2017

Assessing Models using Monte Carlo Simulations

Simon Oh
University of Pennsylvania

Follow this and additional works at: http://repository.upenn.edu/spur

Part of the Finance and Financial Management Commons

Recommended Citation


This paper is posted at ScholarlyCommons. http://repository.upenn.edu/spur/15
For more information, please contact repository@pobox.upenn.edu.
Assessing Models using Monte Carlo Simulations

Abstract
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 show that geometric brownian motion alone cannot generate a majority of the patterns in the distribution of stock returns and wealth creation. Our paper represents an often overlooked departure from the traditional way of validating asset pricing models, in which implications are derived, parameters calibrated, and magnitudes compared to empirical data. Instead, we seek to leverage the power of large numbers by conducting numerous simulations and assessing the probability that they contain our realized stock market.

Keywords
Monte Carlo, Brownian Motion, Skewness, Wealth Creation

Disciplines
Business | Finance and Financial Management

This working paper is available at ScholarlyCommons: http://repository.upenn.edu/spur/15
Assessing Models using Monte Carlo Simulations

Sangmin Oh*       Jessica Wachter†

September 18, 2017

Abstract

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 show that geometric brownian motion alone cannot generate a majority of the patterns in the distribution of stock returns and wealth creation. Our paper represents an often overlooked departure from the traditional way of validating asset pricing models, in which implications are derived, parameters calibrated, and magnitudes compared to empirical data. Instead, we seek to leverage the power of large numbers by conducting numerous simulations and assessing the probability that they contain 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 illustrate that geometric brownian motion alone cannot generate a majority of the patterns in the distribution of stock returns and wealth creation. Our paper represents a fresh departure from the traditional way of validating asset pricing models, in which implications are derived, parameters calibrated, and magnitudes compared to empirical data. Instead, we seek to leverage the power of large numbers by conducting multiple simulations and assessing the probability that they contain our realized stock market.

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. We achieve this through hypotheses 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. Furthermore, it is also 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. Specifically, we restrict our attention to the 431 stocks that have been
in CRSP database throughout the sample period of January 1970 to December 2000. We show that simulations assuming geometric brownian motion fail to generate most of the statistics examined in this paper — in fact, they fail spectacularly when it comes to the cross-sectional skew of monthly returns. The model does succeed, however, in generating a fat-tailed distribution of individual firm’s market cap growth during our sample period. This outcome seems to suggest that the asymmetric distribution of firm size does not necessitate a similar asymmetric distribution of returns.

Our simulations also yield several interesting observations. For example, the skewness of pooled monthly returns in CRSP is far greater than what most simulations generate, whereas the standard deviation is far smaller than the values from the simulations. Given that both are measures of cross-sectional dispersion, the contrast is surprising. Furthermore, the concentration of stock market wealth creation in a few stocks is much more severe in our simulations than the data in CRSP. Given that the skewness of CRSP returns is much greater than what simulations generate, this fact is also counter-intuitive. Together, they seem to suggest there is another force at work other than the distribution of returns that gives rise to the asymmetric distribution of firm sizes and wealth creation.

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.
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\(^1\). This failure has led to many modifications to the features of the original model, including alternate assumptions on preferences\(^2\), rare disasters\(^3\), liquidity risk\(^4\) and market imperfections\(^5\). 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\(^6\) however, has led to a set of new unconditional multi-factor models\(^7\). 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

---

\(^1\)See Mehra and Prescott (1985)
\(^2\)See Abel (1990), Bansal and Yaron (2004), Campbell and Cochrane (1999), Epstein and Zin (1991)
\(^3\)See Barro (2006), Gabaix (2012), and Wachter (2013)
\(^4\)See Alvarez and Jermann (2001)
\(^6\)See Fama and French (2004)
captures the distribution of individual stock returns around the market return. Recent literature has been paying attention to its role in forecasting market returns\textsuperscript{8} implications for asset managers\textsuperscript{9} and pricing of the cross-section of stock returns\textsuperscript{10}.

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\textsuperscript{11} 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\textsuperscript{12}.

