Is Index Concentration an Inevitable Consequence of Market-Capitalization Weighting?

Abstract

Market-cap-weighted equity indexes are ubiquitous. However, there are growing concerns that such indexes are increasingly concentrated in a few stocks. We ask: Does market-cap weighting inevitably lead to increased concentration over time? The question of inevitability arises from research that suggests the possibility of dominance by a few firms over time via a variety of plausible causal mechanisms. We study concentration in major equity market indexes over time and show that, despite recent concerns, concentration is not yet at levels that may be problematic, and for some indexes was higher in the past. Monte Carlo simulations calibrated to market data provide insight into various approaches to slow concentration, albeit at the expense of higher turnover.

Lisa R. Goldberg
Head of Research, Aperio by BlackRock. Three Harbor Drive, Suite 204, Sausalito, CA 04965, 415.339.4300. E-mail: lisa.goldberg@blackrock.com.

Ananth Madhavan
Head of Academic Engagement, External Affairs, BlackRock.

Harrison Selwitz
Quantitative Researcher, Aperio by BlackRock.

Alexander Shkolnik
Assistant Professor, University of California, Santa Barbara.

August 3, 2022

This material is for academic use and is not intended as the basis of investment advice or recommendations. The authors thank Xuan Che for expert research assistance, and Andrew Ang, Samara Cohen, Jennifer Hsui, Jason Ribando, and Nogie Udevbulu for helpful comments.

Many traditional equity indexes—originally intended as performance benchmarks to capture the entire market in each region, asset class, or sector—are weighted by market capitalization.[1] Market-cap-weighted schema have strong justification on theoretical and practical grounds. Theory comes from the capital asset pricing model (Sharpe 1964) where investors hold combinations of the market portfolio and riskless bonds. From a practical perspective, although all indexes need to periodically rebalance and reconstitute, cap-weighted indexes are naturally self-rebalancing, reducing turnover and transaction costs relative to alternative weighting schema.[2] Despite their advantages, concerns that cap-weighted indexes are overly concentrated in the largest stocks or in certain sectors are increasingly common.[3] Indeed, concerns over concentration and loss of diversification have led index providers to consider alternative index construction methodologies.

We ask: Will the use of market-cap weighting in indexes *inevitably* lead to increased dominance over time by the largest firms? The question of inevitability is suggested by research that shows, under a variety of possible and plausible causal mechanisms, that the distribution of firm size may be increasingly dominated over time by just a few firms. We begin by defining an intuitive metric for index concentration at a point in time. We then examine the time series of concentration for major equity indexes over the 25 years from 1996–2021. We find that the MSCI Emerging Markets Index and the S&P 500 index have seen the largest increases in concentration, with current levels above those of the peak of the dot-com era in 1999, but no such trend is evident for the MSCI EAFE Index. Given the prominence of the S&P 500 index, we examine in detail the concentration and turnover in that index from 1975 onward. Interestingly, although concentration in the S&P 500 Index is the highest in a quarter century, it was even higher in the 1970s.

Next, using the theory of power laws, we turn to the question of the time horizons over which concentration may become problematic. A random variable follows a power law distribution if the probability of a value exceeding x is proportional to x to the power of a negative constant. Power law distributions occur widely in physics, biology, earth and planetary sciences, demography, and the social sciences, including economics and finance. The theory provides a natural way to study index concentration, as well as to explore the impact of alternative index construction methodologies, including diversity weighting and two capping approaches, index constructors might take to mitigate concentration. Monte Carlo simulations calibrated to market data provide insight into an approach to slowing concentration, albeit at the expense of higher turnover.

Our paper contributes to the academic literature on index weighting and the impact of flows into index funds. Cap weighting has been questioned on both theoretical and empirical grounds.[4] Jiang et al. (2020) provide a model with noise traders where flows into index funds raise the prices of large stocks in the index disproportionately more than the prices of small stocks, implying higher expected returns for a small-minus-large portfolio. Note that this concern is distinct from the discussion about smart beta indexes, which is a different topic altogether. Several empirical studies have discussed how size and sector bias may limit diversification and how dominance by stocks whose valuations have risen quickly may add to risk.[5]

Questions of concentration have also gained more prominence recently because of the growth of custom indexes that seek exposure to a particular subindustry, style factor, or theme.[6] The proliferation of exchange-traded funds (ETFs) has also played an important role in the growth of custom indexes. These new indexes typically use market-cap weighting within their geographies or industries but may use alternative weighting schemes to avoid being dominated by the largest firms in their segments, whether a subindustry or style factor.[7] As a result, flows into these ETFs may result in large volumes and price movements in the smaller stocks of the index.

A final concern with traditional equity indexes is in how countries or firms are selected for index inclusion. Even well-established indexes based on published index construction methodologies have subjective elements. The S&P 500 index, for example, uses fundamental rules (four quarters of earnings, float, and liquidity requirements) to determine inclusion but ultimately relies on subjective judgement from a nine-person committee. Further, criteria for index inclusion selected by index providers does change over time and is also subjective in nature, depending on factors such as whether the company has dual-class shares, domicile, etc.

We note in passing that our analysis is distinct from fundamental investing, which weights securities on accounting or fundamental characteristics such as profits or book value, to avoid investing in stocks that are potentially overpriced and hence may dominate a traditional market-cap-weighted index. Hsu (2006), Arnott and Hsu (2008), and Arnott et al. (2008) provide a rationale for alternative indexing, Chow et al. (2011) provide a survey, and Perold (2007), and

Graham (2012) offer counterarguments. We do not make any statements about the return properties of alternative weighting schema (see Fernholz 1999), choosing to focus on the time-series of concentration.

I. Index Breadth and Concentration

i. Measurement

Figure 1 shows the weight of each stock in the S&P 500 index as proxied by the iShares S&P 500 Index ETF (ticker: IVV) as of December 31, 2020, where stocks are ranked with 1 being the greatest weight. (Because of corporate actions, the index has more than 500 constituents; this variation is not uncommon.) We observe that the largest few stocks account for a significant fraction of the index relative to the tail, a characteristic of power law distributions.

Simple measures of concentration (such as the weight of the top five or ten constituents) miss the distribution outside the extreme largest stocks. One measure that looks at the whole distribution is breadth, defined as the inverse of the Herfindahl-Hirschman Index (HHI) over n stocks in the universe:

$$B_n = 1/\sum_{i=1}^{n}(S_i)^2 \tag{1}$$

where S_i is the weight (in decimal) of stock $i = 1, \dots, n$. HHI is widely used as a concentration metric in industrial organization, but breadth is a more intuitive metric for our purposes.[8] In an equally weighted index of 500 stocks, the weight of stock i is $S_i = 0.002$ so $B = 500$; a fund holding a single asset, e.g., a gold ETF, would have $B = 1$. Using equation (1), we find that the breadth of the S&P 500 is 67.5 names as of December 31, 2021. This figure in itself is not particularly revealing, but breadth gives rise to a natural measure of concentration that we use to compare across indexes at a point in time or to measure changes in a particular index over time.

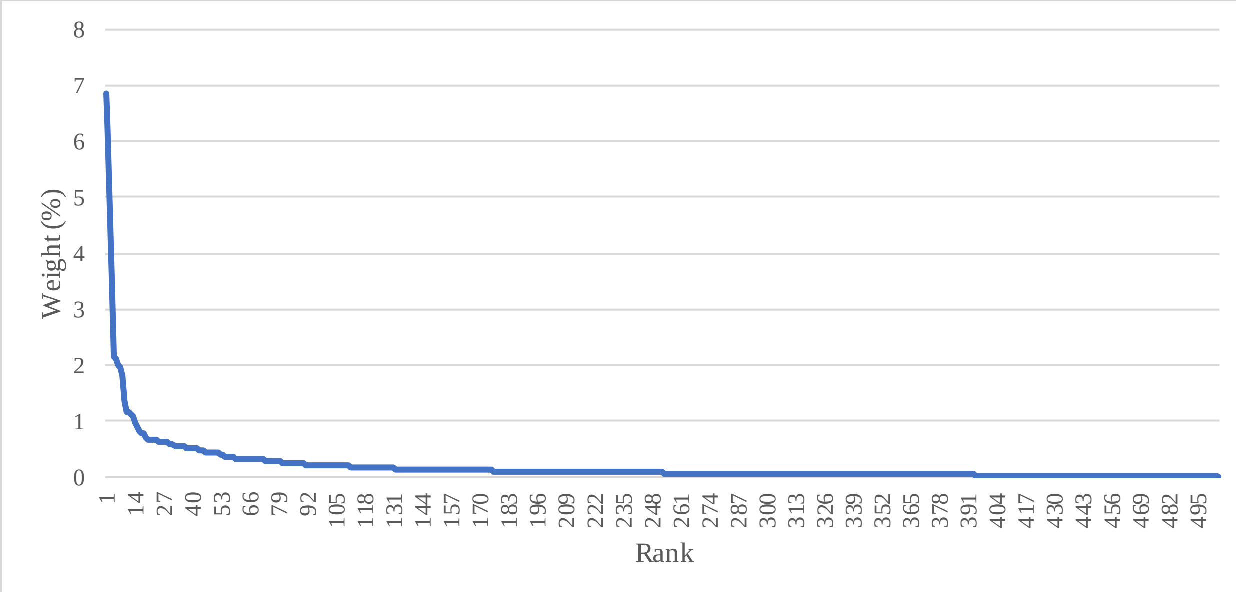

Figure 1 S&P 500 index weights as of December 31, 2020, by rank.

Source: BlackRock, based on index holdings data from MSCI on December 31, 2021.

Specifically, we scale B by the number of constituents, with $n \gg 1$ and define concentration as:

$$C_n = 1 - (B_n/n), \tag{2}$$

Concentration lies between 0 and 1. The case of $C_n = 0$ corresponds to an equal-weighting scheme, whereas $C_n \approx 1$ corresponds to dominance by one asset. Industry rotation may mitigate the impact of concentration because the composition of the topmost index ranks is not stable. In March 2020, Berkshire Hathaway was the fifth-largest stock until it was displaced by Alphabet, which itself was ultimately displaced by the entry of Tesla on December 21, 2020, as the largest weighting ever added to the index.[9] Indeed, none of the top five names in the S&P 500 index in December 2000 were in the top five in December 2020, as shown in Table 1.

