Index concentration and rich-get-richer phenomenon using reinforced urn processes

Author:
Nossaiba Kheiri

Supervisor:
Prof. Lisa Goldberg

March 25, 2023

A thesis submitted in fulfillment of the requirements
for the degree of Bachelor of Science

in the

Department of Economics at the University of California,Berkeley

Joint work with Prof. Lisa Goldberg, Prof. Alex Shkolnik and Harrison Selwitz.

Statement of Academic Integrity

I, the undersigned Nossaiba Kheiri, hereby certify on my honor that:

- The results presented in this report are the product of my own work.
- I am the original creator of this report.
- Where I have consulted the published work of others, this is always clearly attributed.
- I have not used sources or results from third parties without clearly stating thus and referencing them according to the recommended rules for providing bibliographic information.

I hereby declare that this work contains no plagiarized material.

Signature: Nossaiba Kheiri

Date and Place: 19/03/2023 - Paris, France

ÉCOLE POLYTECHNIQUE

Abstract

Department of Economics at the University of California,Berkeley

Bachelor of Science

Index concentration and rich-get-richer phenomenon using reinforced urn processes

Joint work with Prof.Lisa Goldberg, Prof.Alex Shkolnik and Harrison Selwitz.

This paper attempts to understand the mechanisms behind index concentration using power laws as a measure of index concentration. Our study is based on (Goldberg et al., 2022) where reflecting geometric Brownian motion is used to model the observed concentration. We introduce a reinforced urn processes **(RUP)** approach that may generate hypothetical indexes matching empirical market cap-weighted indexes. We present two narratives that might explain the **rich-get-richer** phenomenon observed for market cap weighted indexes: Born rich, get richer, and get richer with help along the way. The Barabasi Albert model that we extend using a Polya urn approach produces hypothetical indexes that match empirical market cap-weighted indexes in important ways.

Contents

List of Figures

Introduction

Indexes are a highly influential benchmark that shapes decision-making. In particular, financial indexes receive a wide attention and convey information about markets as a benchmark of performance and funds tracking. A vast portion of equity indexes is weighted according to the size of its total market capitalization.

However market-cap weighted indexes [1] tend to be concentrated in the largest stocks.Indeed, (Goldberg et al., 2022) shows empirical evidence on concentration by studying the time series of concentration over the past 25 years of four major equity indexes.Tracking index concentration has gained a wider importance with the concentration increase in financial indexes present in the $S\&P$ 500 for example,giving rise to inequalities characterized by **rich-get- richer** dynamics.
"The five biggest stocks (in the $S\&P$ 500) had never represented more than 18.2%, last seen at the height of the 'dot-com' bubble in March 2000. Currently Apple, Microsoft, Amazon, Facebook and Alphabet represent 21% of the index, almost double the long-term average of 12.5%".(Debru, 2020)

Concentration may arise for multiple reasons, including differences in market capitalization, industry dominance, and investor sentiment... Indeed, particular industries or companies that perform well (rapid growth and increased investor demand) may become over-represented in the index. And has important implications for investors, market efficiency, and stability.

On one hand, a highly concentrated index can have the potential for higher yield for investors who hold the dominant stocks if those top-performing stocks continue to perform well. On the other hand, it increases the risk of the index due to a highly plausible downturn or negative shock due to reduced diversification of the index.

[1] Market cap weighted indexes refer to indexes where the weight of a stock listed in this index is determined by the **the market capitalization** (the market value of a company's stocks traded on the stock market) of the stock.

How do these indexes evolve? What might drive this concentration we observe? While each day new stocks are listed why do few stocks achieve wider importance? Understanding the causes and consequences of concentration is a complex problem.

Random processes can generate hypothetical indexes that mimic the concentration in actual indexes. Reinforced urns processes offer an algorithmic approach to understanding how concentration arises in financial indexes, by modeling the complex interactions between individual securities and the larger index and capturing the feedback loops that can reinforce or reduce concentration over time.

Our paper aims to provide alternative explanations of the observed concentration in indexes. We attempt to explain the concentration phenomenon using random processes that arise naturally in various contexts and are used to model the internet, trade, political persuasion... Moreover, the paper contributes to the literature by suggesting models that can be used to test different policy interventions and market structures, to see how they might affect concentration levels and dynamics. This approach can help policymakers and investors to better understand the risks and opportunities associated with concentration, and to design strategies that promote greater market efficiency and stability.