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\textsuperscript{13} Recently, Gabaix (2016) has effectively used the observed skewness in firm sizes to examine how standard economic theories fit with the empirical data.

\textsuperscript{8}See Garcia et al. (2014), Goyal and Santa-Clara (2003)
\textsuperscript{9}See Gorman, Sapra, and Weigand (2010)
\textsuperscript{10}See Verousis and Voukelatos (2015)
\textsuperscript{11}See Harvey and Siddique (2000)
\textsuperscript{13}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 at daily frequency from January 1970 to December 2000 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.

For our simulation, 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.

### 3.2 Parameter Estimation

Geometric brownian motion starts with the following stochastic differential equation:

\[
\frac{dS(t)}{S(t)} = \mu dt + \sigma dW(t)
\]
where $S(t)$ is the price of a security at time $t$, $\mu$ the drift parameter, $\sigma$ the volatility parameter, and $W(t)$ the value of a Wiener process at time $t$. Its major implication is that log returns are normally distributed:

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

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

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

The second and third methods are based on direct estimates from all daily returns in our sample period January 1970 to December 2000. They are obtained from equations (4) and (5) respectively:

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

$$\mu = \ln[E[R]]$$

Figure 1 illustrates the distribution of $\mu$ for the 431 firms in our universe.

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

$$\hat{\sigma} = \sqrt{\frac{1}{n-1}\sum_{t=1}^{n}(r_t - \bar{r})^2}$$

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

$$\ln R_t = \alpha + \beta \ln R_{M,t} + \epsilon_t$$  \hspace{1cm} (7)

Figure 3 illustrates the distribution of $\sigma$ and $\beta$ for the 431 firms in our universe.

### 3.3 Simulation Overview

In implementing the simulation of stock prices, we wish to differentiate a market-wide shock from an idiosyncratic shock to each firm. Having estimated the $\sigma_i$ parameter for firm $i$, we can decompose it into a systematic component and an idiosyncratic component:

$$\sigma_i = \beta_i \sigma_M + \sigma_\epsilon$$  \hspace{1cm} (8)

where $\sigma_M$ denotes the market return volatility and $\sigma_\epsilon$ denotes idiosyncratic volatility. Therefore, the stock price at time $t$ following geometric brownian motion can be expressed as:

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

where $\epsilon_M$ and $\epsilon_i$ represent the market shock and idiosyncratic shock for firm $i$ respectively. For each day in the simulation, we draw 432 times from $N(0,1)$ — one for the market-wide shock $\epsilon_M$ and 431 for idiosyncratic shock $\epsilon_i$ corresponding to each firm in our universe. Using the realizations of the random variable, we update the daily stock price according to equation (9).

Each simulation consists of daily stock price simulations for all 431 firms between January 1970 to December 2000. We conduct the simulation 300 times.
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
for the 431 common stocks in CRSP from January 1970 to December 2000. The steps
we take are the following:

1. Choose a statistic $\zeta$. 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 $\zeta_{data}$ from the CRSP data on 431 stocks in our universe.

3. For each of the 300 stock market simulations, compute $\zeta_{sim}^{i}$ corresponding to the
$i$th simulation. Consequently, we obtain a sampling distribution of $\zeta_{sim}^{i}$ consisting
of 300 independent observations.

4. We test the null hypothesis that $\zeta_{data}$ represents a random sample from the distri-
bution of $\zeta_{sim}^{i}$. 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.

Unlike 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 all 431 stocks from January 1970 to
December 2000. We examine the mean, standard deviation, skewness, and percentage
of returns that are positive. We also study the percentage of returns that are greater
than the value-weighted and equal-weighted market returns.\(^{14}\)

\(^{14}\)Because our universe consists of 431 stocks, the market return is also computed as the weighted
sum of the 431 individual returns, not the entire CRSP universe.
The second category of statistics is based on monthly cross-sectional skewness, defined as the skewness of monthly returns for the 431 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 $\alpha$ of the power law distribution fitted to the distribution of individual firm’s wealth creation. The estimate of $\alpha$, 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 z-scores and p-values corresponding to each statistic computed for the 431 firms. Simulations in Panel A are conducted with the $\mu$ 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. This result is not a surprise - expected returns are a sole function of $\mu$ which is directly estimated from our sample data.