Table 1 Rotation in top five names of S&P 500 index, 2000–2021.

December 2000	December 2010	September 2020	January 2021
General Electric	Exxon Mobil	Apple	Apple
Exxon Mobil	Apple	Microsoft	Microsoft
Pfizer	Microsoft	Amazon	Amazon
Cisco	General Electric	Facebook	Facebook
Citigroup	Chevron	Alphabet	Tesla

Sources: BlackRock; Bloomberg, as of December 31, 2021; and Debru (2020).

ii. Empirical evidence on concentration

To study the time series of concentration, we focus first on the past 25 years and four major equity indexes: S&P 500; Russell 2000; MSCI EAFE; and MSCI Emerging Markets. Figure 2 shows the history concentration (C) for the four indexes from 1996–2021. Since indexes are not directly investible, we use the respective iShares ETFs as proxies[10] and study their actual holdings as of December 31 of each year from 1996–2021. For one very large index, EAFE, there is no evidence that concentration increased from 1995–2021, and indeed concentration appears to have decreased.

Concentration was highest in emerging markets, first decreasing from 1996 to 2010 and then rising thereafter. But for the S&P 500, breadth declined from a maximum of 141.5 on December 31, 2016, to 67.5 on December 31, 2021, while concentration rose from 0.72 to 0.87 in that period, the highest on record. We see rising concentration in the Russell 2000, but with a value of 0.54 as of December 31, 2021 (and breadth of 934 names), this index was less concentrated than the others. We turn now to a deeper analysis of concentration in the largest securities in an index. We use research on power laws to understand if there is systematic tendency for the upper tail to become more concentrated over time.

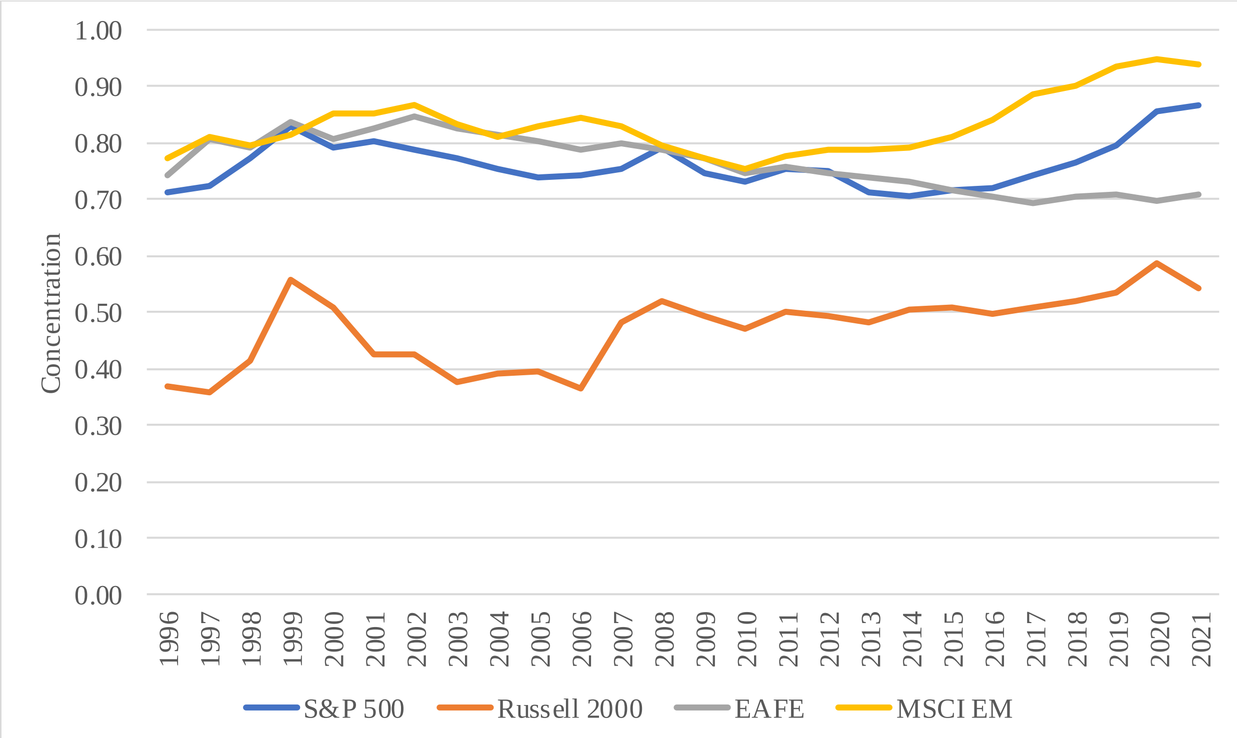

Figure 2 Time series in concentration from 1996–2021. Concentration is computed using actual iShares holdings of funds tracking these indexes as of December 31 of each year.

Source: Authors' estimates based on data from BlackRock.

II. Power Laws and Index Concentration

i. Definitions

The distributions of the populations of cities, earthquake magnitudes, solar flares, frequency of use of words in language, firm sizes, income, and wealth have been modeled using power laws.[11] In financial markets, extreme return events are more likely and larger in magnitude than suggested by normality as witnessed by events such as the Quant Quake of 2007, Global Financial Crisis of 2009, Flash Crash of 2010, Covid Crisis of 2020, etc.[12] To our knowledge, this analysis is the first application of power laws to study index weighting schema.

ii. Modeling extreme events

Mathematically, a power law is a relation of the type $Y = \zeta X^{-\alpha}$, where Y and X are variables of interest, α is the power law exponent, and ζ is a constant. Pareto's law (see Pareto [1896] and Mandelbrot [1960]) has the cumulative distribution function:

$$P(X > x) = \zeta x^{-\alpha}, \tag{3}$$

for $X > x_{min}$ where $x_{min} > 0$, ζ is a normalized constant to the area under the distribution function sums to 1, and $\alpha > 0$ is the exponent of the power law. This idea dates to Pareto's analysis of the proportion of individuals with an income above a certain level x. A lower α means a higher degree of inequality in the distribution, i.e., a greater probability of finding very large values. A special but important case is Zipf's law where $\alpha = 1$. A linguist, Zipf (1949) found that the frequency of the i^{th} most common word in English text, denoted by S_i, is inversely proportional to its rank i out of n, so that we obtain:

$$S_i \sim 1/i. \tag{4}$$

Tests typically involve inverting Zipf's law in equation (4) to obtain:

$$P(S_i > x) \approx Rank(i) = \zeta x^{-\alpha} \tag{5}$$

Equation (5) implies that:

$$\ln(i) = \ln(\zeta) - \alpha \cdot \ln(S_i) \tag{6}$$

The power law coefficient in equation (6) can be estimated via a linear regression of log rank on log size.

iii. Example: US city sizes

Gabaix (1999) examines the size of US cities and finds support for Zipf's law.[13] He ranks city size (1 = New York City, 2 = Los Angeles, ...) and regressed the log rank on the log of city population and estimates the power coefficient $\alpha \approx 1$. Figure 3 shows an update of Gabaix (1999) using US city population as of June 2020. Each dot is a city, and the line shown is the ordinary least squares (OLS) estimate of equation (5).

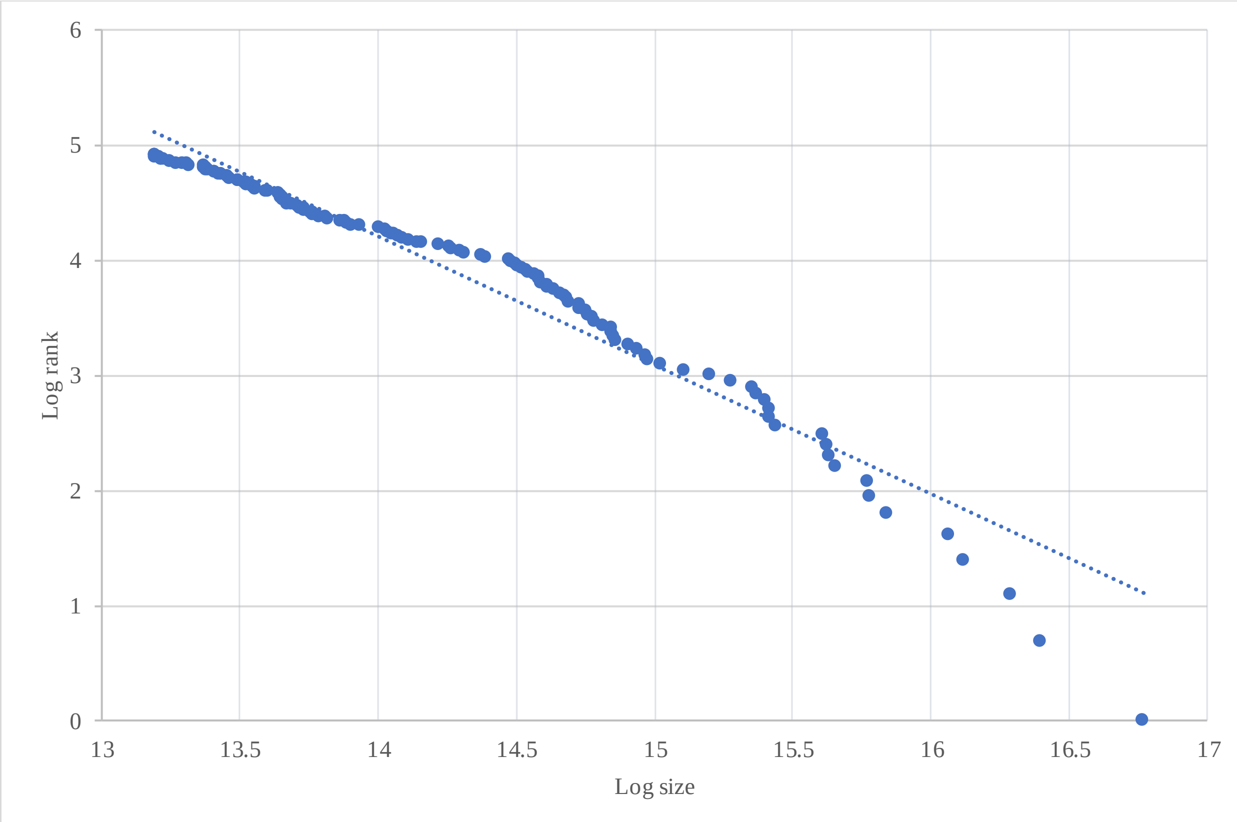

Figure 3 Log rank–log size regression for size of US cities: June 2020.

