of stocks in any given simulation to 2,440. Therefore, for each simulation, we start by drawing 2,440 stocks without replacement. Furthermore, to reflect the history of each firm in our data, we assign the probability of stock i being drawn as

$$\frac{N_i}{\sum_{i=1}^{16,087} N_i}$$

where N_i refers to the number of months the security is available in CRSP. Once the 2,440 firms are selected for each simulation, the price process follows equation (9)

4 Tests with Sampling Distribution

We conduct tests of geometric brownian motion using the sampling distributions from the simulations. Our empirical testing ground is the returns and wealth creation data in CRSP. The steps we take are the following:

1. Choose a statistic ζ. It can be any statistic that can be computed from the returns and wealth creation data. In our study, we examine statistics pertinent to three broad categories of data: pooled monthly returns, time-series of monthly cross-sectional skews, and distribution of wealth creation.

2. Compute ζ_{data} from the CRSP data that corresponds to the simulations.

3. For each of the N stock market simulations, compute ζ^i_{sim} corresponding to the ith simulation. Consequently, we obtain a sampling distribution of ζ^i_{sim} consisting of N independent observations.

4. We test the null hypothesis that ζ_{data} represents a random sample from the distribution of ζ^i_{sim} . The rejection of the null hypothesis implies that the simulations are unlikely to generate the observed stock market outcome. In other words, it rejects the assumption underlying the simulations.

Expanding on previous studies in the literature that focus on a select group of statistics to test the validity of models, we employ an extensive set of measures. The first category of statistics pertains to the pooled monthly returns of stocks. We examine the mean, standard deviation, skewness, and percentage of returns that are positive.

The second category of statistics is based on monthly cross-sectional skewness, defined as the skewness of monthly returns for the firms in any given month. Computing the skewness for each month therefore yields a time-series of monthly cross-sectional skewness. We examine the mean and standard deviation of such time-series, which captures the persistence of the returns skewness in the cross-section.

Finally, we are interested in the distribution of wealth creation by individual firms. Wealth is measured in three ways: market cap growth, cumulative returns, and aggregate investor wealth computed using a metric in Bessembinder (2017). Given the asymmetric distribution of wealth creation observed in empirical data, the first statistic we compute is the parameter α of the power law distribution fitted to the distribution of individual firm's wealth creation. The estimate of α, however, is sensitive to the choice of cutoff (see Appendix D). To overcome this potential uncertainty, we compute a second statistic, the percentage of wealth contributed to overall wealth by the top ten stocks with greatest wealth creation. The higher the percentage, the more concentrated the wealth creation in a few number of stocks and the greater the asymmetry.

4.1 Pooled Monthly Returns

Table 1 reports selected statistics for the pooled CRSP common stock returns for different time horizons and different universe. For all CRSP stocks, monthly returns are highly skewed with skewness greater than 6 in both the population period and our sample period. Consistent with Bessembinder (2017), we verify that more than half of the monthly returns are negative.

When the universe is restricted to only the 431 stocks, the mean and median monthly return increases while the standard deviation decreases. The skewness drops significantly from 6.418 to 6.608 to 0.828, still indicating a positive skew with a smaller magnitude. This contrast is not surprising — the 431 stocks have long lives, having been in existence throughout the 30 years of our sample period. The monthly returns seem to be clustered around a higher mean with a lower probability of obtaining extreme positive returns. Albeit interesting, the empirical distribution of the pooled returns is of a secondary concern to this study; the primary objective is to examine its features in the sampling distribution obtained from our simulations.

Table 2 reports the p-values corresponding to each statistic in Type 1 Simulations. Simulations in Panel A are conducted with the μ parameter obtained using CAPM; simulations in panel B and C are conducted with direct estimates from individual stock returns. In all three panels, the null hypothesis is rejected at the 0.01 significance level for four of the six statistics computed on pooled monthly returns, indicating that they cannot be considered a random sample from the sampling distribution. The two statistics for which the null hypothesis cannot be rejected are the mean of pooled monthly returns and the percentage of returns that are positive.

The sign of the z-scores is worth a closer look. The z-score of the standard deviation is extremely negative, indicating that the standard deviations in the sampling distribution are mostly greater than the value observed in our data. On the other hand, the z-score of the skewness is extremely positive for all three methods, implying that our stock market exhibits skewness much greater than the skewness from our simulations. This contrast is puzzling as both standard deviation and skewness measure the dispersion of returns.

For the percentage of stock returns greater than the value-weighted and equal-weighted market return, the null hypothesis is also rejected; the simulations underestimate the percentage of stocks that outperform the market. Also, note that the simulations also underestimate the skewness of stock returns. The two observations initially seem at contrast: simulations underestimate the number of firms that outperform the market, yet they also underestimate the probability of obtaining extreme positive returns. This calls for a mechanism in which the minority of firms beating the market cancel the effect of the majority of firms underperforming the market, but without extreme positive returns. Such mechanism is not immediately obvious.

Table 3 reports the z-scores and p-values corresponding to each statistic in type 2 Simulations. In the first panel when CAPM is used to estimate the μ parameter, we see that the simulations seem to reasonably generate the percentage of positive monthly returns observed in CRSP. For all the remaining statistics, the null hypothesis is rejected at the 0.01 significance level. In particular, the simulations fail egregiously at generating the skewness observed in CRSP data, whose p-value is orders of magnitude smaller than those for other rejected statistics. This observation indicates that the skewness of pooled monthly returns is a cross-sectional feature that a model should seek to be able to match.

In sum, the simulations seem to fail at generating, with reasonable probability, majority of the statistics on pooled monthly returns. The only area in which they seem to somewhat succeed is the mean and percentage of positive returns. In particular, both sets of simulations fail notably when the skewness of pooled returns is considered, and this observation renders new significance to skewness as an important cross-sectional feature of stock returns.

4.2 Time-series of Monthly Cross-sectional Skewness

We define *monthly cross-sectional skew* of month t for n firms as the following:

$$\gamma_t^{cs} = \frac{\frac{1}{n}\sum_{i=1}^{n}(r_{i,t} - \bar{r}_t)^3}{\left[\frac{1}{n}\sum_{i=1}^{n}(r_{i,t} - \bar{r}_t)^2\right]^{3/2}} \tag{10}$$

where $r_{i,t}$ is firm i's monthly return for month t and $\bar{r}_t = \sum_{i=1}^{n} r_{i,t}$. Unlike time-series skewness, cross-sectional skewness captures the dispersion of stock returns as a snapshot at each point in time. Computing this metric for T months yields a time-series of monthly cross-sectional skew. The statistics that we examine is the mean and standard deviation of the time-series, shown in equations (11) and (12) respectively:

$$\frac{1}{T}\sum_{t=1}^{T}\gamma_t^{cs} \tag{11}$$

$$\sqrt{\frac{1}{T}\sum_{t=1}^{T}\left(\gamma_t^{cs} - \frac{1}{T}\sum_{t=1}^{T}\gamma_t^{cs}\right)^2} \tag{12}$$

Table **??** reports selected statistics for the pooled CRSP common stock returns for different time horizons and different universe. For all CRSP stocks, the mean of monthly cross-sectional skewness is 1.490 for the population period and 3.293 for the sample period. The standard deviation of monthly cross-sectional skewness is quite high, with 2.012 and 4.058 respectively. Of the 1,086 months from September 1926 to December 2016, 72 months exhibit negative cross-sectional skewness. While the skewness is generally persistent throughout our time period, it is misleading to argue that the distribution of returns for any given month is positively skewed.

When the universe is restricted to only the 431 stocks, the mean of monthly cross-sectional skewness drops to 0.774 and the standard deviation to 1.041. Intuitively, the skewness should increase as the number of firms in the universe increases — the more firms there are to attain extreme returns, the more likely that the returns will be positively skewed. For this argument to hold, the likelihood of an extreme positive return should be on average greater than the likelihood of an extreme negative return. While this topic warrants a stricter investigation, explanations in favor may include the lower bound on stock prices[15], increased correlations during crises[16], and positive skewness in sector-specific return shocks[17].