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 out-
perform 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.

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
succeed is the mean and percentage of positive returns, a result that seems trivial.

4.2 Time-series of Monthly Cross-sectional Skewness

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

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

(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:

$$\frac{1}{T} \sum_{t=1}^{T} \gamma_{cs}^{t}$$

(11)

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

(12)

Table 3 reports selected statistics for the pooled CRSP common stock returns for dif-
ferent 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\textsuperscript{15}, increased correlations during crises\textsuperscript{16} and positive skewness in sector-specific return shocks\textsuperscript{17}.

Table\textsuperscript{4} reports the z-scores and p-values corresponding to the mean and standard deviation of monthly cross-sectional skew computed for the 431 firms. 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. In short, geometric brownian motion alone cannot generate the characteristics of monthly cross-sectional skewness observed in the actual stock market.

\textsuperscript{15}Stock prices cannot go below zero.
\textsuperscript{16}See Campbell, Koedijk, and Kofman (2002)
\textsuperscript{17}See Duffee (2001)
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 $\Delta MC_t$ from time 0 to time $t$ is defined as

$$\Delta MC_t = \frac{p_t N_t}{p_0 N_0}$$  \hspace{1cm} (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

$$CR_t = \prod_{i=0}^{t} (1 + r_i) - 1$$  \hspace{1cm} (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:

$$AWC_t = \prod_{t=1}^{T} \left[ I_{t-1}(r_t - r_f)(1 + r_f)^{T-t} \right]$$  \hspace{1cm} (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 5, 6, and 7 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). In number, the ten stocks represent 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. The details of the estimation
can be found in Appendix D.

Table 8 reports the z-scores and p-values corresponding to the estimated $\alpha$ for each
wealth creation metric. For market cap growth, the null hypothesis is not rejected at
the 0.01 significance level in all three panels, indicating that the asymmetric distribu-
tion 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 $\alpha$. Smaller $\alpha$ 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.

Table 9 reports the z-scores and p-values corresponding to the percentage con-
tributed 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.

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. The sign of the z-score implies that the simulations overestimate the contribu-
tion 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 brown-
nian motion, the simulations seem to produce scenarios in which the concentration of

\footnote{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.}

\footnote{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.}
wealth creation is much more severe than what the actual stock market exhibits.

In sum, we have mixed results in trying to reject the null hypotheses using both the power law coefficient estimates and the percentage contribution of top ten stocks to total wealth. Notably, we cannot reject the null hypothesis when the market cap growth is used as a measure of wealth creation. The assumption of geometric brownian motion seems to be sufficient in generating the asymmetric distribution of market capitalization observed in the U.S. stock market.

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 cross-sectional skewness of monthly stock returns. But it does seem to be sufficient in generating the 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.

Several puzzles arise from the results of our simulations. First, the skewness of pooled monthly returns in CRSP is far greater than what most simulations generate, whereas the standard deviation is far smaller than the values from simulations. Given that both are measures of cross-sectional dispersion, this deep contrast is surprising. Second, geometric brownian motion alone has generated outcomes in which the concentration of wealth creation is much more severe than what the actual stock market exhibits. Given that the skewness of the CRSP returns is much greater than what simulations predict, this observation seems to suggest that there is another force at
work other than the distribution of returns that drives the asymmetric distribution of value creation.

Our simulations include some simplifications 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. 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, there may be contentions regarding the estimation of the simulation parameters. Given our focus on 431 firms that have been in existence throughout our sample period, the estimates of the drift and volatility parameters are subject to a survivorship bias. This fact, however, does not directly jeopardize the validity of our inferences as we are comparing the simulated results to the actual returns distribution of only 431 firms, not the entire CRSP universe. Nonetheless, it would be interesting to include in our estimation the stocks that are not necessarily extant throughout the full sample.