Source: Authors' estimates based on data from the Statistical Abstract of the United States (2020).

Consistent with Gabaix (1999), we focus on the top 135 metro regions and estimate a power law coefficient of α = 1.12 and R^2 = 0.95.[14] This is slightly higher in (absolute) magnitude than the estimate in Gabaix (1999). We also observe that Zipf does not fit well for the top ten cities shown on the right-hand side of the plot, although it does fit the lower tail well. When we turn to indexes, we will again see deviation from linearity in log rank/log size plots. There, we find that the relatively substantial tails of the distributions are well fit with power laws.

While the OLS estimates are unbiased and consistent, the estimated standard error (0.022) needs modification because of the autocorrelation induced by the ranking procedure. Gabaix and Ibragimov (2011) show that the correct standard error is:

$$SE = |\alpha|\sqrt{(2/n)} \tag{7}$$

where α is the OLS estimate of the slope coefficient in equation (6). This calculation yields an adjusted standard error of 0.136, so we cannot reject the null hypothesis that the distribution is Zipf. The regression is, however, very sensitive to the choice of truncation point. When we use 415 metro regions as our cutoff, we obtain $\alpha = 0.81$, and the null hypothesis that the power law coefficient is Zipf is rejected.[15] We conclude that Zipf may be a good descriptor in the tails but may not capture the size distribution of the largest metro areas.

iv. Using power law distributions to model index concentration

Let us revisit the results on the four indexes of interest using the regression approach of equation (6) to investigate whether power laws capture the concentration in the largest names. Figure 4 shows a scatter plot of log rank on log size for four public equity indexes on December 31, 2021. There is a pronounced concave relationship between log rank and log size for all four indexes,[16] and each has a substantial, linear tail. This relationship is present throughout our data set. The concavity means that the distribution of firm sizes cannot, in its entirety, be well fit with a power law. Instead, we use formula (6) to fit power laws to the tails for the distributions, where the log rank/log size plots are approximately linear.[17] For each index on each date, we compute the mean m and standard deviation s of log size. Then, we fit a regression line to observations for which log size exceeds the cutoff $m + s$.

The table below the scatter plots in Figure 4 shows the estimated power law coefficients, the standard error of the estimate computed by formula (7), the number of observations in each tail, and the goodness of fits measured by R-squared. The fits are excellent, with R-squared values greater than or equal to 0.97. We reject the Zipf distribution for three of the four indexes since the value of 1 is outside the interval alpha +/ SE. For the S&P 500, the value 1 is within the interval, and we fail to reject the Zipf distribution as a descriptor of the lower tail of the distribution, i.e., the stocks on the right of Figure 1.

Index	Power law coefficient	Standard error	Number of observations	R^2
S&P 500	1.33	0.21	82	0.99
Russell 2000	3.08	0.22	377	0.97
MSCI EAFE	1.87	0.22	139	0.99
MSCI EM	1.40	0.14	207	0.99

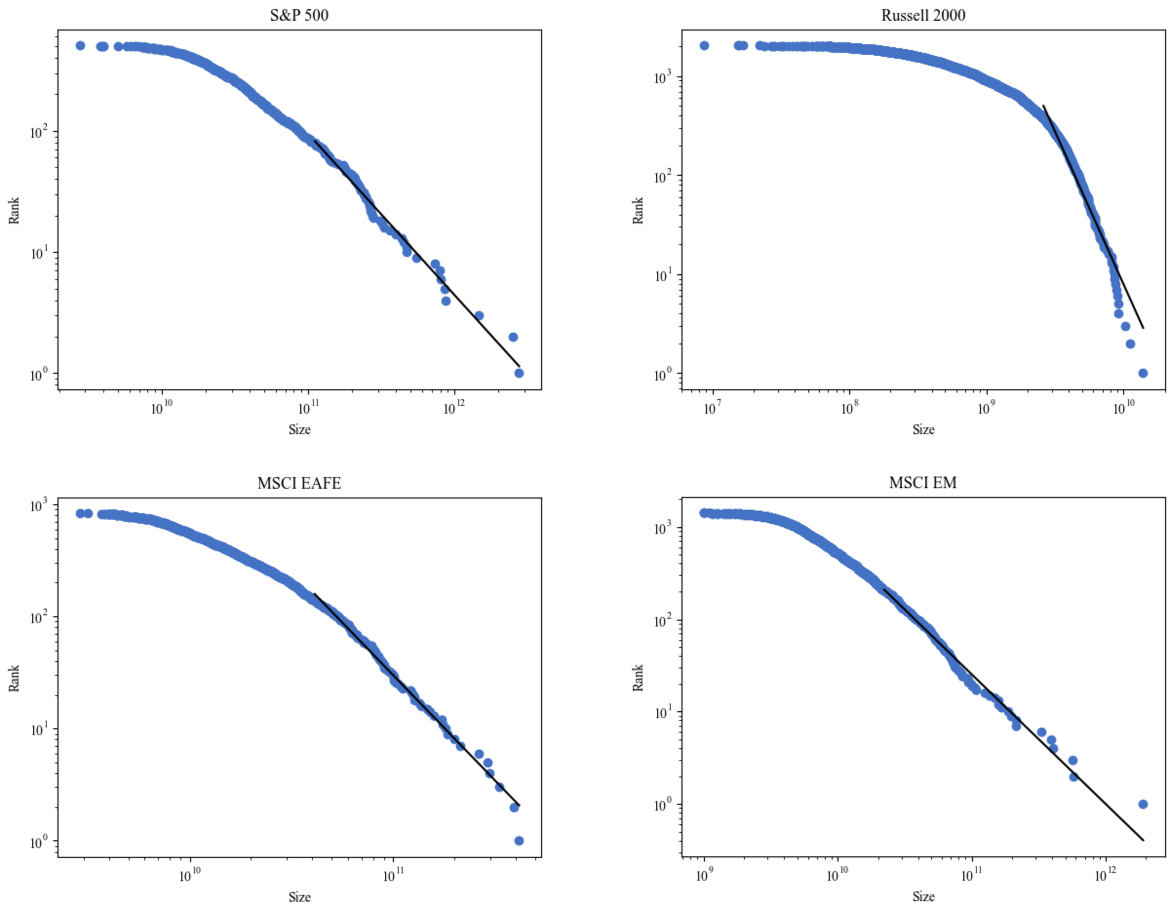

Figure 4 Rank-size scatter plots for four public equity indexes: December 2021.
Source: Authors' estimates based on data from BlackRock as of December 31, 2021.

Figure 5 plots the estimated power law coefficient α by year for the four indexes. A comparison of Figures 2 and 5 illustrates the inverse relationship between concentration and power law coefficients. Concentration was lowest and estimated power law coefficients were highest and least stable for the Russell 2000, the one index in the group that excludes the largest securities in its market.

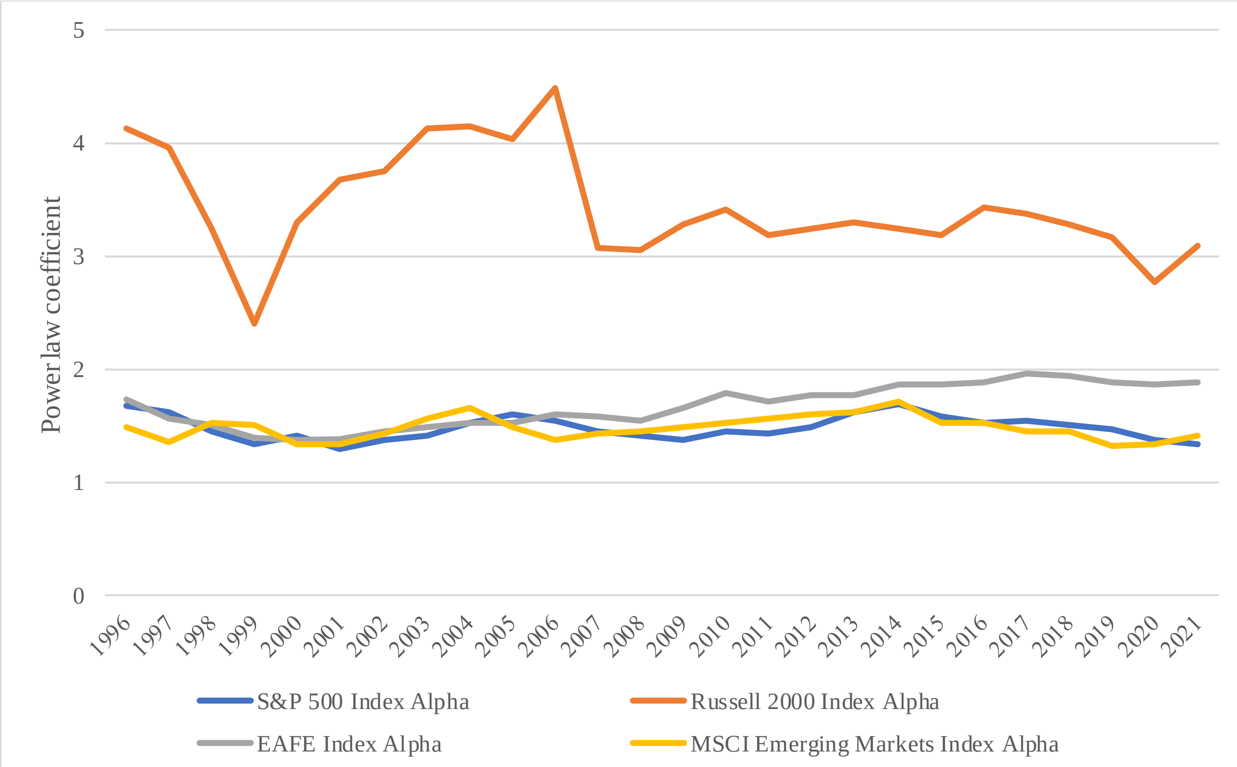

Figure 5 Estimated power law coefficients from 1996–2021 for major equity indexes.
Source: Authors' estimates based on data from BlackRock as of December 31 of each year.