Table 4 reports the p-values corresponding to the mean and standard deviation of monthly cross-sectional skew corresponding to Type 1 Simulations. In all three panels, the null hypothesis is strongly rejected - the simulations fail to generate values that can reasonably correspond to the observed value in our data. Most notably, the simulations severely underestimate the average cross-sectional skewness of the actual stock market, indicating that geometric brownian motion alone cannot generate the characteristics of monthly cross-sectional skewness observed in the actual stock market.

4.3 Wealth Creation

In this section, we focus on three different metrics of wealth creation: growth in market cap, cumulative return, and aggregate wealth creation as measured by equation in Bessembinder (2017). First, the market cap growth ΔMC_t from time 0 to time t is

[15] Stock prices cannot go below zero.
[16] See Campbell, Koedijk, and Kofman (2002)
[17] See Duffee (2001)

defined as

$$\Delta MC_t = \frac{p_t N_t}{p_0 N_0} \tag{13}$$

where p_t and p_0 are the stock prices at time t and 0 and N_t and N_0 are the number of shares outstanding at time t and time 0. Furthermore, the cumulative return CR_t from time 0 to time t is defined as

$$CR_t = \prod_{i=0}^{t}(1 + r_i) - 1 \tag{14}$$

where r_i denotes the holding period return for month i. Finally, the aggregate wealth creation AWC_t from time 0 to time t is given as:

$$AWC_t = \prod_{t=1}^{T}\left[I_{t-1}(r_t - r_f)(1 + r_f)^{T-t}\right] \tag{15}$$

in which the beginning of period market capitalization is used in the role of I_t. The idea behind the metric and its derivation is shown in Appendix C. All three metrics are computed using CRSP data.

Table 6, 8, and **??** report the wealth creation of each firm from January 1970 to December 2000, listing the 10 stocks with the greatest wealth creation among the 431 firms. For all three metrics, the asymmetric nature of wealth creation is clear: ten stocks account for 39.5% of total market cap growth, 19.5% of total cumulative returns, and 41.3% of aggregate wealth creation as measured by equation (15). This observation is more shocking sincet he ten stocks represent in number a mere 2% of the 431 firms in our universe.

To quantify the degree of asymmetry exhibited in these distribution, we fit a power law distribution to the data on wealth creation and estimate the associated coefficient.

We estimate the parameter using maximum likelihood. [18] The details of the estimation can be found in Appendix D.

Table 12 reports the p-values corresponding to the estimated α in type 1 Simulations. For market cap growth, the null hypothesis is not rejected at the 0.01 significance level in all three panels, indicating that the asymmetric distribution of market cap growth in our stock market can reasonably be attained from our simulations. On the other hand, the null hypothesis is rejected when wealth creation is measured using cumulative return and equation (15). Furthermore, the sign of the z-scores tells us that the sampling distribution from the simulations underestimates the α. Smaller α implies a larger probability of obtaining extreme values - in this case, a larger probability that a few firms are responsible for a majority of wealth creation. In other words, the simulations generate scenarios in which the asymmetry in wealth creation is much more severe than what we observe in CRSP data.

For the type 1 simulations, table 19 also reports the p-values corresponding to the percentage contributed to overall wealth by the top ten stocks with the greatest wealth creation. The analysis is not carried out for equation (15) as many simulations yield absurdly positive or negative values for the statistic.[19]

Evidence in this table is somewhat dubious. We can only reject the null hypothesis for methods 2 and 3 in which the cumulative return is used as a measure of wealth creation. Comparison against the percentile values implies that the simulations overestimate the contribution total wealth by the top ten stocks. This fact, consistent with the observation from the table of power law coefficients, is quite surprising: with just geometric brownian motion, the simulations seem to produce scenarios in which

[18] Methods based on a least squares fit are not suitable to estimate the parameter because they require additional assumptions about the data set. See Hanel et al. (2017) for more discussion.

[19] Because equation (15) is based on *excess* returns, it is entirely possible for the wealth created to be negative. When there are many firms with total negative wealth created, the percentage contributed by the top ten stocks cannot be calculated, thereby rendering inference on it undesirable.

the concentration of wealth creation is much more severe than what the actual stock market exhibits.

Similar analysis is carried out for type 2 simulations. Table 15 first reports the p-values corresponding to the estimated α. The first column represents the power law distribution coefficient estimated for 16,087 firms that constitute our universe, and the next columns illustrate the distribution of the same parameters estimated from our simulations. Using market cap growth and the metric in Bessembinder (2017) as a measure of wealth creation, we see that the null hypothesis is rejected; the simulations produce a more severe asymmetry in wealth creation. The only case in which the simulations successfully generate the empirical stock market is when cumulative return is used as a wealth metric with CAPM.

One possible objection to the preceding analysis is that the empirical calculations are based on 16,087 firms, whereas each simulation only contains 2,440 firms. To address this concern, we conduct a similar analysis based on firm sizes that explicitly matches the number of firms. Specifically, we choose a time point in CRSP such that the number of firms in existence almost matches the median number of firms – January 1972 with 2,435 firms – and examine the distribution of firm sizes. We fit a power law distribution to the firm sizes and compare the coefficient to the distribution of similar coefficients obtained from the simulations. Table **??** reports the p-values corresponding to the aforementioned analysis. Once again, we see that the simulations severely underestimate the estimated α and therefore produces a wealth asymmetry much greater than the one found in empirical data. We reject the null hypotheses at 0.01 significance level for all three methods of parameter estimation.

In sum, we find that the simulations generated by assuming geometric brownian motion produce scenarios in which wealth creation is more concentrated and asymmetric than the actual stock market. Consequently, this observation shows that such asymmetry is a feature of our stock market that does not mandate additional shocks such as the entry of new firms or disastrous shocks.

5 Conclusion

In this paper, we illustrate the success and limitations of geometric brownian motion by employing a wider variety of statistics on empirical stock market data and hypothesis testing using sampling distributions. Simulations assuming geometric brownian motion fail to generate most of the statistics examined in this paper, especially when it comes to the skewness of monthly stock returns. But it does seem to be sufficient in generating a fat-tailed distribution of market cap growth during the sample period. This implies that the asymmetric distribution of firm sizes observed in the market does not necessitate a similar asymmetric distribution of returns.

One major puzzle arises from the results of our simulations. The simulations heavily underestimate the skewness in both pooled distribution of returns and the monthly cross-section. Yet they overestimate the asymmetry in wealth creation by individual stocks. Together, they suggest that the asymmetric distribution of firm size does not necessitate a similar asymmetric distribution of returns. In fact, there seems to be another force at work other than the distribution of returns that gives rise to the asymmetric distribution of firm sizes and wealth creation, and any serious model should incorporate a process that successfully reconciles this discrepancy.

Our simulations include some immediate directions for extension that must be noted. First, we have ignored the introduction of new listings and exclusion of firms due to delisting, spin-offs, or bankruptcy. Consequently, our simulations do not account for the large price movements associated with initial public offerings and the accompanying high volatility of initial returns.[20] Furthermore, incorporating the disappearance and separation of firms can play a key role in understanding the wealth created by each firm, as an investor with stake in these firms are significantly affected by such firm activities.

Second, we could bolster the underlying assumption by incorporating time-varying beta or volatility for individual stocks. Also, an alternate modification is including rare events that result in sudden jumps in market prices. The approach undertaken in this paper, one based on simulations and inferences using sampling distributions, can be readily applied in both cases.

[20]See Lowry, Officer, and Schwert (2010)

References

[1] S Rao Aiyagari and Mark Gertler. Asset returns with transactions costs and uninsured individual risk. *Journal of Monetary Economics*, 27(3):311–331, 1991.

[2] Fernando Alvarez and Urban J Jermann. Quantitative asset pricing implications of endogenous solvency constraints. *The Review of Financial Studies*, 14(4):1117–1151, 2001.

[3] Diego Amaya, Peter Christoffersen, Kris Jacobs, and Aurelio Vasquez. Does realized skewness predict the cross-section of equity returns? *Journal of Financial Economics*, 118(1):135–167, 2015.