Finally, immediate extensions to the simple geometric brownian motion can be considered. One possible modification is dropping the assumption that the drift or volatility is constant. 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 sampling distributions, can be readily applied in both cases.

---

20See Lowry, Officer, and Schwert (2010)
References


Table 1: Selected Statistics on Pooled Monthly Returns

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>0.011</td>
<td>0.012</td>
<td>0.013</td>
</tr>
<tr>
<td>Median</td>
<td>0.000</td>
<td>0.0000</td>
<td>0.008</td>
</tr>
<tr>
<td>Standard Deviation</td>
<td>0.170</td>
<td>0.185</td>
<td>0.100</td>
</tr>
<tr>
<td>Skewness</td>
<td>6.418</td>
<td>6.608</td>
<td>0.828</td>
</tr>
<tr>
<td>% Positive</td>
<td>49.3%</td>
<td>46.5%</td>
<td>53.1%</td>
</tr>
<tr>
<td>% ≥ VW Mkt Return</td>
<td>46.4%</td>
<td>45.4%</td>
<td>48.5%</td>
</tr>
<tr>
<td>% ≥ EQ Mkt Return</td>
<td>46.0%</td>
<td>45.4%</td>
<td>48.6%</td>
</tr>
</tbody>
</table>

Source: CRSP

Notes: The table reports selected statistics on pooled CRSP common stock monthly returns for different time horizons and different universe of stocks.
### Table 2: Inference on Pooled Monthly Returns with Sampling Distribution

<table>
<thead>
<tr>
<th>Panel A: Method 1</th>
<th>Value</th>
<th>z-score</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>0.013</td>
<td>2.00</td>
<td>0.0227</td>
</tr>
<tr>
<td>Standard Deviation</td>
<td>0.100</td>
<td>-8.13</td>
<td>0.0000</td>
</tr>
<tr>
<td>Skewness</td>
<td>0.828</td>
<td>9.10</td>
<td>0.0000</td>
</tr>
<tr>
<td>% Positive</td>
<td>53.1%</td>
<td>1.43</td>
<td>0.0757</td>
</tr>
<tr>
<td>% ≥ VW Mkt Return</td>
<td>48.5%</td>
<td>11.3</td>
<td>0.0000</td>
</tr>
<tr>
<td>% ≥ EQ Mkt Return</td>
<td>48.6%</td>
<td>4.59</td>
<td>0.0000</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Panel B: Method 2</th>
<th>Value</th>
<th>z-score</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>0.013</td>
<td>-0.15</td>
<td>0.4388</td>
</tr>
<tr>
<td>Standard Deviation</td>
<td>0.100</td>
<td>-9.39</td>
<td>0.0000</td>
</tr>
<tr>
<td>Skewness</td>
<td>0.828</td>
<td>6.99</td>
<td>0.0000</td>
</tr>
<tr>
<td>% Positive</td>
<td>53.1%</td>
<td>-0.70</td>
<td>0.2400</td>
</tr>
<tr>
<td>% ≥ VW Mkt Return</td>
<td>48.5%</td>
<td>7.12</td>
<td>0.0000</td>
</tr>
<tr>
<td>% ≥ EQ Mkt Return</td>
<td>48.6%</td>
<td>5.71</td>
<td>0.0000</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Panel C: Method 3</th>
<th>Value</th>
<th>z-score</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>0.013</td>
<td>-0.10</td>
<td>0.4605</td>
</tr>
<tr>
<td>Standard Deviation</td>
<td>0.100</td>
<td>-9.05</td>
<td>0.0000</td>
</tr>
<tr>
<td>Skewness</td>
<td>0.828</td>
<td>6.43</td>
<td>0.0000</td>
</tr>
<tr>
<td>% Positive</td>
<td>53.1%</td>
<td>-0.64</td>
<td>0.2606</td>
</tr>
<tr>
<td>% ≥ VW Mkt Return</td>
<td>48.5%</td>
<td>8.52</td>
<td>0.0000</td>
</tr>
<tr>
<td>% ≥ EQ Mkt Return</td>
<td>48.6%</td>
<td>5.63</td>
<td>0.0000</td>
</tr>
</tbody>
</table>