Table 2 shows, for each index, averages over time of the estimated power law coefficient, the standard error computed by formula (7), the number of observations in the tail, and the goodness of fit measured by R-squared. The fits are very high throughout the study period, with average R-squared values ranging from 0.94 for the Russell 2000 Index to 0.99 for the MSCI EM Index. The Zipf distribution is rejected throughout the study period for three of the four indexes, and it is borderline for the S&P 500 index.

Table 2 Average statistics for estimated power law coefficients from 1996–2021 for major equity indexes.

Index	Power law coefficient	Standard error	Number of observations	R^2
S&P 500	1.48	0.24	78.54	0.96
Russell 2000	3.43	0.26	349.15	0.94
MSCI EAFE	1.67	0.19	162.81	0.96
MSCI EM	1.48	0.18	143.69	0.99

Source: Authors' estimates based on data from 1996–2021 sourced from BlackRock as of December 31, 2021.

v. Zipf and current index concentration

The discussion so far indicates that although the tails of market-cap distributions are well described by power laws, the coefficients are significantly different from 1. Here, we compare empirical index weights directly to weights predicted by a Zipf distribution.

The weight of constituent *i* under Zipf is given by:

$$S_i = 1/(i \sum_{j=1}^{n} 1/j), \tag{8}$$

where n is the number of constituents. Define $H_n^z = \sum_{j=1}^{n} 1/j \approx \ln(n) + \gamma$, where $\gamma = 0.577\ldots$ is the Euler-Mascheroni constant. The Zipf weights are (for $n > 1$) given by:

$$S_i = 1/iH_n^z \tag{9}$$

which is decreasing in n. For the largest five constituents from n, we obtain: $S_1 = 1/H_n^z$, $S_2 = 1/2H_n^z$,…, $S_5 = 1/5H_n^z$. So, applying Zipf's Law to the S&P 500, the top five constituents with $n = 500$ would have weights, in percent, of 14.69, 7.34, 4.90, 3.67, and 2.94, respectively. Figure 6 compares the top 20 theoretical and actual weights (in percent) as of December 31, 2021, based on holdings of the iShares fund that seeks to track this index.

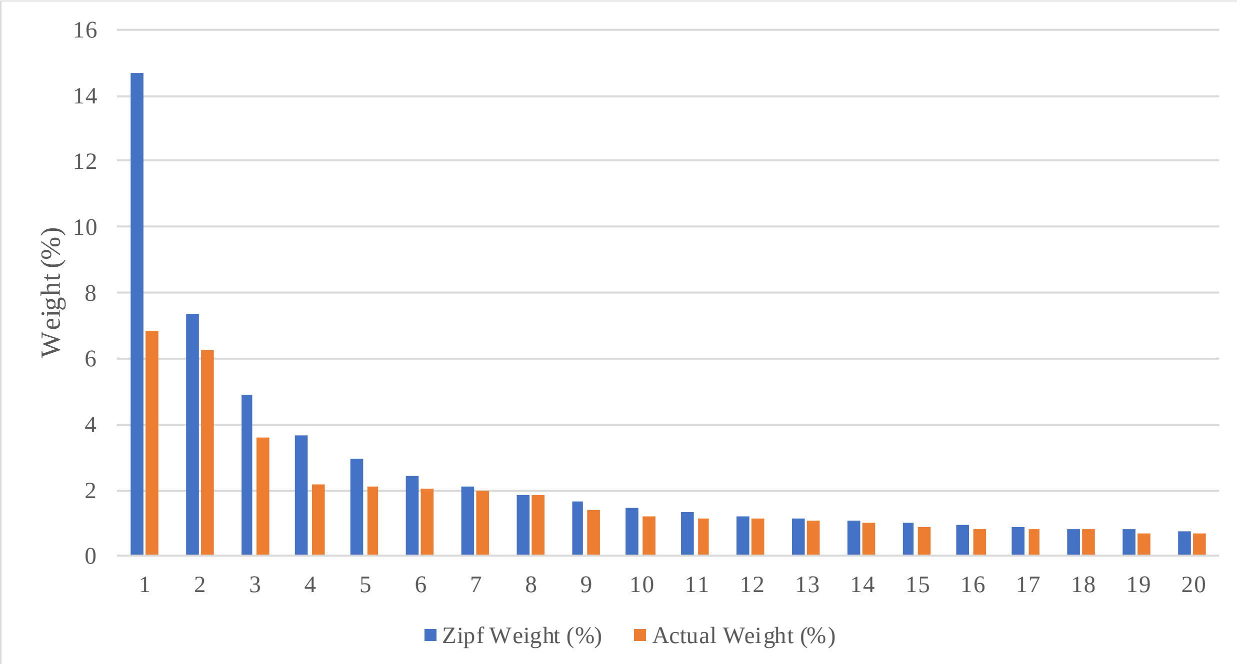

Figure 6 Top 20 Zipf and actual weights for the S&P 500 index.
Source: Authors' estimates as of December 31, 2021, based on index holdings data from MSCI.

Figure 7 shows that we are still very far from Zipf in terms of concentration in the upper tail, that is the left-most 20 stocks of Figure 1 with $n = 500$. Indeed, the implied breadth of the Zipf distribution is $B = 28.2$, implying a high concentration $C = 0.95$, well above the actual index concentration. The actual top five holding weights (in percent, as of December 31, 2020) were 6.68, 5.30, 4.38, 2.07, and 1.69. As with the analysis of city size in Figure 3, we observe deviations from the Zipf weights in the largest stocks. However, Zipf fits the tail of the distribution well.

Will expanding the number of constituents in the index (n) mitigate concentration? Since the Zipfian weight $S_i = 1/iH_n^z$ from equation (9) is decreasing in n, this seems plausible. However, the answer is no. Figure 7 shows the Zipfian weights of the top security (lower line) and the top ten securities (upper line) as a function of n. We see that the weights of the top securities quickly decrease as n increases but approach a constant level asymptotically. So, even if the S&P 500 were expanded to 2,000 stocks, we still would not be able to lower the Zipf weight of the top stock to below 10% or the weight of the top ten to below 35%.

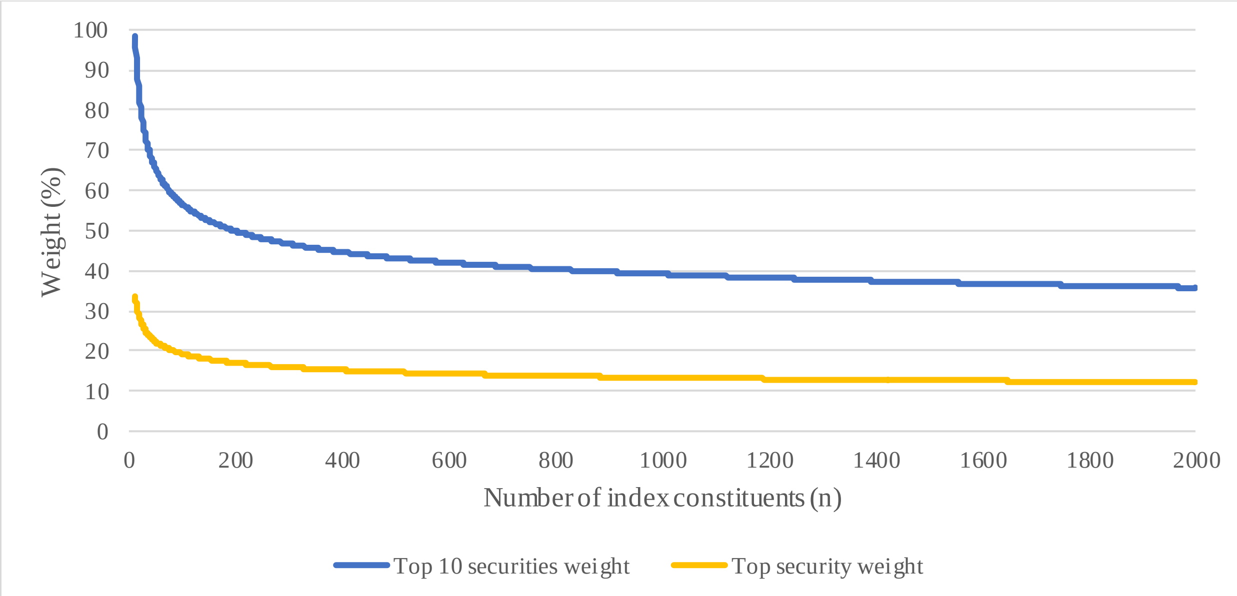

Figure 7 Zipfian weight of the top security (bottom line) and top ten securities (upper line) as a function of the number of constituents n.

Source: Authors' computations based on equation (9).

In summary, although a power law may a be good descriptor of the tail of the distribution of index weights, we still have a long way to go before we get to levels of concentration they eventually will imply in the largest stocks. This is not to say that the growth dynamics explored above may result in rising concentration over time. We turn now to an analysis of various proposals to reduce concentration in indexes.