[4] Andrew Ang, Robert J Hodrick, Yuhang Xing, and Xiaoyan Zhang. The cross-section of volatility and expected returns. *The Journal of Finance*, 61(1):259–299, 2006.

[5] Robert L Axtell. Zipf distribution of us firm sizes. *science*, 293(5536):1818–1820, 2001.

[6] Ravi Bansal and Amir Yaron. Risks for the long run: A potential resolution of asset pricing puzzles. *The Journal of Finance*, 59(4):1481–1509, 2004.

[7] Robert J Barro. Rare disasters and asset markets in the twentieth century. *The Quarterly Journal of Economics*, 121(3):823–866, 2006.

[8] Hendrik Bessembinder. Do stocks outperform treasury bills? 2017.

[9] John Y Campbell and John H Cochrane. By force of habit: A consumption-based explanation of aggregate stock market behavior. *Journal of political Economy*, 107(2):205–251, 1999.

[10] Rachel Campbell, Kees Koedijk, and Paul Kofman. Increased correlation in bear markets. *Financial Analysts Journal*, 58(1):87–94, 2002.

[11] Mark M Carhart. On persistence in mutual fund performance. *The Journal of finance*, 52(1):57–82, 1997.

[12] Jennifer Conrad, Robert F Dittmar, and Eric Ghysels. Ex ante skewness and expected stock returns. *The Journal of Finance*, 68(1):85–124, 2013.

[13] George M Constantinides, John B Donaldson, and Rajnish Mehra. Junior can't borrow: A new perspective on the equity premium puzzle. *The Quarterly Journal of Economics*, 117(1):269–296, 2002.

[14] Kent Daniel and Sheridan Titman. Market reactions to tangible and intangible information. *The Journal of Finance*, 61(4):1605–1643, 2006.

[15] Gregory R Duffee et al. Asymmetric cross-sectional dispersion in stock returns: evidence and implications. Federal Reserve Bank of San Francisco, 2001.

[16] Larry G Epstein and Stanley E Zin. Substitution, risk aversion, and the temporal behavior of consumption and asset returns: An empirical analysis. *Journal of political Economy*, 99(2):263–286, 1991.

[17] Eugene F Fama and Kenneth R French. The capital asset pricing model: Theory and evidence. *The Journal of Economic Perspectives*, 18(3):25–46, 2004.

[18] Xavier Gabaix. Variable rare disasters: An exactly solved framework for ten puzzles in macro-finance. *The Quarterly journal of economics*, 127(2):645–700, 2012.

[19] Xavier Gabaix. Power laws in economics: An introduction. *The Journal of Economic Perspectives*, 30(1):185–205, 2016.

[20] René Garcia, Daniel Mantilla-Garcia, and Lionel Martellini. Idiosyncratic risk and the cross-section of stock returns. *July*, 5:57, 2011.

[21] Larry R Gorman, Steven G Sapra, and Robert A Weigand. The cross-sectional dispersion of stock returns, alpha, and the information ratio. *The Journal of Investing*, 19(3):113–127, 2010.

[22] Amit Goyal and Pedro Santa-Clara. Idiosyncratic risk matters! *The Journal of Finance*, 58(3):975–1007, 2003.

[23] Rudolf Hanel, Bernat Corominas-Murtra, Bo Liu, and Stefan Thurner. Fitting power-laws in empirical data with estimators that work for all exponents. *PloS one*, 12(2):e0170920, 2017.

[24] Campbell R Harvey and Akhtar Siddique. Conditional skewness in asset pricing tests. *The Journal of Finance*, 55(3):1263–1295, 2000.

[25] John Heaton and Deborah J Lucas. Evaluating the effects of incomplete markets on risk sharing and asset pricing. *Journal of political Economy*, 104(3):443–487, 1996.

[26] Alan Kraus and Robert H Litzenberger. Skewness preference and the valuation of risk assets. *The Journal of Finance*, 31(4):1085–1100, 1976.

[27] Michelle Lowry, Micah S Officer, and G William Schwert. The variability of ipo initial returns. *The Journal of Finance*, 65(2):425–465, 2010.

[28] Rajnish Mehra and Edward C Prescott. The equity premium: A puzzle. *Journal of monetary Economics*, 15(2):145–161, 1985.

[29] Todd Mitton and Keith Vorkink. Equilibrium underdiversification and the preference for skewness. *The Review of Financial Studies*, 20(4):1255–1288, 2007.

[30] Robert Novy-Marx. Is momentum really momentum? *Journal of Financial Economics*, 103(3):429–453, 2012.

[31] Pavel Savor and Mungo Wilson. How much do investors care about macroeconomic risk? evidence from scheduled economic announcements. *Journal of Financial and Quantitative Analysis*, 48(2):343–375, 2013.

[32] Thanos Verousis and Nikolaos Voukelatos. Cross-sectional dispersion and expected returns. 2015.

[33] Jessica A Wachter. Can time-varying risk of rare disasters explain aggregate stock market volatility? *The Journal of Finance*, 68(3):987–1035, 2013.

Table 1: Selected Statistics on Pooled Monthly Level Returns

	All CRSP Stocks (1926.07 - 2016.12)	16,807 Select Stocks (1926.07 - 2016.12)	392 Select Stocks (1973.01 - 2016.12)
Mean (in %)	1.111	1.315	1.330
Median (in %)	0.000	0.000	0.975
Standard Deviation (in %)	16.95	16.95	10.07
Skewness	6.418	5.846	1.256
% Positive	49.32	48.94	54.78
% $\geq$ VW Mkt Return	46.39	46.71	50.28
% $\geq$ EQ Mkt Return	45.95	46.13	49.43

Source: CRSP

Notes: The table reports selected statistics on pooled CRSP common stock monthly level returns for different time horizons and different universe of stocks. The first column examines pooled monthly returns of all CRSP common stocks from July 1926 to December 2016. The second column concerns pooled monthly returns of all CRSP common stocks with at least 60 monthly returns from July 1926 to December 2016. The third column concerns pooled returns of all CRSP common stock for which less than 5 daily returns are missing between the period January 1973 to December 2016.

Table 2: Inference on Pooled Monthly Returns (Simulations with 392 Firms)

	Empirical Value	Simulated Values					Simulation Population
		Min	5th	Median	95th	Max	
Panel A: Method 1 (CAPM)							
$E[R]-1$ (in %)	1.330	0.397	0.539	0.835	1.135	1.520	0.826
$\sigma[R]$ (in %)	10.07	10.51	10.66	10.77	10.87	10.95	10.76
$skew[R]$	1.256	0.443	0.472	0.510	0.549	0.602	0.511
$E[\log R]$ (in %)	0.837	-0.168	-0.027	0.262	0.556	0.987	0.258
$\sigma[\log R]$ (in %)	9.87	10.38	10.54	10.64	10.73	10.81	10.64
$skew[\log R]$	-0.407	-0.126	-0.106	-0.079	-0.052	-0.026	-0.078
% $\log R > 0$	54.78	49.69	50.23	51.40	52.75	54.23	51.38
Panel B: Method 2 (Direct Estimation using Expectation of Log Returns)							
$E[R]-1$ (in %)	1.330	0.889	1.128	1.408	1.717	2.065	1.407
$\sigma[R]$ (in %)	10.07	10.71	10.76	10.86	10.96	11.03	10.86
$skew[R]$	1.256	0.493	0.517	0.557	0.604	0.634	0.561
$E[\log R]$ (in %)	0.837	0.323	0.556	0.834	1.142	1.476	0.831
$\sigma[\log R]$ (in %)	9.87	10.50	10.55	10.64	10.73	10.81	10.64
$skew[\log R]$	-0.407	-0.096	-0.064	-0.039	-0.011	0.013	-0.037
% $\log R > 0$	54.78	51.73	52.48	53.66	54.96	56.23	53.66

Source: CRSP and simulations

Notes: Using the universe of 392 firms, we conduct 400 monthly simulations of the stock market for the two methods of estimating μ detailed in Appendix B. Sampling distributions of each statistic are obtained from the simulations. The first column shows the statistic for the 392 firms from January 1973 to December 2016. The next five columns show the distribution of the statistic obtained from the simulations. The last column illustrates the statistic for the pooled values of 100 simulations. For each statistic, the null hypothesis is that the statistic represents a random sample from the sampling distribution.