**Source**: CRSP and simulations

**Notes**: We conduct 300 simulations of the stock market for each of three methods of estimating $\mu$ detailed in Appendix B. Sampling distributions of each statistic are obtained from the simulations. The first column shows the statistic for the 431 firms from January 1970 to December 2000. For each statistic, the null hypothesis is that the obtained statistic represents a random sample from the sampling distribution. To test the hypothesis, we compute the z-scores and the corresponding p-values.
**Table 3: Selected Statistics on Monthly Cross-sectional Skewness**

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Mean</strong></td>
<td>1.490</td>
<td>3.293</td>
<td>0.774</td>
</tr>
<tr>
<td><strong>Standard Deviation</strong></td>
<td>2.012</td>
<td>4.058</td>
<td>1.041</td>
</tr>
</tbody>
</table>

*Source:* CRSP

*Notes:* The table reports selected statistics on monthly cross-sectional skewness, defined as the skewness of monthly returns for all 431 firms in any given month. Computing the skewness for each month yields a time-series of monthly skewness for the period. We report the mean and standard deviation of monthly cross-sectional skewness for different time horizons and different universe of stocks.
Table 4: Inference on Monthly Cross-sectional Skewness with Sampling Distribution

<table>
<thead>
<tr>
<th>Panel A: Method 1</th>
<th>Value</th>
<th>z-score</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>0.774</td>
<td>9.96</td>
<td>0.0000</td>
</tr>
<tr>
<td>Standard Deviation</td>
<td>1.041</td>
<td>12.2</td>
<td>0.0000</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Panel B: Method 2</th>
<th>Value</th>
<th>z-score</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>0.774</td>
<td>8.16</td>
<td>0.0000</td>
</tr>
<tr>
<td>Standard Deviation</td>
<td>1.041</td>
<td>11.4</td>
<td>0.0000</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Panel C: Method 3</th>
<th>Value</th>
<th>z-score</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>0.774</td>
<td>7.57</td>
<td>0.0000</td>
</tr>
<tr>
<td>Standard Deviation</td>
<td>1.041</td>
<td>10.6</td>
<td>0.0000</td>
</tr>
</tbody>
</table>

Source: CRSP and simulations

Notes: We conduct 300 simulations of the stock market for each of three methods of estimating $\mu$ detailed in Appendix B. Sampling distributions of each statistic are obtained from the simulations. The first column shows the mean and standard deviation of monthly cross-sectional skewness for the 431 firms from January 1970 to December 2000. For each statistic, the null hypothesis is that the obtained statistic represents a random sample from the sampling distribution. To test the hypothesis, we compute the z-scores and the corresponding p-values.
### Table 5: Wealth Creation: Market Cap Growth

<table>
<thead>
<tr>
<th>Company Name</th>
<th>Value Created</th>
<th>% of Total Value</th>
<th>Cumulative %</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tyco Labs</td>
<td>2239.07</td>
<td>15.58%</td>
<td>15.58%</td>
</tr>
<tr>
<td>Scientific Atlanta</td>
<td>843.63</td>
<td>5.87%</td>
<td>21.45%</td>
</tr>
<tr>
<td>Unilever</td>
<td>792.92</td>
<td>5.52%</td>
<td>26.97%</td>
</tr>
<tr>
<td>Industrial National</td>
<td>327.71</td>
<td>2.28%</td>
<td>29.25%</td>
</tr>
<tr>
<td>Alpha Industries</td>
<td>286.82</td>
<td>2.00%</td>
<td>31.25%</td>
</tr>
<tr>
<td>Northwest Bancorp</td>
<td>280.29</td>
<td>1.95%</td>
<td>33.20%</td>
</tr>
<tr>
<td>Walgreens Boots</td>
<td>263.30</td>
<td>1.83%</td>
<td>35.03%</td>
</tr>
<tr>
<td>Time Warner</td>
<td>259.38</td>
<td>1.81%</td>
<td>36.84%</td>
</tr>
<tr>
<td>Automatic Data</td>
<td>192.73</td>
<td>1.34%</td>
<td>38.18%</td>
</tr>
<tr>
<td>Mid Continent Tel</td>
<td>188.86</td>
<td>1.31%</td>
<td>39.49%</td>
</tr>
</tbody>
</table>