III. Economic Basis for Concentration

A considerable body of work has gone into explaining why power laws are found in so many diverse areas. In this section, we sketch one explanation for rising concentration in indexes—a phenomenon known as Gibrat's law. Following Gabaix (1999),[18] consider a stock i whose market capitalization (scaled to the total market cap of the index) at time t is denoted $S_{i,t}$. These weights evolve randomly $S_{i,t+1} = r_{i,t+1}\, S_{i,t}$ where $r_{i,t+1} = \exp(g_{i,t+1})$ and $g_{i,t+1}$ is a random growth rate (not necessarily normal) where we normalize the mean to zero. This is known as Gibrat's law. Since $\sum_{i=1}^{N} S_{i,t} = 1$, we have $\mathrm{E}[r] = \int_0^\infty r f(r)\, dr = 1$ where $f(r)$ is the distribution function of r, which in turn derives from the distribution function of the random growth rates. At time t, the distribution of weights is $G_t(S) = P(S_t > x)$. If there exists a stationary distribution[19] $G(S)$ of index weights, then $G_{t+1}(S) = G_t(S) = G(S)$. Now recall that:

$$G_{t+1}(S) = P(S_{t+1} > S) = P(r_{t+1}S_t > S)$$

$$= G_t(\frac{S}{r}) = \int_0^\infty G(\frac{S}{r}) f(r)\, d\gamma \quad (10)$$

Try $G(S) = b/S$ where b is a constant and check that this works (using our earlier result that $\int_0^\infty rf(r)dr = 1$). The long-run steady-state distribution is Zipf (aka Pareto). Consequently, it does not have a mean or any higher moments. But Zipf may be useful for understanding the extreme portion of the probability distribution above a certain size. Scale invariance drives this result. Of course, this example is just to motivate why power laws might be useful in modeling concentration in indexes.[20] We do not make a statement that this outcome is inevitable, as explained in more detail below.

Suppose each firm in an index grows randomly so that $S_{i,t+1} = \gamma_{i,t+1}\, S_{i,t}$ where:

$$\gamma_{i,t+1} = \exp(g_{i,t+1} + e_{i,t+1}) \quad (11)$$

Here, $g_{i,t+1}$ is interpreted as the organic growth rate (which could be positive or negative) plus a random shock $e_{i,t+1}$ so that:

$$\ln(S_{i,T}) = \ln(S_{i,0}) + \sum g_{i,t} + \sum e_{i,t} \quad (12)$$

As T gets large, the distribution of size becomes a log normal distribution. Unlike Zipf, the moments of a log normal distribution are all well-defined. If indexation accounts for a constant fraction of total market cap, the distribution of dollar market capitalization of individual firms in the universe can be directly translated to weights. But for large T, the distribution is log-normal albeit with huge variance and a few funds dominate the distribution. Gabaix (1999) shows that with a minimum threshold for viable inclusion size (quite plausible in the index context), we get a Pareto distribution.

IV. Mitigating Concentration

Index providers too have experimented with ways to diversify and have made changes to their methodologies to reflect new developments. We examine the impact of alternative weighting schemes on the concentration of cap-weighted indexes. A popular scheme is equal weighting (e.g., weight of $1/N$), which has the advantage of simplicity, albeit with high turnover.[21] Our focus here is on approaches that preserve cap weighting to some extent. We first consider the MSCI 25/50 rule, which mandates that no single issuer should account for more than 25% of an index, and the aggregate weight of issuers each exceeding 5% of the index should not exceed 50% of the index's total assets. We next consider the UCITS 5/10/40 rule, which is similar in form: it limits ownership in a single issuer to 10% and caps total ownership of issuers with weight exceeding 5% to 40%. Alternative approaches derive from mathematical theory and include entropy weighting and diversity weighting. Define S_i to be the market cap weight of i. Then, the entropy weight is:

$$e_i = -S_i \ln(S_i) / \sum_{j=1}^{n} -S_j \ln(S_j) \tag{13}$$

Entropy weighting has strong support from information theory. Diversity weighting (see Fernholz 2005) is defined for $0 < p < 1$ as:

$$\pi_i = S_i^p / \sum_{j=1}^{n} S_j^p \tag{14}$$

As p ranges from 0 to 1, diversity weighting interpolates between equal weighting and market-cap weighting. For $p > 0$, the diversity weight is monotonic in the cap weight, so the ranking of securities is the same for cap weighting and diversity weighting. In contrast, the entropy function is not monotonic; it diminishes ranks of large and small securities relative to mid-size securities. To build intuition, suppose the index consists of only 3 constituents with weights ordered from largest to smallest of 0.5, 0.3, and 0.2, respectively. Then, the corresponding weights under entropy weighting using equation (13) are 0.337, 0.351, and 0.313, respectively. Under diversity weighting (with $p = 0.5$) we obtain 0.415, 0.322, and 0.263, respectively, using equation (14). So, both schemes down-weight the largest security and up-weight the smallest. This example also demonstrates entropy weighting's non-monotonicity; the ranks of the top two cap-weighted holdings are reversed. Given the generality of diversity weighting, we will focus on this weighting scheme in our subsequent analysis.

Fernholz (2005) defines a summary measure of Diversity $D(p)$ with $p > 0$ as:

$$D(p) = \left(\sum_{i=1}^{n} S_i^p\right)^{1/p}. \tag{15}$$

In our simple 3 constituent example above with $p = 0.5$, $D(p) = 2.897$. By contrast, with $p = 1$ under market cap weighting, $D(p) = 1$.[22] The properties of diversity weighting strategies are developed in Fernholz et al. (1998), Fernholz (1999), Fernholz (2002), and Fernholz (2005). Diversity weighting is a generalization of equal weighting, which is studied in Booth and Fama (1992) and Fernholz and Maguire (2007). Fernholz (1999) argues that diversity weighting will outperform cap weighting over the long run. This finding may also reflect returns to a size factor arising from the up-weighting of the smallest stocks and down-weighting of the largest stocks under diversity weighting.

i. Detailed analysis

We examine the ability of various rules to mitigate concentration in the S&P 500 index, expanding the date range from January 31, 1975, to December 31, 2021, with a monthly data frequency using data from MSCI. Over this period, neither the MSCI 25/50 nor UCITS 5/10/40 rules were binding on the S&P 500 index, so we focus our empirical analysis on diversity weighting.

In Figure 8, we show the time evolution of concentration in the S&P 500 index and its diversity-weighted counterpart. Trends in concentration, breadth, and power law exponent of the diversity-weighted and cap-weighted S&P 500 index are in sync, but the absolute levels are materially different. At the end of 2021, for example, the breadth of the S&P 500 index, or effective number of securities, was 67.5, while diversity weighting increased that value to 327.1. The corresponding power law coefficient increased from 0.89 to 1.78. We also include the equal-weighted index for comparisons in breadth, concentration, and turnover.

Diminished concentration in a cap-weighted index comes at the cost of excess turnover, which was generally increased by diversity and equal weighting, as shown in Figure 9. Exceptions include periods in which large-cap securities entered or exited the index. In these cases, the reduction in weights of these securities under diversity weighting trumped the excess turnover. One such example is the entrance of Tesla in December 2020. As of December 31, 2020, Tesla's weight in the S&P 500 index was 1.69%. Diversity-weighted, Tesla would comprise 0.72% of the index, enough reduction to bring the total month-to-month turnover down below that of the cap-weighted index. The enormous spike in turnover in the equal-weighted index in 1976 is explained by the large number of additions and deletions in that year. Figure 8 illustrates the positive association between breadth and power law. In our data sample, the correlation between the two measures is 0.45.

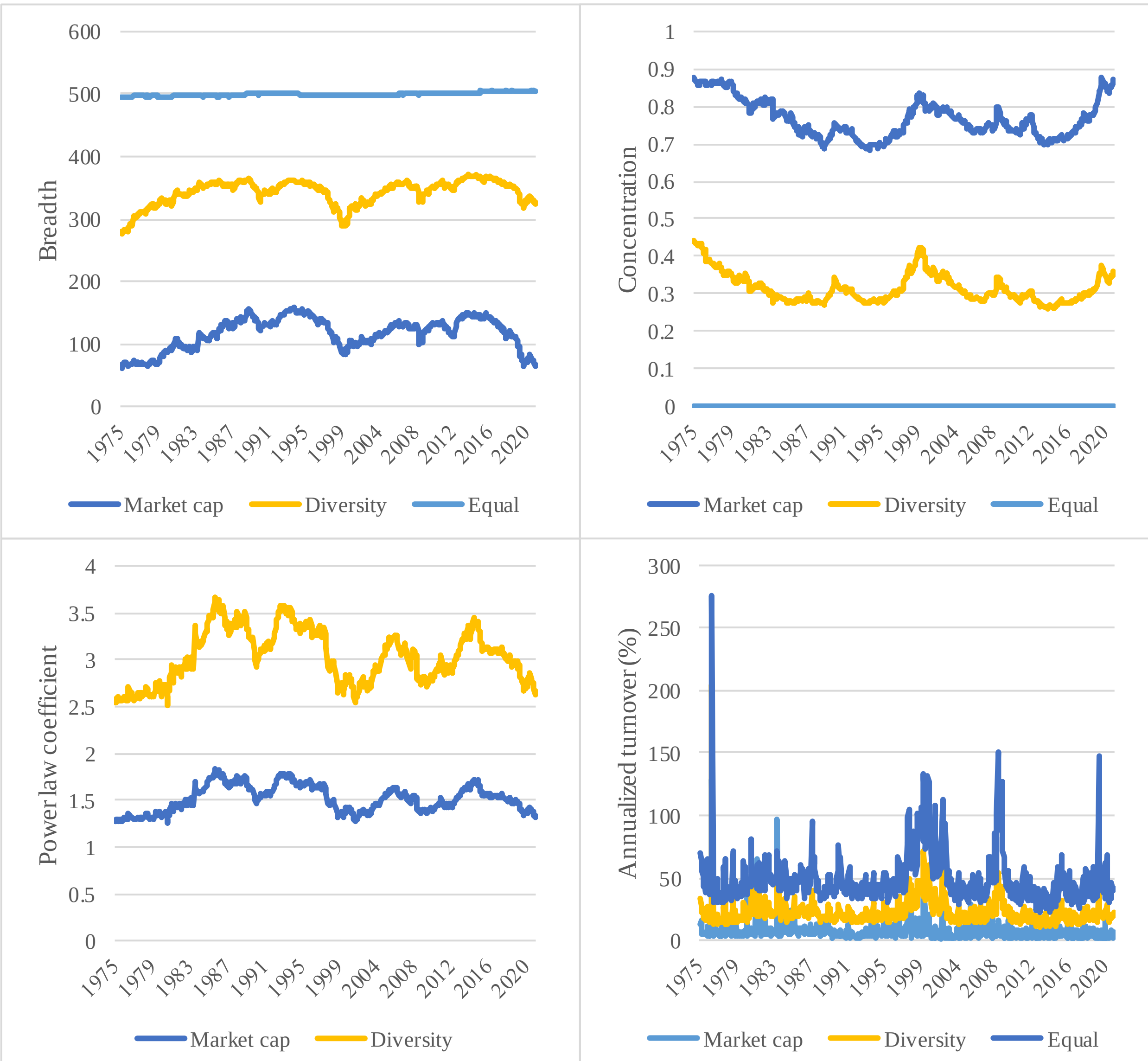

Figure 8 Concentration metrics and turnover for historical S&P 500.