Table 3: Inference on Pooled Monthly Returns (Simulations with 16,807 Firms)

	Empirical Value	Simulated Values					Simulation Population
		Min	5th	Median	95th	Max	
Panel A: Method 1 (CAPM)							
$E[R]-1$ (in %)	1.315	0.451	0.713	1.046	1.360	1.590	
$\sigma[R]$ (in %)	16.95	15.95	16.18	16.46	16.75	16.97	
$skew[R]$	5.846	0.864	0.893	0.946	1.045	1.180	
$E[\log R]$ (in %)	0.099	-0.858	-0.570	-0.258	0.058	0.291	
$\sigma[\log R]$ (in %)	15.44	15.64	15.86	16.11	16.37	16.61	
$skew[\log R]$	-0.195	-0.209	-0.192	-0.166	-0.142	-0.131	
% $\log R > 0$	48.94	49.42	49.22	50.14	51.06	51.63	
Panel B: Method 2 (Direct Estimation using Expectation of Log Returns)							
$E[R]-1$ (in %)	1.315	0.809	1.022	1.308	1.618	1.915	
$\sigma[R]$ (in %)	16.95	16.11	16.25	16.51	16.77	16.95	
$skew[R]$	5.846	0.823	0.864	0.921	0.997	1.164	
$E[\log R]$ (in %)	0.099	-0.491	-0.290	0.004	0.314	0.593	
$\sigma[\log R]$ (in %)	15.44	15.75	15.90	16.16	16.39	16.53	
$skew[\log R]$	-0.195	-0.248	-0.228	-0.191	-0.166	-0.153	
% $\log R > 0$	48.94	49.69	50.40	51.19	52.08	52.83	

Source: CRSP and simulations

Notes: Using the universe of 16,087 firms, we conduct 400 monthly simulations of the stock market for the two methods of estimating μ detailed in Appendix B. Each simulation consists of 2,440 stocks sampled from the universe whose probability of being drawn is proportional to the length of its returns history. Sampling distributions of each statistic are obtained from the simulations. The first column shows the statistic for the 16,087 firms from January 1973 to December 2016. The next five columns show the distribution of the statistic obtained from the simulations. The last column illustrates the statistic for the pooled values of 100 simulations. For each statistic, the null hypothesis is that the statistic represents a random sample from the sampling distribution.

Table 4: Inference on Monthly Cross-sectional Skew (Simulations with 392 Firms)

	Average Monthly Cross-sectional Skew ($\bar{\gamma}_{cs}$)	$\bar{\gamma}_{cs}$ from Simulated Values					% of Months with $\bar{\gamma}_{cs} \geq$ Max
		Min	5th	50th	95th	Max	
Panel A: Method 1 (CAPM)							
$skew[R]$	0.884	0.414	0.437	0.475	0.508	0.529	54.55
Panel B: Method 2 (Direct Estimation using Expectation of Log Returns)							
$skew[R]$	0.884	0.473	0.494	0.527	0.566	0.585	52.47

Source: CRSP and simulations

Notes: We conduct 400 monthly simulations of the stock market for both methods of estimating μ detailed in Appendix B. The first column shows the average monthly cross-sectional skewness ($\bar{\gamma}_{cs}$) for the 392 firms from January 1973 to December 2016. The next five columns illustrate the distribution of $\bar{\gamma}_{cs}$ obtained from simulations. The final column reports the percentage of months in the period January 1973 to December 2016 in which $\bar{\gamma}_{cs}$ is greater than the maximum $\bar{\gamma}_{cs}$ obtained from the simulations.

Table 5: Inference on Monthly Cross-sectional Skew (Simulations with 16,087 Firms)

	Average Monthly Cross-sectional Skew	Simulated Values					% of Months with $skew[R] \geq$ Max
		Min	5th	50th	95th	Max	
Panel A: Method 1 (CAPM)							
$skew[R]$	2.381	0.828	0.853	0.899	0.967	1.032	68.41
Panel B: Method 2 (Direct Estimation using Expectation of Log Returns)							
$skew[R]$	2.381	0.789	0.818	0.872	0.930	0.990	69.25

Source: CRSP and simulations

Notes: We conduct 400 monthly simulations of the stock market for both methods of estimating μ detailed in Appendix B. The first column shows the average monthly cross-sectional skewness ($\bar{\gamma}_{cs}$) for the 16,087 firms from July 1926 to December 2016. The next five columns illustrate the distribution of $\bar{\gamma}_{cs}$ obtained from simulations. The final column reports the percentage of months in the period July 1926 to December 2016 in which $\bar{\gamma}_{cs}$ is greater than the maximum $\bar{\gamma}_{cs}$ obtained from the simulations.

Table 6: Top Ten Market Cap Growth Among the 392 CRSP Stocks

Company Name	PERMNO	Growth	% of Total Value	Cumulative %
Southwest Airlines Co	58683	7631.95	13.67%	13.67%
Skyworks Solutions Inc	45911	3922.21	7.03%	20.69%
Applied Materials Inc	14702	2381.65	4.27%	24.96%
Unilever	28310	1522.11	2.73%	27.69%
Tyson Foods	77730	1399.29	2.51%	30.19%
Thermo Fisher Scientific Inc	62092	1360.05	2.44%	32.63%
Johnson Controls International PLC	45356	1335.74	2.39%	35.02%
Intel Corp	59328	1240.23	2.22%	37.24%
Analog Devices Inc	60871	1183.89	2.12%	39.36%
Wal Mart Stores Inc	55976	1100.91	1.97%	41.33%

Source: CRSP

Notes: The table reports market cap growth for the 392 CRSP stocks from January 1973 to December 2016. Results pertain to the 10 stocks with the greatest market cap growth. Market cap is computed as the closing price of each month multiplied by the number of shares outstanding, as available in CRSP. Market cap growth is computed as the market cap in December 2016 divided by the market cap in January 1973. The company name displayed is that associated with the PERMNO for the most recent CRSP record.

Table 7: Top Ten Market Cap Growth Among the 16,087 CRSP Stocks

Company Name	PERMNO	Growth	% of Total Value	Cumulative %
Vulcan Materials Co	15202	68993.6	9.90%	9.90%
Pepsico Inc	13856	31842.7	4.57%	14.47%
Boeing Co	19561	26294.6	3.77%	18.25%
Schlumberger Ltd	14277	16022.0	2.30%	20.54%
Altria Group Inc	13901	14123.4	2.03%	22.57%
Johnson & Johnson	22111	11154.7	1.60%	24.17%
General Dynamics Corp	12052	10968.1	1.57%	25.75%
Pfizer Inc	21936	9062.1	1.30%	27.05%
Wyeth	15667	8295.1	1.19%	28.24%
Precision Castparts Corp	63830	8081.6	1.16%	29.40%

Source: CRSP

Notes: The table reports market cap growth for the 16,087 CRSP stocks from July 1926 to December 2016. Results pertain to the 10 stocks with the greatest market cap growth. Market cap is computed as the closing price of each month multiplied by the number of shares outstanding, as available in CRSP. Market cap growth is computed as its market cap on the most recent available date divided by the market cap in the earliest available date. The company name displayed is that associated with the PERMNO for the most recent CRSP record.

Table 8: Top Ten Cumulative Return Among the 392 CRSP Stocks

Company Name	PERMNO	Growth	% of Total Value	Cumulative %
Holly Frontier Corp	32803	6312.39	6.85%	6.85%
Kansas City Southern	12650	5512.66	5.98%	12.83%
Southwest Airlines Co	58683	4510.04	4.89%	17.72%
Eaton Vance Corp	31500	2679.43	2.91%	20.63%
Wal Mart Stores Inc	55976	1621.78	1.76%	22.38%
Altria Group Inc	13901	1563.88	1.70%	24.08%
Tyson Foods Inc	77730	1422.68	1.54%	25.62%
Walgreen Boots Alliance Inc	19502	1382.41	1.50%	27.12%
Aqua America Inc	52898	1315.67	1.43%	28.55%
Humana Inc	48653	1304.17	1.41%	29.97%

Source: CRSP

Notes: The table reports cumulative returns among the 392 CRSP stocks from January 1973 to December 2016. Results pertain to the 10 stocks with the greatest cumulative returns. Cumulative return is computed as the product of total monthly returns throughout the period. The company name displayed is that associated with the PERMNO for the most recent CRSP record.