**Source:** CRSP

**Notes:** The table reports market cap growth of each firm from January 1970 to December 2000. Results pertain to the 10 stocks with the greatest market cap growth among the 431 firms. Market cap is computed as the closing price of each day multiplied by the number of shares outstanding, as available in CRSP.
Table 6: Wealth Creation: Cumulative Return

<table>
<thead>
<tr>
<th>Company Name</th>
<th>Value Created</th>
<th>% of Total Value</th>
<th>Cumulative %</th>
</tr>
</thead>
<tbody>
<tr>
<td>Kansas City South.</td>
<td>861.4</td>
<td>4.69%</td>
<td>4.69%</td>
</tr>
<tr>
<td>Walgreens Boots</td>
<td>531.4</td>
<td>2.90%</td>
<td>7.59%</td>
</tr>
<tr>
<td>Altria Group Inc.</td>
<td>361.9</td>
<td>1.97%</td>
<td>9.56%</td>
</tr>
<tr>
<td>UST Inc.</td>
<td>302.6</td>
<td>1.65%</td>
<td>11.21%</td>
</tr>
<tr>
<td>Enron Corp</td>
<td>280.0</td>
<td>1.53%</td>
<td>12.73%</td>
</tr>
<tr>
<td>Scientific Atlanta</td>
<td>273.4</td>
<td>1.49%</td>
<td>14.22%</td>
</tr>
<tr>
<td>Kroger Company</td>
<td>261.1</td>
<td>1.42%</td>
<td>15.65%</td>
</tr>
<tr>
<td>Tyco Labs</td>
<td>257.1</td>
<td>1.40%</td>
<td>17.05%</td>
</tr>
<tr>
<td>Cincinnati Bell</td>
<td>229.4</td>
<td>1.25%</td>
<td>18.30%</td>
</tr>
<tr>
<td>Wrigley William Jr.</td>
<td>215.8</td>
<td>1.18%</td>
<td>19.47%</td>
</tr>
</tbody>
</table>

Source: CRSP

Notes: The table reports cumulative return of each firm from January 1970 to December 2000. Results pertain to the 10 stocks with the greatest cumulative return among the 431 firms. Cumulative return is computed by linking daily gross returns, thereby assuming the reinvestment of dividends but no additional purchase of the shares.
Table 7: Wealth Creation: Metric from Bessembinder (2017)

<table>
<thead>
<tr>
<th>Company Name</th>
<th>Value Created (Mil)</th>
<th>% of Total Value</th>
<th>Cumulative %</th>
</tr>
</thead>
<tbody>
<tr>
<td>General Electric</td>
<td>511,434.4</td>
<td>10.2%</td>
<td>10.2%</td>
</tr>
<tr>
<td>Standard Oil</td>
<td>351,510.2</td>
<td>7.0%</td>
<td>17.1%</td>
</tr>
<tr>
<td>Merck &amp; Co</td>
<td>235,356.6</td>
<td>4.7%</td>
<td>21.8%</td>
</tr>
<tr>
<td>Pfizer Chas &amp; Co</td>
<td>171,155.6</td>
<td>3.4%</td>
<td>25.2%</td>
</tr>
<tr>
<td>Philip Morris &amp; Co</td>
<td>166,290.2</td>
<td>3.3%</td>
<td>28.5%</td>
</tr>
<tr>
<td>Coca Cola</td>
<td>160,707.6</td>
<td>3.2%</td>
<td>31.7%</td>
</tr>
<tr>
<td>Johnson &amp; Johnson</td>
<td>141,756.4</td>
<td>2.8%</td>
<td>34.5%</td>
</tr>
<tr>
<td>P&amp;G</td>
<td>139,123.7</td>
<td>2.8%</td>
<td>37.3%</td>
</tr>
<tr>
<td>Ford Motor</td>
<td>108,007.3</td>
<td>2.1%</td>
<td>39.4%</td>
</tr>
<tr>
<td>Hewlett Packard</td>
<td>939,166.7</td>
<td>1.9%</td>
<td>41.3%</td>
</tr>
</tbody>
</table>