Source: Authors' estimates using monthly market cap and index holdings data from MSCI.

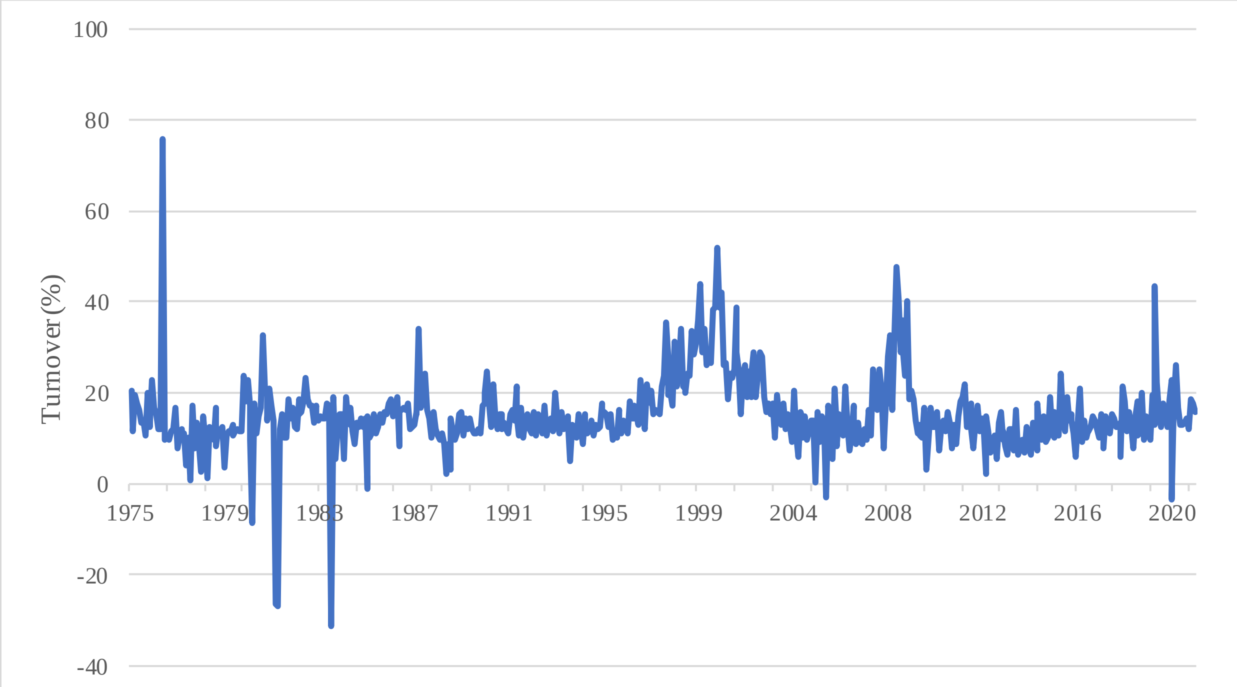

Figure 9 Difference in annualized turnover between market-cap and diversity-weighted S&P 500 index.

Source: Authors' estimates using monthly market cap and index holdings data from MSCI.

ii. Using simulation to evaluate concentration mitigation schemes

Simulation sheds light on concentration mitigation schemes. We follow Gabaix (1999), which models the evolution of normalized market cap, the fraction of the market accounted for by a firm, with a reflecting geometric Brownian motion, which we specify in formulas (16)-(20).[23] We begin with the growth rate distribution, which is expressed in terms of an arithmetic Brownian motion that is 0 at time 0:

$$dX_t = m_t + \sigma B_t. \tag{16}$$

The evolution of normalized market caps over time is given by

$$S_t = S_{min}\, exp(X_t + L_t) \tag{17}$$

where $L_t = inf_0 \leq s \leq t \cdot min(0, X_s)$, and $S_{\min}$ is the smallest fraction of the market that can be realized by a firm. In this setting, S hits the barrier $S_{\min}$ precisely when the driving Brownian motion

X hits a new low. As explained in Gabaix (1999), when the underlying Brownian motion's drift is negative and given by:

$$m = (-\sigma^2/2) * 1/(1 - S_{min}/S_{avg}), \tag{18}$$

the distribution of S achieves a steady-state power law distribution with exponent

$$\alpha = 1/(1 - S_{min}/S_{avg}) \tag{19}$$

Here, S_{avg} is the expected value of S_t, which is the same for each time t. In a market with N firms, we set S_{avg}= 1/N, which gives us the steady-state power law exponent:[24]

$$\alpha = 1/(1 - NS_{min}) \tag{20}$$

To fully specify our model, we provide values for parameters, which are calibrated to market data. In the simulations shown below, we set volatility sigma to 0.118, which is calibrated from monthly normalized growth of firms in the S&P 500 index between January 1975 and January 2022. This translates to an annualized volatility of approximately 41%, which is reasonable for individual stocks. Portfolio volatility is, of course, much lower. We set the product NS_{min}= 0.35, leading to a power law exponent of a little more than 1.5. Our hypothetical index is taken to be the 500 largest firms in a universe of N= 4,000 securities. Simulation results are based on 1,000 experiments.

Continuing our focus from the empirical section on cap-weighting and diversity weighting with p = 0.5, we show breadth, concentration, power law exponent, and annualized turnover for our simulated indexes in Figure 10.[25] Across our 1,000 experiments, diversity weighting elevated the median breadth of the simulated cap-weighted index from 123 to 407. The corresponding median values of concentration were 0.75 and 0.19. Diversity weighting raised the power law exponent of the simulated cap-weighted index from 1.51 to 3.03.[26] The cost of diversity can be measured in turnover, whose annualized median value was 40% for the simulated cap-weighted index and 124% for its diversity-weighted counterpart.

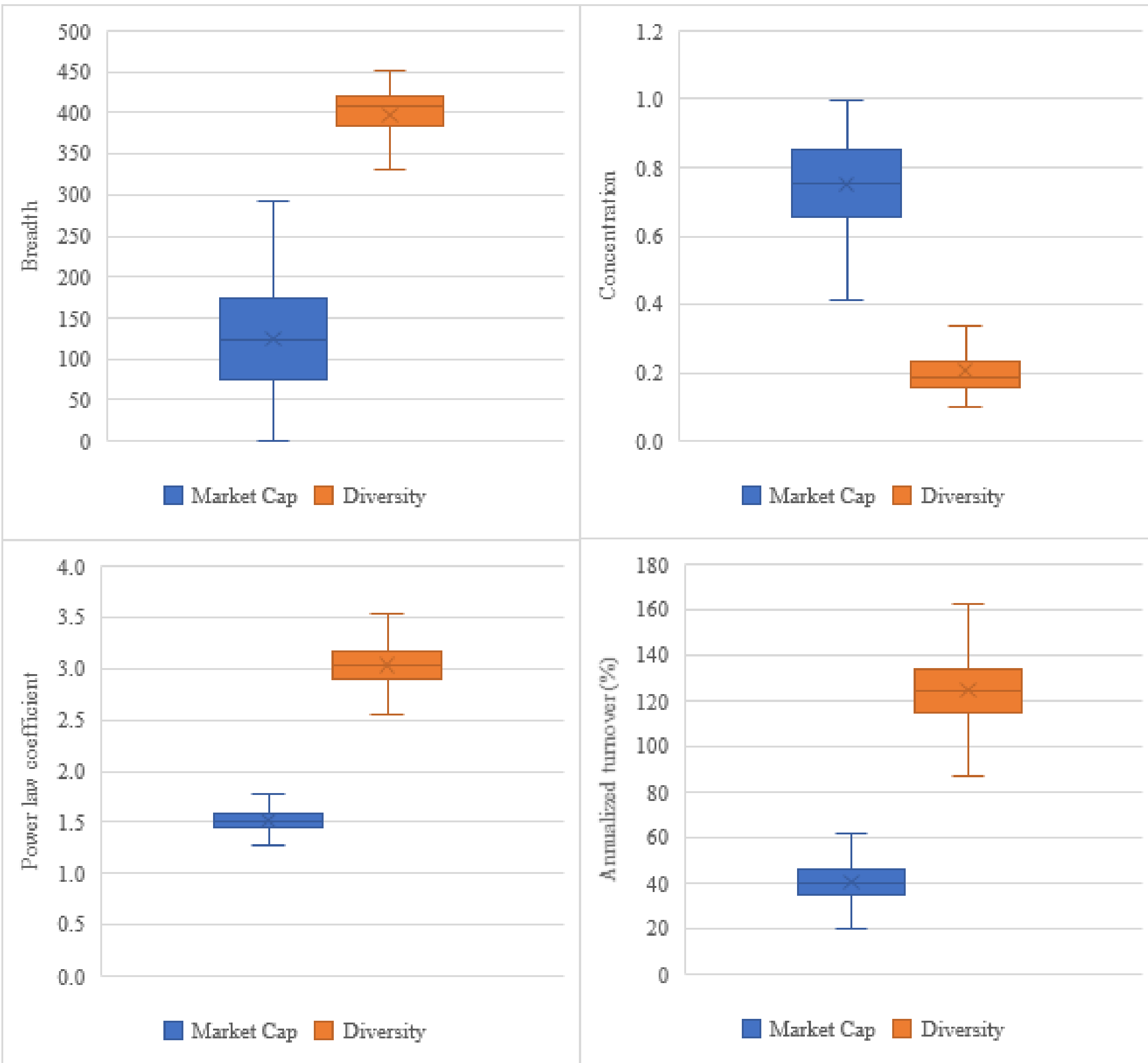

Figure 10 Concentration metrics and turnover for simulated indexes.