Table 9: Top Ten Cumulative Return Among the 16,087 CRSP Stocks

Company Name	PERMNO	Growth	% of Total Value	Cumulative %
Altria Group Inc	13901	2655968.8	53.05%	53.05%
Vulcan Materials Co	15202	215689.6	4.31%	57.36%
Boeing Co	19561	150389.9	3.00%	60.37%
International Business Machs Cor	12490	123210.6	2.46%	62.83%
Kansas City Southern	12650	100475.7	2.01%	64.83%
General Dynamics Corp	12052	96926.8	1.94%	66.77%
Walgreens Boots Alliance Inc	19502	71387.9	1.43%	68.20%
Coca Cola Co	11308	70744.31	1.41%	69.61%
Wyeth	15667	54700.0	1.09%	70.70%
Universal Corporation	16555	48921.0	0.98%	71.68%

Source: CRSP

Notes: The table reports cumulative returns among the 16,087 CRSP firms from July 1926 to December 2016. Results pertain to the 10 stocks with the greatest cumulative returns. Cumulative return is computed as the product of total monthly returns throughout the period that is available for each firm. The company name displayed is that associated with the PERMNO for the most recent CRSP record.

Table 10: Inference on Distribution of Wealth Creation using Empirical x_{min} (Simulations with 392 Firms)

	$\hat{x}_{min}$	$\hat{\alpha}$	$\hat{\alpha}$ from Simulated Values				
			Min	5th	50th	95th	Max
Panel A: Method 1 (CAPM)							
Market Cap Growth	65.84	1.925	1.406	1.687	1.998	2.657	5.496
Cumulative Return	176.6	2.353	1.478	1.763	2.133	3.699	24.70
Panel B: Method 2 (Direct Estimation using Expectation of Log Returns)							
Market Cap Growth	65.84	1.925	1.241	1.348	1.495	1.645	1.889
Cumulative Return	176.6	2.353	1.296	1.428	1.592	1.765	1.932

Source: CRSP and simulations

Notes: We conduct 400 monthly simulations of the stock market for the two methods of estimating μ detailed in Appendix B. For each method, we consider the distribution of lifetime wealth creation as measured by market cap growth and cumulative return. Using the steps outlined in Appendix C, we estimate the power law parameter α for each distribution of wealth creation. Instead of estimating the optimal x_{min} for each simulation, we use the $\hat{x}_{min}$ value estimated from the power law parameter estimation on CRSP stocks, reported in the first column. We use these values of x_{min} to estimate the power law parameter α for each distribution of wealth creation in simulations. The second column shows the α estimated for the distribution of wealth creation of the 392 firms from January 1973 to December 2016. The next five columns show the distribution of the statistic obtained from the simulations.

Table 11: Inference on Distribution of Wealth Creation using Empirical x_{min} (Simulations with 16,807 Firms)

	$\hat{x}_{min}$	$\hat{\alpha}$	$\hat{\alpha}$ from Simulated Values				
			Min	5th	50th	95th	Max
Panel A: Method 1 (CAPM)							
Market Cap Growth	255.6	2.005	1.189	1.244	1.368	1.512	1.723
Cumulative Return	3.491	1.586	1.126	1.156	1.220	1.327	1.440
Panel B: Method 2 (Direct Estimation using Expectation of Log Returns)							
Market Cap Growth	255.6	2.005	1.099	1.126	1.159	1.193	1.218
Cumulative Return	3.491	1.586	1.075	1.092	1.114	1.139	1.158

Source: CRSP and simulations

Notes: We conduct 400 monthly simulations of the stock market for the two methods of estimating μ detailed in Appendix B. For each method, we consider the distribution of lifetime wealth creation as measured by market cap growth and cumulative return. Using the steps outlined in Appendix C, we estimate the power law parameter α for each distribution of wealth creation. Instead of estimating the optimal x_{min} for each simulation, we use the $\hat{x}_{min}$ value estimated from the power law parameter estimation on CRSP stocks, reported in the first column. We use these values of x_{min} to estimate the power law parameter α for each distribution of wealth creation in simulations. The second column shows the α estimated for the distribution of wealth creation of the 16,087 firms from July 1926 to December 2016. The next five columns show the distribution of the statistic obtained from the simulations.

Table 12: Inference on Distribution of Wealth Creation using Optimal x_{min} (Simulations with 392 Firms)

	$\hat{\alpha}$	$\hat{\alpha}$ from Simulated Values				
		Min	5th	50th	95th	Max
Panel A: Method 1 (CAPM)						
Market Cap Growth	1.925	1.554	1.698	1.958	2.476	3.075
Cumulative Return	2.353	1.554	1.698	1.958	2.476	3.075
Panel B: Method 2 (Direct Estimation using Expectation of Log Returns)						
Market Cap Growth	1.925	1.463	1.557	1.699	1.959	2.904
Cumulative Return	2.353	1.463	1.557	1.699	1.959	2.904

Source: CRSP and simulations

Notes: We conduct 400 monthly simulations of the stock market for the two methods of estimating μ detailed in Appendix B. For each method, we consider the distribution of lifetime wealth creation as measured by market cap growth and cumulative return. Using the steps outlined in Appendix C, we estimate the power law parameter α for each distribution of wealth creation. Similarly, sampling distributions of $\hat{\alpha}$ are obtained from the simulations. The first column shows the $\hat{\alpha}$ for the distribution of wealth creation of the 392 firms from January 1973 to December 2016. The next five columns show the distribution of the statistic obtained from the simulations For each estimated alpha, the null hypothesis is that it represents a random sample from the sampling distribution of estimated alphas in our simulations..

Table 13: Distribution of Thresholds from α Estimation (Simulations with 392 Firms)

	$\hat{x}_{min}$	$\hat{x}_{min}$ from Simulated Values				
		Min	5th	50th	95th	Max
Panel A: Method 1 (CAPM)						
Market Cap Growth	65.84	0.821	5.259	30.86	239.9	1942.3
Cumulative Return	176.6	0.821	5.259	30.86	239.9	1942.3
Panel B: Method 2 (Direct Estimation using Expectation of Log Returns)						
Market Cap Growth	65.84	25.45	69.55	430.7	3115.2	15385.4
Cumulative Return	176.6	25.45	69.55	430.7	3115.2	15385.4

Source: CRSP and simulations

Notes: We report the estimated threshold values used to estimate the power law parameter α in Table 12. The first column shows the $\hat{x}_{min}$ used in estimating the power law parameter for the distribution of wealth creation of the 392 firms from January 1973 to December 2016. The next three columns show the distribution of $\hat{x}_{min}$ from the simulations.

Table 14: Inference on Distribution of Wealth Creation using Median x_{min} (Simulations with 392 Firms)

	$\hat{\alpha}$	$\hat{x}_{min}$	$\hat{\alpha}$ from Simulated Values				
			Min	5th	50th	95th	Max
Panel A: Method 1 (CAPM)							
Market Cap Growth	1.770	30.86	1.342	1.599	1.895	2.395	3.089
Cumulative Return	1.603	30.86	1.342	1.599	1.895	2.395	3.089
Panel B: Method 2 (Direct Estimation using Expectation of Log Returns)							
Market Cap Growth	2.541	430.68	1.370	1.508	1.668	1.860	2.073
Cumulative Return	2.238	430.68	1.370	1.508	1.668	1.860	2.073

Source: CRSP and simulations

Notes: For the same 400 monthly simulations as in Table 12, we estimate the power law parameter α for each distribution of wealth creation. Instead of estimating the optimal x_{min}, we use the median $\hat{x}_{min}$ value obtained from the power law parameter estimation on simulations, as reported in Table 12. The first column shows the α estimated for the distribution of wealth creation of the 392 firms from January 1973 to December 2016. The next five columns show the distribution of the statistic obtained from the simulations For each estimated alpha, the null hypothesis is that it represents a random sample from the sampling distribution of estimated alphas in our simulations.