Source: CRSP

Notes: The table reports wealth creation of each firm from January 1970 to December 2000 as measured by equation (15). Results pertain to the 10 stocks with the greatest wealth creation among the 431 firms from January 1970 to December 2000.
Table 8: Inference on Power Law Distribution Parameter with Sampling Distribution

<table>
<thead>
<tr>
<th>Panel</th>
<th>Method</th>
<th>$\hat{\alpha}$</th>
<th>z-score</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>1</td>
<td>2.373</td>
<td>0.532</td>
<td>0.3272</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Market Cap Growth</td>
<td>2.373</td>
<td>0.532</td>
<td>0.3272</td>
</tr>
<tr>
<td></td>
<td>Cumulative Return</td>
<td>3.025</td>
<td>2.276</td>
<td>0.0058</td>
</tr>
<tr>
<td></td>
<td>Bessembinder (2017)</td>
<td>3.307</td>
<td>3.725</td>
<td>0.0003</td>
</tr>
<tr>
<td>B</td>
<td>2</td>
<td>2.373</td>
<td>1.988</td>
<td>0.0234</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Market Cap Growth</td>
<td>2.373</td>
<td>1.988</td>
<td>0.0234</td>
</tr>
<tr>
<td></td>
<td>Cumulative Return</td>
<td>3.025</td>
<td>5.081</td>
<td>0.0000</td>
</tr>
<tr>
<td></td>
<td>Bessembinder (2017)</td>
<td>3.307</td>
<td>6.419</td>
<td>0.0000</td>
</tr>
<tr>
<td>C</td>
<td>3</td>
<td>2.373</td>
<td>2.064</td>
<td>0.0194</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Market Cap Growth</td>
<td>2.373</td>
<td>2.064</td>
<td>0.0194</td>
</tr>
<tr>
<td></td>
<td>Cumulative Return</td>
<td>3.025</td>
<td>5.310</td>
<td>0.0000</td>
</tr>
<tr>
<td></td>
<td>Bessembinder (2017)</td>
<td>3.307</td>
<td>6.714</td>
<td>0.0000</td>
</tr>
</tbody>
</table>

Source: CRSP and simulations

Notes: We conduct 300 simulations of the stock market for each of the three methods estimating $\mu$ detailed in Appendix B. For each method, we consider the distribution of lifetime wealth creation as measured by market cap, cumulative return, and Bessembinder (2017) metric. Using the steps outlined in Appendix C, we estimate the power law parameter $\alpha$ for each distribution of wealth creation. Similarly, sampling distributions of $\alpha$ are obtained from the simulations. The first column shows the $\alpha$ estimated for the distribution of wealth creation of the 431 firms from January 1970 to December 1999. For each estimated alpha, the null hypothesis is that it represents a random sample from the sampling distribution of estimated alphas in our simulations. To test the hypothesis, we compute the z-scores and the corresponding p-values.
Table 9: Inference on Wealth Contribution of Top Ten Stocks with Sampling Distribution

<table>
<thead>
<tr>
<th>Panel A: Method 1</th>
<th>% Contributed</th>
<th>z-score</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Market Cap Growth</td>
<td>39.5%</td>
<td>0.027</td>
<td>0.4971</td>
</tr>
<tr>
<td>Cumulative Return</td>
<td>18.3%</td>
<td>-2.058</td>
<td>0.0162</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Panel B: Method 2</th>
<th>% Contributed</th>
<th>z-score</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Market Cap Growth</td>
<td>39.5%</td>
<td>-1.207</td>
<td>0.1137</td>
</tr>
<tr>
<td>Cumulative Return</td>
<td>18.3%</td>
<td>-2.471</td>
<td>0.0070</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Panel C: Method 3</th>
<th>% Contributed</th>
<th>z-score</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Market Cap Growth</td>
<td>39.5%</td>
<td>-1.297</td>
<td>0.0973</td>
</tr>
<tr>
<td>Cumulative Return</td>
<td>18.3%</td>
<td>-2.632</td>
<td>0.0042</td>
</tr>
</tbody>
</table>