Source: Authors' computations.

So, with turnover approximately three times higher in the simulated diversity-weighted index, the corresponding transaction cost drag will similarly be higher by a factor of three. Assuming costs of 0.2% in individual stocks, this amounts to an annual drag of 0.248% vs. 0.08% under cap weighting. With, say, $24 trillion following major indexes, this amounts to additional costs of $40 billion per annum, a huge amount to index investors.

In Figure 11, we look at the dependence of index concentration on the diversity weight exponent *p*. The figure shows this dependence for the S&P 500 index on December 31, 1993, when the market was relatively diversified, and on December 31, 2021, when the market was relatively concentrated.

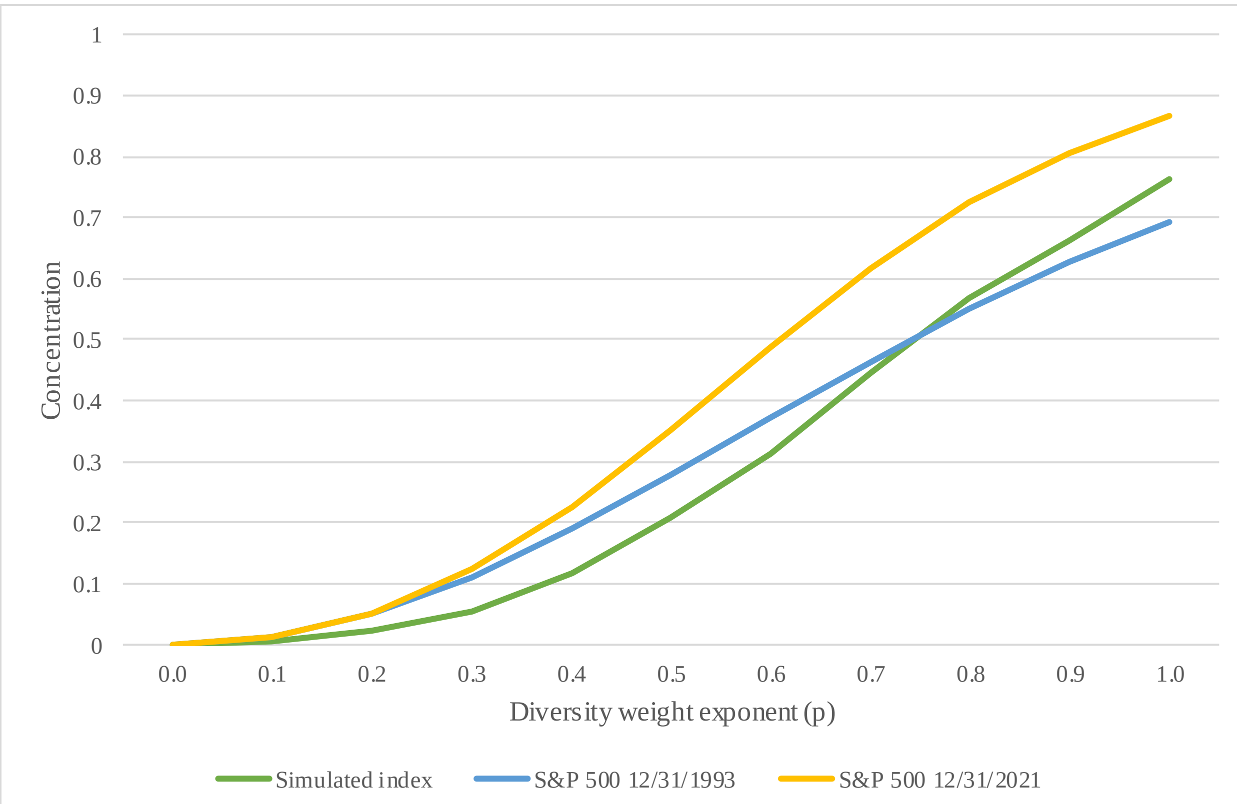

Figure 11 Dependence of concentration on diversity weight exponent for the S&P 500 index on two dates and a simulated index

Source: Authors' computation and market cap and index holdings data from MSCI.

As we vary the diversity weight exponent p, we obtain a *concentration curve*, which shows how the concentration of a diversity weighted index varies for a fixed set of market caps as p ranges between 0 and 1. We show the curve for the cap weights taken from the median path of our experiment along with two curves based on empirical cap weights of the S&P 500 index on two different dates. Moving from equal weights (p = 0) to cap weights ($p = 1$), concentration emerges more slowly for the median simulated index than the empirical indexes. For cap weights, concentration of the median simulated index, 0.76, lies in between the December 1993 value of 0.69 and the December 2021 value of 0.87.

Finally, we look at the time it takes to go from a starting distribution to a steady-state distribution of market caps in our simulation. We run three experiments distinguished by assumption on initial state. The simplest is that all firms begin at zero growth, which corresponds to an equally weighted initial distribution of normalized market caps. We also use trailing one-year growth rates for firms in the S&P 500 as of December 31,1993, when the market was relatively diversified, and as of December 31, 3021, when the market was relatively concentrated. The median time to steady state decreased from 550 months to 530 months to 496 months for the three cases. In other words, the time to a steady-state power law was 40–45 years in our simulation.

V. Conclusion

Market-capitalization weighting is a traditional and popular approach to constructing indexes. Developed initially for performance measurement, market-cap-weighted indexes are common benchmarks for equity index–tracking funds and ETFs. The relatively low turnover in cap-weighted indexes together with their association with the market portfolio are important rationales for their widespread adoption. However, there are increased concerns that some cap-weighted indexes may suffer from overly concentrated positions in a few stocks, limiting diversification and exposing investors to the risks that some of the largest holdings are overpriced.

We find that some, but not all, major indexes have seen increases in concentration in the past quarter century. To put this in context, we examine the historical concentration of major equity indexes. It turns out that concentration varied dramatically across indexes, and the current levels are not unusual by historical standards. For example, the S&P 500 index was more concentrated in December 1975 than in December 2021.

The academic literature on concentrated phenomena, including distributions of words, sizes of cities, and other areas in which the largest players dwarf the smaller ones offers several plausible mechanisms that may help to explain concentration in firm size and its evolution over time. These mechanisms suggest that the eventual distribution of cap weights in a market or index may be approximated by power laws in the tails. We generally reject the Zipf distribution for current indexes, meaning we are far from the possible extremes of concentration suggested by theory, particularly for the largest constituents.

To assess both the efficacy of a basic theory that predicts concentration and the cost of concentration mitigation schemes, we develop a simulation based on a geometric Brownian motion with a reflecting barrier. This model fits empirical data remarkably well, suggesting that market concentration may, to some degree, be the inevitable consequence of dynamic forces. The framework has several practical uses too. For example, asset managers may use this approach to assess the trade-off between lower concentration and higher turnover. We find that concentration mitigation schemes are effective, but they generate additional turnover and may reduce investors' returns relative to cap weighting, with higher transaction costs. Our framework offers a potential to gauge the relative merits of these mitigation efforts. As usual, there is no free lunch, and we conclude that market-cap weighting is not likely to be displaced in the near term.

References

Amenc, N., F. Goltz, A. Lodh, and L. Martellini. (2012). "Diversifying the Diversifiers and Tracking the Tracking Error: Outperforming Cap-Weighted Indices with Limited Risk of Underperformance," *Journal of Portfolio Management* **38**, 72–88.

Amenc, N., F. Goltz, L. Martellini, and P. Retkowsky. (2011). "Efficient Indexation: An Alternative to Cap-Weighted Indices," *Journal of Investment Management* **9**, 1–23.

Arnott, R. D., and J. C. Hsu. (2008). "Noise, CAPM and the Size and Value Effects," *Journal of Investment Management* **6**, 1–11.

Arnott, R. D., J. C. Hsu, and J. West. (2008). *The Fundamental Index: A Better Way to Invest.* Hoboken: John Wiley & Sons, Inc.

Axtell, R. (2001). "Zipf Distribution of US Firm Sizes," *Science* **293**, 1818–20.

Bernstein, P. (2003). "Points of Inflection: Investment Management Tomorrow." *Financial Analysts Journal* **59**, 18-23

Bohn, J., L. R. Goldberg, and S. Ulucam. (2022). "Sustainable Investing from a Practitioner's Viewpoint: What's in Your ESG Portfolio?" *Journal of Investment Management*, forthcoming.

Booth, D. G., and E. F. Fama. (1992). "Diversification Returns and Asset Contributions," *Financial Analysts Journal* **48**, 26–32.

Bouchaud, J-P., Farmer, J. D., and F. Lillo. (2009). "How Markets Slowly Digest Changes in Supply and Demand." In *Handbook of Financial Markets: Dynamics and Evolution*, eds. K. R. Schenk-Hoppé and T. Hens, 57–160. Amsterdam: North Holland Publishing Co.

Chow, T., J. C. Hsu, V. Kalesnik, and B. Little. (2011). "A Survey of Alternative Equity Index Strategies," *Financial Analysts Journal* **67**, 37–57.

Debru, P. (2020). "S&P 500 or S&P 5? Is the Current Concentration Sustainable Long Term?" ETF Stream, November 5. Available at: https://www.etfstream.com/features/sp-500-or-sp-5-is-the-current-concentration-sustainable-long-term/.

DeMiguel, V., L. Garlappi, and R. Uppal. (2009). "Optimal Versus Naive Diversification: How Inefficient Is the 1/*N* Portfolio Strategy?" *Review of Financial Studies* **22**, 1915–53.

Embrechts, P., C. Klüppelberg, and T. Mikosch. (1997). *Modelling Extremal Events: for Insurance and Finance*. Berlin, Heidelberg: Springer-Verlag.

Fernholz, R. (1999). "On the Diversity of Equity Markets," *Journal of Mathematical Economics* **31**, 393–417.

Fernholz, R. (2002). *Stochastic Portfolio Theory*. New York: Springer-Verlag.

Fernholz, R. (2005). “Stock Market Diversity,” INTECH Working Paper.