Table 15: Inference on Distribution of Wealth Creation using Optimal x_{min} (Simulations with 16,087 Firms)

	$\hat{\alpha}$	$\hat{\alpha}$ from Simulated Values				
		Min	5th	50th	95th	Max
Panel A: Method 1 (CAPM)						
Market Cap Growth	2.005	1.296	1.363	1.448	1.547	1.750
Cumulative Return	1.586	1.296	1.363	1.448	1.547	1.750
Panel B: Method 2 (Direct Estimation using Expectation of Log Returns)						
Market Cap Growth	2.005	1.177	1.198	1.219	1.243	1.261
Cumulative Return	1.586	1.177	1.198	1.219	1.243	1.261

Source: CRSP and simulations

Notes: We conduct 400 monthly simulations of the stock market for the two methods of estimating μ detailed in Appendix B. For each method, we consider the distribution of lifetime wealth creation as measured by market cap growth and cumulative return. Using the steps outlined in Appendix C, we estimate the power law parameter α for each distribution of wealth creation. Similarly, sampling distributions of $\hat{\alpha}$ are obtained from the simulations. The first column shows the $\hat{\alpha}$ for the distribution of wealth creation of the 16,087 firms from July 1926 to December 2016. The next five columns show the distribution of the statistic obtained from the simulations For each estimated alpha, the null hypothesis is that it represents a random sample from the sampling distribution of estimated alphas in our simulations.

Table 16: Distribution of Thresholds from α Estimation (Simulations with 16,087 Firms)

	$\hat{x}_{min}$	$\hat{x}_{min}$ from Simulated Values				
		Min	5th	50th	95th	Max
Panel A: Method 1 (CAPM)						
Market Cap Growth	255.6	13.47	129.4	2387.0	80,721.7	862,975.2
Cumulative Return	3.491	13.47	129.4	2387.0	80,721.7	862,975.2
Panel B: Method 2 (Direct Estimation using Expectation of Log Returns)						
Market Cap Growth	255.6	446.67	3,880.7	70,780.6	8,309,391	3,562,204,252
Cumulative Return	3.491	446.67	3,880.7	70,780.6	8,309,391	3,562,204,252

Source: CRSP and simulations

Notes: We report the estimated threshold values used to estimate the power law parameter α in Table 15. The first column shows the $\hat{x}_{min}$ used in estimating the power law parameter for the distribution of wealth creation of the 16,087 firms from July 1926 to December 2016. The next three columns show the distribution of $\hat{x}_{min}$ from the simulations.

Table 17: Inference on Distribution of Wealth Creation using Median x_{min} (Simulations with 16,087 Firms)

	$\hat{\alpha}$	$\hat{x}_{min}$	$\hat{\alpha}$ from Simulated Values				
			Min	5th	50th	95th	Max
Panel A: Method 1 (CAPM)							
Market Cap Growth	2.078	2,387.0	1.232	1.304	1.433	1.560	1.767
Cumulative Return	1.729	2,387.0	1.232	1.304	1.433	1.560	1.767
Panel B: Method 2 (Direct Estimation using Expectation of Log Returns)							
Market Cap Growth	N/A	70,780.6	1.156	1.187	1.216	1.239	1.263
Cumulative Return	2.042	70,780.6	1.156	1.187	1.216	1.239	1.263

Source: CRSP and simulations

Notes: For the same 400 monthly simulations as in Table 12, we estimate the power law parameter α for each distribution of wealth creation. Instead of estimating the optimal x_{min}, we use the median $\hat{x}_{min}$ value obtained from the power law parameter estimation on simulations, as reported in Table 12. The first column shows the α estimated for the distribution of wealth creation of the 392 firms from January 1973 to December 2016. The next five columns show the distribution of the statistic obtained from the simulations For each estimated alpha, the null hypothesis is that it represents a random sample from the sampling distribution of estimated alphas in our simulations.

Table 18: Inference on Distribution of Market Cap (Simulations with 16,087 Firms)

	Empirical Value	Simulated Values				
		Min	5th	50th	95th	Max
Panel A: Method 1 (CAPM)						
Market Cap	2.289	1.296	1.363	1.448	1.547	1.750
Panel B: Method 2 (Direct Estimation using Expectation of Log Returns)						
Market Cap	2.289	1.177	1.198	1.219	1.243	1.261

Source: CRSP and simulations

Notes: We conduct 400 monthly simulations of the stock market for both methods of estimating μ detailed in Appendix B. For each method, we examine the distribution of firm sizes as measured by market cap and estimate the power law parameter α of the distribution. Sampling distributions of $\hat{\alpha}$ are obtained from the simulations. The first column shows the parameter estimated for the distribution of market caps for all 2,395 firms in January 1972. The next five columns illustrate the distribution of $\hat{\alpha}$ obtained from the simulations.

Table 19: Inference on Wealth Contribution of Top Ten Stocks (Simulations with 392 Firms)

	Empirical Value	Simulated Values				
		Min	5th	50th	95th	Max
Panel A: Method 1 (CAPM)						
Market Cap Growth	41.33	27.16	34.35	49.64	74.88	92.97
Cumulative Return	29.97	27.16	34.35	49.64	74.88	92.97
Panel B: Method 2 (Direct Estimation using Expectation of Log Returns)						
Market Cap Growth	41.33	39.06	52.44	74.06	97.37	99.96
Cumulative Return	29.97	39.06	52.44	74.06	97.37	99.96

Source: CRSP and simulations

Notes: We conduct 400 monthly simulations of the stock market for both methods of estimating μ detailed in Appendix B. For each method, we examine the percentage of wealth contributed by the top ten stocks with wealth creation as measured by market cap and cumulative return. Sampling distributions of the percentages are obtained from the simulations. A greater percentage of wealth contributed by the ten stocks implies a greater asymmetry exhibited in the distribution of wealth creation.

Table 20: Inference on Wealth Contribution of Top Ten Stocks (Simulations with 16,087 Firms)

	Empirical Value	Simulated Values				
		Min	5th	50th	95th	Max
Panel A: Method 1 (CAPM)						
Market Cap Growth	29.40	71.48	88.99	99.63	99.99	99.99
Cumulative Return	71.67	71.48	88.99	99.63	99.99	99.99
Panel B: Method 2 (Direct Estimation using Expectation of Log Returns)						
Market Cap Growth	29.40	98.69	99.92	99.99	99.99	99.99
Cumulative Return	71.67	98.69	99.92	99.99	99.99	99.99

Source: CRSP and simulations

Notes: We conduct 400 monthly simulations of the stock market for both methods of estimating μ detailed in Appendix B. For each method, we examine the percentage of wealth contributed by the top ten stocks with wealth creation as measured by market cap and cumulative return. Sampling distributions of the percentages are obtained from the simulations. A greater percentage of wealth contributed by the ten stocks implies a greater asymmetry exhibited in the distribution of wealth creation.