Source: CRSP

Notes: We conduct 300 simulations of the stock market for each of the three methods estimating \( \mu \) detailed in Appendix B. For each method, we examine the percentage of wealth contributed by the top ten stocks with the greatest wealth creation as measured by market cap and cumulative return. We do not compute the percentage using the metric from Bessembinder (2017) due to negative wealth created observed in many simulations. Sampling distributions of the percentages are obtained from the simulations. For each percentage, the null hypothesis is that it represents a random sample from the sampling distribution of the percentages contributed in our simulations. To test the hypothesis, we compute the z-scores and corresponding p-values.
Figure 1: Distribution of monthly $\mu$ for each method. For each method of estimating $\mu$ detailed in Appendix B, we plot the histogram of monthly $\mu$s for the 431 firms in our universe. Method 1 uses CAPM; methods 2 and 3 rely on direct estimates from the returns data.
Panel A. Distribution of $E[r]$ estimated using Method 1

Panel B. Distribution of $E[r]$ estimated using Method 2

Panel C. Distribution of $E[r]$ estimated using Method 3

Figure 2: Distribution of monthly $E[r]$ for each method. For each method of estimating $\mu$ detailed in Appendix B, we plot the histogram of monthly expected returns for the 431 firms in our universe. Method 1 uses CAPM; methods 2 and 3 rely on direct estimates from the returns data.
Figure 3: Distribution of $\beta$ and monthly $\sigma$. We plot the histogram of market beta $\beta$ and monthly volatility $\sigma$ for the 431 firms in our universe. $\beta$ is estimated from a regression of log stock returns on log market returns, and $\sigma$ is estimated from historical volatility of each stock.
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 \( t \), \( \mu \) the drift parameter, \( \sigma \) the volatility parameter, and \( W(t) \) the value of a Wiener process at time \( t \). Applying Ito’s lemma to \( d\ln S(t) \):

\[
d\ln S(t) = \frac{1}{S(t)} dS(t) - \frac{1}{2} \frac{1}{S(t)^2} dS(t)^2
\]

\[
= \frac{1}{S(t)} S(t)[\mu dt + \sigma dW(t)] - \frac{1}{2} \frac{1}{S(t)^2} S(t)^2 [\sigma^2 dW(t)^2]
\]

\[
= \mu dt + \sigma dW(t) - \frac{1}{2} \sigma^2 dt
\]

Integrating each side,

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

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

\[
\ln R \sim N \left( \left( \mu - \frac{1}{2} \sigma^2 \right) t, \sigma^2 t \right)
\]

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

\[
E[R] = E[\exp(\ln R)] = \exp \left( \left( \mu - \frac{1}{2} \sigma^2 + \frac{1}{2} \sigma^2 \right) t \right) = \exp(\mu t)
\]
Appendix B. Estimation of Drift Parameter ($\mu$)

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 $\beta$ 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:

$$\mu - r_f = \beta (\log E[R_M] - r_f)$$

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

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

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

Method 2. Direct Estimation via Expectation of Log Returns

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

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

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

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

Method 3. Direct Estimation via Expectation of Total Returns

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

$$E[R] = \exp(\mu t)$$

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

$$\mu = \ln[E[R]]$$
Appendix 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 investor’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:

$$W_t - W_0(1 + r_f)^T = \prod_{t=1}^{T} I_{t-1}(r_t - r_f)(1 + r_f)^{T-t}$$
Appendix D. Estimation of Power Law Parameter

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

\[
p(x) = C x^{-\alpha}
\]

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

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

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

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

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

Estimating \( \alpha \) requires the choice of \( x_{\text{min}} \). We use the minimization of Kolmogorov-Smirnov statistic \( D \), defined as

\[
D = \max_x |F(x|\alpha, x_{\text{min}}) - F(x)|
\]

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

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