In section 1 we give a general overview of the economics of index concentration and the rich get richer phenomenon and its relation to power laws and other concentration measures. We then present in section 2 reinforced urn processes. Section 3, presents different Polya urn models that we use to understand the concentration phenomenon. The urn setting creates systems that imitate the real world to achieve coordination over time which translates into limiting behavior that the system reaches as a result of micro-level interaction rules. Such systems are referred to in the literature as **self-organizing systems** or **lock-in** phenomenon in industrial and consumer behavior.

In section 4, we introduce the Barabási–Albert model, a network science algorithm, that we adapt to concentration modeling using an analogy to Polya urns. In the last section, we conclude the results of our paper.

Discussion of the approach

0.1 The economics of index concentration - A rich-get-richer phenomenon

The **Normal distribution** is observed in many systems in reality and has been a natural "guess" for many models. However, many extreme events arise in real life which suggest magnitudes higher than the ones suggested by the normality assumptions, for example in economics and finance.Extreme events arise in the economy such as inflation, massive layovers in tech " the Quant Quake of 2007, Global Financial Crisis of 2009, Flash Crash of 2010, Covid Crisis of 2020..." (Goldberg et al., 2022).

The economic theory presents the **the lock-in effect** as a plausible explanation of the extreme imbalances observed in reality or constant **popularity** of few firms.(Eurich,Burtsher 2014), For example, if most people use a particular social media platform, it becomes increasingly difficult for new players to enter the market... In other words, initial advantages can lead to cumulative advantages over time, making it increasingly difficult to catch up. This effect is closely related to rich get richer models.

The economist W. Brian Arthur proposed urn models to model firms' concentration. While using simple urn models, His conclusions resonated with the economics community.

Measures of concentration

It is important to look at ways to quantify concentration. Our study is based on (Goldberg et al., 2022) where they use reflected geometric Brownian motion to model the observed concentration. The paper also looked at ways to mitigate the concentration using Monte Carlo simulations calibrated to market data. In this study, we present other mechanisms through reinforced urn processes which can also generate concentration.

The (Goldberg et al., 2022) paper presents the Herfindahl-Hirschman Index (HHI) index amongst many other measures of concentration. Although some are more intuitive measures of index concentration than others.We use power laws as a measure of concentration to coincide with literature on stochastic processes.

0.1.1 Power law and Regression

Zipf's law

Zipf's law was first formulated by George Zipf, an American linguist. Zipf found that the "rank(i)- frequency (S_i) [2]relationship" quantifies as

$$S_i \sim \frac{1}{i}$$

Zipf's law also implies that the probability that the value S is greater than some fixed numbers decreases as s increases.

$$P(s) = Pr(S > s) \sim \frac{1}{i}$$

Beyond word frequency in a language, power law appears in other fields such as city populations, earthquake magnitude, income distribution, the number of citations of scientific publications, sociology, economics, and computer science. It is a powerful tool for modeling and analyzing large-scale patterns in complex systems.

Let S_i be the frequency of the ith ranked element. In other words, the share of that item of the ith element represents its "market share".

We present below the general form of a power law equation for the random variables S_i and i, where $\zeta > 0$ is a constant and α is the power law coefficient which can be used as a measure of index concentration.

$$S_i \sim \zeta i^{-\alpha}$$

[2]the relative frequency of words in natural languages.

This refers to the fact that the market share S_i of some stochastic variable, usually a size or frequency, is greater than i its rank , and decays with the growth of i.

We perform a linear regression of log-rank on log size to examine the emergence of a power law in the models we present in this paper.

$$ln(i) = ln(\zeta) - \alpha \cdot ln(S_i) \tag{1}$$

Data overview

The study of 4 of the most popular financial indexes ($S\&P500$, Russel 2000, MSCI EAFE, MSCI EM) in (Goldberg et al., 2022) , using empirical evidence from 1996-2021, suggests that the $S\&P500$ is well fit with a power law (including the tails) the most , amongst the 4 financial indexes, with a power law coefficient lying in the confidence interval.

We restrict this study to the $S\&P500$. We attempt to understand the mechanisms that may govern index concentration. The market shares of each stock in the $S\&P500$ are sampled from a BlackRock dataset , based on index holdings data from MSCI.

0.2 Reinforced urn processes (RUP)

In this section, we present the research design we use to study index concentration. We start by presenting the theory behind reinforcement processes. Reinforced urn processes **(RUP)** give a perspective on how extreme imbalances might arise. The urn scheme, which was first presented by Eggenberger and Polya (1923) to describe contagious diseases, quickly evolved into the template for probabilistic models used in numerous application areas.