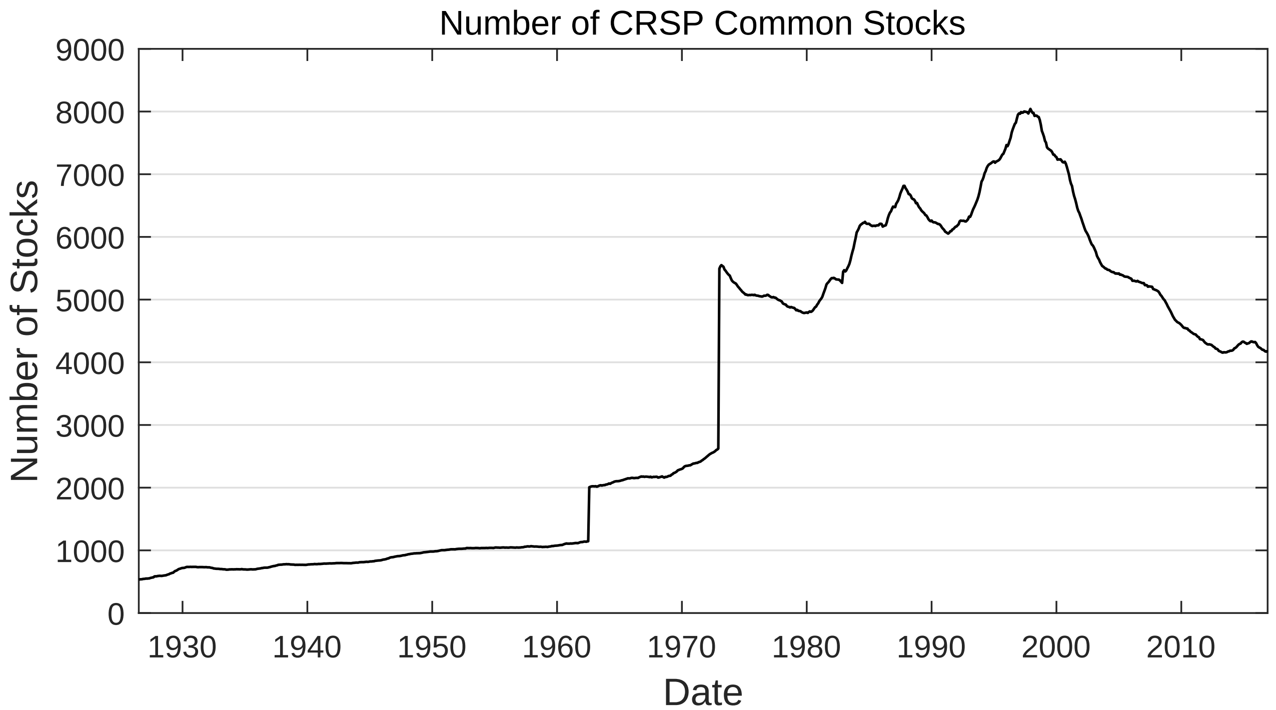

Figure 1: Historical Number of CRSP Common Stocks
On the first day of each month from July 1926 to December 2016, we count the number of unique common stocks in the cross-section, as available in CRSP. The jump between December 1972 and January 1973, from 2,623 to 5,494, corresponds to the establishment of Nasdaq.

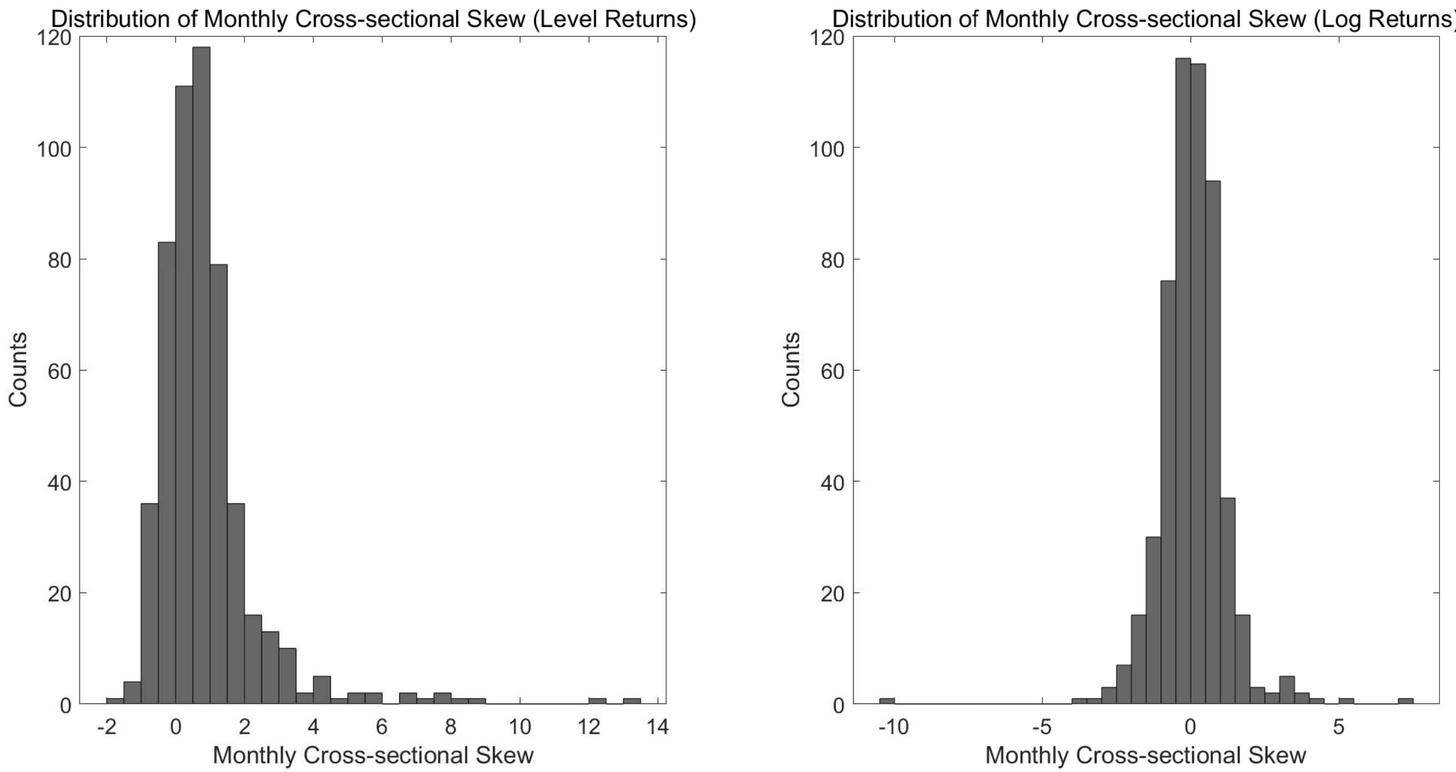

Figure 2: Distribution of Monthly Cross-sectional Skew for the 392 Firms
The figures illustrate the distribution of monthly cross-sectional skewness, defined as the skewness of monthly returns for all 392 firms in any given month.