Fernholz, R., R. Garvy, and J. Hannon. (1998). “Diversity-Weighted Indexing,” *Journal of Portfolio Management* **24**, 74–82.

Fernholz, R., and C. Maguire Jr. (2007). “The Statistics of Statistical Arbitrage,” *Financial Analysts Journal* **63**, 46–52.

Fernholz, R. T., and R. Fernholz. (2020). “Zipf’s Law for Atlas Models,” *Journal of Applied Probability* **57**, 1276–97.

Gabaix, X. (1999). “Zipf’s Law for Cities: An Explanation,” *Quarterly Journal of Economics* **114**, 739–67.

Gabaix, X. (2009). “Power Laws in Economics and Finance,” *Annual Review of Economics* **1**, 55–93.

Gabaix, X., and R. Ibragimov. (2011). “Rank – 1/2: A Simple Way to Improve the OLS Estimation of Tail Exponents,” *Journal of Business Economics and Statistics* **29**, 24–39.

Gibrat, R. (1931). *Les inégalités économiques*. Paris: Recueil Sirey.

Goldberg, L. R., G. Miller, and J. Weinstein. (2008). “Beyond Value at Risk: Forecasting Portfolio Loss at Multiple Horizons,” *Journal of Investment Management* **6**, 73–98.

Goltz, F., and V. Le Sourd. (2010). “Does Finance Theory Make the Case for Capitalization-Weighted Indexing?” EDHEC-Risk Institute.

Graham, J. (2012). “Comment on the Theoretical and Empirical Evidence of Fundamental Indexing,” *Journal of Investment Management*, First Quarter.

Grinold, R. (1992). “Are Benchmark Portfolios Efficient?” *Journal of Portfolio Management* **19**, 34–40.

Harrison, J. M. (2013). *Brownian Models of Performance and Control*. New York: Cambridge University Press.

Haugen, R., and N. Baker. (1991). “The Efficient Market Inefficiency of Capitalization-Weighted Stock Portfolios,” *Journal of Portfolio Management* **17**, 35–40.

Hsu, J. C. (2006). “Cap-Weighted Portfolios Are Suboptimal Portfolios,” *Journal of Investment Management* **4**, 1–10.

Jiang, H., D. Vayanos, and L. Zheng. (2020). “Tracking Biased Weights: Asset Pricing Implications of Value-Weighted Indexing,” NBER Working Paper No. 28253. DOI 10.3386/w28253.

Levy, M. (2016). “It’s Easy to Beat the Market,” *Journal of Investment Management* **14**, 38–50.

Madhavan, A. (2016). *Exchange-Traded Funds and the New Dynamics of Investing*. New York: Oxford University Press.

Madhavan, A., A. Sobczyk, and A. Ang. (2018). "What's in Your Benchmark? A Factor Analysis of Major Market Indexes," *Journal of Portfolio Management* **44**, 46–59.

Malevergne, Y., P. Santa-Clara, and D. Sornette. (2009). "Professor Zipf Goes to Wall Street," National Bureau of Economic Research Working Paper.

Mandelbrot, B. (1960). ''The Pareto-Levy Law and the Distribution of Income,'' *International Economic Review*, 1: 79–106.

Morningstar (2020). "The S&P 500 Grows Ever More Concentrated." Available at https://www.morningstar.com/articles/992504/the-sp-500-grows-ever-more-concentrated.

Pareto, V. (1896). *Cours d'économie politique*. Geneva: Droz.

Perold, A. F. (2007). "Fundamentally Flawed Indexing," *Financial Analysts Journal* **63**, 31–37.

Robertson, A. Z. (2019). "Passive in Name Only: Delegated Management and 'Index' Investing," *Yale Journal on Regulation* **36**, 795–851.

Roll, R. (1977). "A Critique of the Asset Pricing Theory's Tests Part I: On Past and Potential Testability of a Theory," *Journal of Financial Economics* **4**, 129–76.

Saichev, A., Y. Malevergne, and D. Sornette. (2010). *Theory of Zipf's Law and Beyond*. Berlin, Heidelberg: Springer-Verlag.

Santilli, P. (2020). "Tesla Stock Joins the S&P 500: A Game Changer," *Wall Street Journal*, December 21, 2020.

Shankar, S., J. Miller, and P. Balakrishnan. (2015). "Concentration of Trading in S&P 500 Stocks," http://dx.doi.org/10.2139/ssrn.2742363.

Sharpe, W. F. (1964). "Capital Asset Prices: A Theory of Market Equilibrium under Conditions of Risk," *Journal of Finance* **19**, 425–41.

Strongin, S., M. Petsch, and G. Sharenow. (2000). "Beating Benchmarks," *Journal of Portfolio Management* **26**, 11–27.

Tabner, I. (2007). "Benchmark Concentration: Capitalization Weights versus Equal Weights in the FTSE 100 Index," University of Stirling Working Paper.

Zipf, G. (1949). *Human Behavior and the Principle of Least Effort: An Introduction to Human Ecology*. Cambridge: Addison-Wesley.

Endnotes

[1] See, e.g., Madhavan (2016). A prominent exception is the Dow Jones Industrial Average, which is price weighted.

[2] Rebalancing refers to adjusting index weights as specified by the index provider to reflect corporate actions such as mergers, spin-offs, etc. Reconstitution refers to adding or deleting securities from an index based on the index criteria.

[3] See, e.g., Morningstar (2020) at https://www.morningstar.com/articles/992504/the-sp-500-grows-ever-more-concentrated. Debru (2020) notes that as of November 2020, the five biggest stocks in the S&P 500 represented 21% of the index, almost double the long-term average of 12.5%. Academic concerns go back much further, as in Strongin et al. (2000), Bernstein (2003), Tabner (2007), and Malevergne et al. (2009).

[4] Roll (1977) famously questions the common use of cap-weighted indexes as a proxy for the market portfolio. The point is carried further in Bohn et al. (2022), who document a large-cap bias in cap-weighted subportfolios of market indexes. Using the methodology from their paper, they find the largest decile of securities has more than 90% of the active weight of the S&P 500 Index against the Russell 1000 Index in an unpublished calculation.

[5] See, e.g., Amenc et al. (2011) for a review. See also Haugen and Baker (1991), Grinold (1992), Amenc et al. (2012), Tabner (2007), and Malevergne et al. (2009), among others. Levy (2016) compares the market with a large number of randomly constructed and passively held portfolios and finds that 69% of these random portfolios yield higher Sharpe ratios than the market.

[6] Newer, factor, and thematic indexes are not just performance benchmarks but are designed also to be investible. Such indexes may be non–market cap weighted. Robertson (2019) finds a recent proliferation of indexes. She documents "substantial heterogeneity" across indexes and concludes that the overwhelming majority of the indexes in her sample "are used as a primary benchmark by only a single fund."

[7] In the factor space, concentration has grown larger, and investors need to be conscious of unintended factor allocations arising from following common indexes. Madhavan et al. (2018) show there is large time-series variation in the exposures of common indexes, implying that the implicit factor loadings of indexes are time varying. Their analysis breaks down the style factor exposure of each benchmark index using stock-by-stock style scores.

[8] The breadth of a portfolio or an index can be thought of as its effective number of securities. Consider three portfolios consisting of two stocks each, with weight distributions of 0.5–0.5, 0.8–0.2 and 0.99–0.01. The breadths of these portfolios are, respectively, 2.00, 1.47, and 1.02.

[9] See, e.g., Santilli (2020).

[10] We use the actual iShares exchange-traded funds (ETFs) that seek to track these indexes.

[11] Power laws are possibly so prevalent across seemingly diverse applications because of their aggregation properties. The sum of two (independent) power law distributions yields another

power law distribution. The product of two power laws or their max (or min) also yields a power law distribution.

[12] For applications in economics/finance, see Axtell (2001), Bouchard et al. (2009), Gabaix (1999, 2009), Saichev et al. (2010), and Fernholz and Fernholz (2020), among many others. Shankar et al. (2015) use power laws to analyze changes in the distribution of trading volumes of S&P 500 stocks and non-S&P 500 stocks over time.

[13] Zipf's law has been used in index construction in the crypto space. The T3 Crypto-X Power Index, initiated in January 2014, seeks to represent a value of investment in a portfolio of top ten cryptocurrencies by market capitalization with weighting determined by Zipf's law chosen as "a middle ground between investing according to market capitalization, which heavily prioritizes Bitcoin and a few other major cryptocurrencies, and equal distribution which puts too high of a risk on less established assets." Source: https://t3index.com/indices/crypto-xp/.

[14] We take α as the negative of the estimated slope coefficient by convention.

[15] In the original 1999 study, data for only 135 metro regions was available. Today, the US data covers more than 400 cities.

[16] In the language of Fernholz and Fernholz (2020), the plots in Figure 5 are "quasi-Zipfian" and can be fit to rank-based Atlas models.

[17] Examples of models that fit a power law only to the tail of an empirical distribution are found, for example, in Embrechts et al (1997) and Goldberg et al. 2008.

[18] Gibrat (1931) proposes proportional random growth as the source of power law distributions. Other explanations include the transfer of that power law through matching and optimization and network effects.

[19] Gabaix (1999) provides the necessary conditions for the existence of a stable, time-invariant distribution.

[20] A Zipf distribution may also arise through optimization. In designing a language, for example, very long words would be rare while very common words would be short. This may also arise through chance, as for example, with monkeys typing on keyboards with a space marking the end of a word.

[21] A discussion of equally weighted strategies is in DeMiguel et al. (2009).

[22] For example, at the end of 2020, the true breadth of the S&P 500 was 72.5 vs. a diversity score of 327.5 with $p = 0.5$ using equation (15).

[23] See Harrison (2013), Chapter 1.

[24] Lowering the product NS_{min}, by lowering either the number of firms in the simulation or the reflecting barrier, drives the power law exponent closer to 1, increasing instability in estimates of concentration and turnover.

[25] We omit results for MSCI 25/50 rule and the UCITS 5/10/40 rule, which were rarely binding.

[26] While the median power law of the simulated cap-weighted index of 500 firms was 1.51, the corresponding exponent for the simulated full universe of 4000 firms was 1.54, consistent with theory.