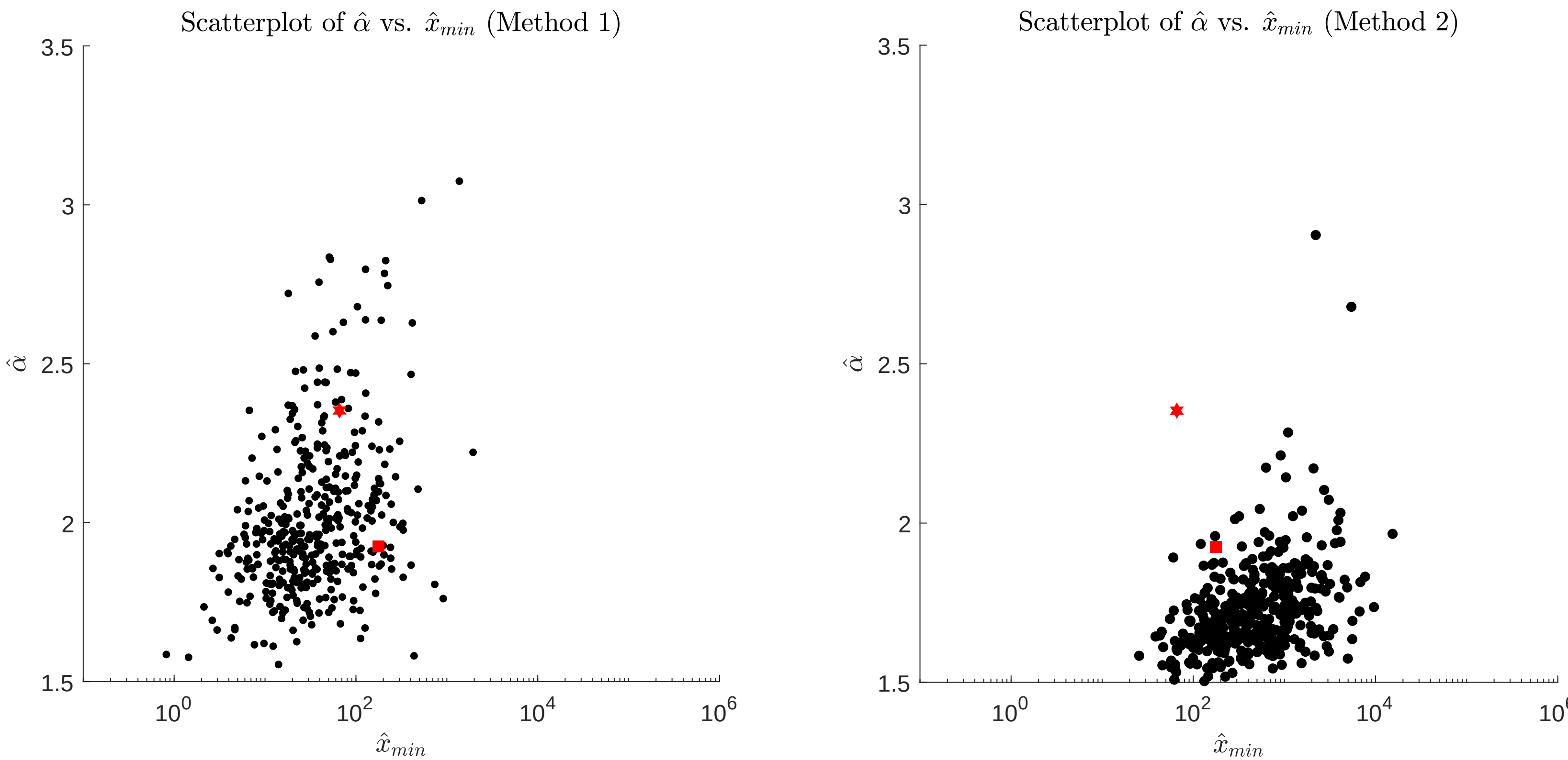

Figure 3: Scatterplot of $\hat{\alpha}$ and $\hat{x}_{min}$ from Empirical Data and Simulations with 392 Firms
The figures illustrate the scatterplot of $\hat{\alpha}$ vs. $\hat{x}_{min}$ computed from the simulations.

Appendix

A Geometric Brownian Motion

Deriving implications of geometric brownian motion starts with the stochastic differential equation:

$$\frac{dS(t)}{S(t)} = \mu dt + \sigma dW(t)$$

where $S(t)$ is the price of a security at time t, μ the drift parameter, σ the volatility parameter, and $W(t)$ the value of a Wiener process at time t. Applying Ito's lemma to $dlnS(t)$:

$$\begin{aligned}
d\log S(t) &= \frac{1}{S(t)}dS(t) - \frac{1}{2}\frac{1}{S(t)^2}dS(t)^2 \\
&= \frac{1}{S(t)}S(t)[\mu dt + \sigma dW(t)] - \frac{1}{2}\frac{1}{S(t)^2}S(t)^2[\sigma^2 dW(t)^2]^2 \\
&= \mu dt + \sigma dW(t) - \frac{1}{2}\sigma^2 dt
\end{aligned}$$

Integrating each side,

$$\log R = \log S(t) - \log S(0) = \left(\mu - \frac{1}{2}\sigma^2\right)t + \sigma W(t)$$

Therefore, we arrive at the normal distribution of log returns:

$$\log R \sim N\left(\left(\mu - \frac{1}{2}\sigma^2\right)t, \sigma^2 t\right)$$

We can also derive an expression for the expected total return, $E[R]$:

$$E[R] = E\left[\exp(\log R)\right] = \exp\left(\left(\mu - \frac{1}{2}\sigma^2 + \frac{1}{2}\sigma^2\right)t\right) = \exp(\mu t)$$

B Estimation of Drift Parameter (μ)

Method 1. CAPM

We assume that the Capital Asset Pricing Model (CAPM) holds for log returns:

$$\log\left[\frac{E[R]}{R_f}\right] = \beta\left[\frac{E[R_M]}{R_f}\right]$$

where R is the total return on a security, R_f the total risk-free return, R_M the total return on the market, and β the market beta of the security. Defining r as $\log R$ and following the notations regarding geometric brownian motion presented in Appendix A, we have:

$$\mu - r_f = \beta\left(\log E[R_M] - r_f\right)$$
$$\mu = (1-\beta)r_f + \beta \log E[R_M]$$

Since the risk-free rate is not constant in our data, we use $E\left[\log R_f\right]$ instead of r_f. Therefore, we arrive at the following expression for μ:

$$\mu = (1-\beta)E\left[\log R_f\right] + \beta \log E[R_M]$$

Method 2. Direct Estimation via Expectation of Log Returns

Recall the expression for the expectation of log returns, $E[\log R]$:

$$\log R = \left(\mu - \frac{1}{2}\sigma^2\right)t + \sigma W(t) \Rightarrow E[\log R] = \left(\mu - \frac{1}{2}\sigma^2\right)t$$

Setting $t = 1$ using daily parameters, we arrive at the following expression for μ:

$$\mu = E\left[\log R\right] + \frac{1}{2}\sigma^2$$

C Aggregate Wealth Creation Metric

Bessembinder (2017) seeks to capture the experience of investors in aggregate and creates a measure of dollar wealth creation for each firm. In this section, we outline the derivation of his metric.

Let W_0 denote the initial wealth of the investor with an investment horizon of T periods. In each period, the investor chooses between a riskless bond with return r_f and a risky investment $\tilde{r} = \tilde{r}_c + \tilde{r}_d$ where $\tilde{r}_c$ is the capital gain and $\tilde{r}_d$ is the dividend yield.

We assume that dividends are returned to the investeor's bond account. We also assume that at time t, the investor take F_t from his bond account and invests it in the risky asset. The wealth in investor's bond account at time t, B_t, evolves according to the following equation:

$$B_t = B_{t-1}(1 + r_f) + I_{t-1}\tilde{r}_d - F_t$$

and the wealth in investor's stock (risky asset) account at time t, I_t, evolves as the following:

$$I_t = I_{t-1}(1 + \tilde{r}_c) + F_t$$

Investor's total wealth can be expressed as $W_t = B_t + I_t$. Therefore:

$$W_t = B_{t-1}(1 + r_f) + I_{t-1}\tilde{r}_d + I_{t-1}(1 + \tilde{r}_c)$$

$$W_t - W_{t-1}(1 + r_f) = I_{t-1}(\tilde{r}_t - r_f)$$

Applying the above equation iteratively and using realized returns, we have:

$$W_t - W_0(1 + r_f)^T = \prod_{t=1}^{T} I_{t-1}(r_t - r_f)(1 + r_f)^{T-t}$$

D Estimation of Power Law Parameter

The power law distribution has the following probability function defined for $x \geq x_{min}$:

$$p(x) = Cx^{-\alpha}$$

where C is a constant and α the power law parameter. It is possible to derive an expression for C through normalization, only when $\alpha > 1$:

$$1 = \int_{x_{min}}^{\infty} p(x)dx = C \int_{x_{min}}^{\infty} \frac{dx}{x^{\alpha}} = \frac{C}{\alpha - 1} x_{min}^{-\alpha+1}$$

$$C = (\alpha - 1) x_{min}^{\alpha-1}$$

Substituting into the original equation, we therefore have the following expression, defined only for $\alpha > 1$ and $x \geq x_{min}$:

$$p(x) = \frac{\alpha - 1}{x_{min}} \left(\frac{x}{x_{min}} \right)^{-\alpha}$$

Estimating α requires the choice of x_{min}. We use the minimization of Kolmogorov-Smirnov statistic D, defined as

$$D = \max_{x} |F(x|\alpha, x_{min}) - F(x)|$$

where $F(x|\alpha, x_{min})$ denotes the cdf of the power law distribution and $F(x)$ the cdf of the data. Once the optimal x_{min} is determined, we estimateα using the standard maximum likelihood approach following Newman (2004). The estimate of α is therefore given as:

$$\hat{\alpha} = 1 + \left[\sum_{i=1}^{n} \log \left(\frac{x_i}{x_{min}} \right) \right]^{-1}$$