In this paper, we present different reinforcement models that we use to understand the concentration phenomenon. The urn setting creates systems that imitate the real world to achieve coordination over time which translates into limiting behavior that the system reaches as a result of micro-level interaction rules. Such systems are referred to in the literature as **self-organizing systems** or **lock-in** phenomenon in industrial and consumer behavior.

0.3 A Pòlya urn model to model index concentration

We consider the **generalized Pòlya urn scheme (GPU)** presented in (Pemantle, 2007). The urn initially contains $k \geq 2$ types of balls (colors, types of stocks...). At each iteration, a ball of type i is sampled from the urn at random and replaced along with A_{ij} balls of type j, with $A_{ij} \geq \delta_{ij}$, in other words: for$j \neq i : A_{ij} \geq 0$ and for $j = i : A_{ij} \geq -1$.

It comes quite natural to consider only non-negative reinforcement values for $j \neq i$ to ensure the non-negativity of the number of balls of type j at the end of each iteration. And allow values in [-1,0) for the same type of reinforcement where the ball drawn can be not entirely replaced.

We thus have the following k by k **reinforcement matrix**:

$$A = \begin{pmatrix} A_{1,1} & A_{1,2} & \cdots & A_{1,k} \\ A_{2,1} & A_{2,2} & \cdots & A_{2,k} \\ \vdots & \vdots & \ddots & \vdots \\ A_{k,1} & A_{k,2} & \cdots & A_{k,k} \end{pmatrix}$$

We suppose that all the types (stocks, securities) enter the market at roughly the same time n=0. Let $R_{j(n)}$ be the number of balls of color j at time n with $R_{j(0)}$ the initial allocation of type j balls present initially in the urn. And let $X_{nj} = \frac{R_{j(n)}}{\sum_{j=1}^{k} R_{j(n)}}$ which represents the 'market' share of type j. Such that : $R_{j(n+1)} = R_{j(n)} + A_{ij}$ for $n \geq 0$ and $0 \leq i, j \leq k$.

In this paper, we analyze the dynamics of the evolution of the **'market' share vector** X_n for each type in the urn which we consider in our analysis as replicating the behavior of the stock market.

Rich-get richer: Gets richer with help along the way...

In this section, we present an alternative plausible causal mechanism that might be useful to understand how concentration emerges in the largest securities for financial indexes.

Consider the generalized Pòlya urn scheme (GPU) with **strictly positive reinforcement matrix**[3] $A_{ij} > 0$ (also the balls non-drawn get reinforced according to the type of the drawn ball), the limiting distribution of the market share vector X_n is deterministic regardless of the initial distribution as shown by (Pemantle, 2007) [Theorem 2.3].

Theorem 2.3 (Pemantle, 2007): *In a GPU with all $A_{ij} > 0$, the vector X_n converges almost surely to π, where π is the* ***unique positive left eigenvector***[4] *of A normalized by $|\pi| := \sum_{i=1} \pi_i = 1$.*

Thus, the limiting distribution of X_n is determined by the unique positive left eigenvector of the reinforcement matrix. The eigenvector is normalized to be in the unit sphere to ensure that π is a probability measure. The existence of such an eigenvector is guaranteed by the **Perron-Frobenius Theorem**. The Perron-Frobenius theory shows that a positive real square matrix has a unique largest real eigenvalue whose eigenvector can be chosen to have strictly positive coordinates.

In this case, the model is **deterministic, regardless of the initial distribution of balls**, and the Perron-Frobenius theorem applies. One can relate this result to the scenario where firms grow with help along the way, this help outweighs the effect of the initial allocation.

Simulation

In this section, we present a simulation of the model described in the previous section and evaluate the result using empirical data from the $S\&P500$.

Reverse engineering the reinforcement matrix

As presented in the previous section, the reinforcement matrix must be strictly positive. This condition is ensured through the following construction.

[3] A positive matrix A is a real or integer matrix $(A)_{ij}$ for which each matrix element is a positive number, i.e., $A_{ij} > 0$ for all i, j.

[4] A left eigenvector is defined as a row vector X_L satisfying $X_L * A = \lambda_L * X_L$

We construct the reinforcement matrix 'artificially' given data from the current distribution of the stocks in the index. The simulation is then run to check if the simulated indexes of the algorithm matches the weights of the actual indexes. Given market capitalization distribution, we work backward to construct a reinforcement matrix whose normalized unique positive left eigenvector is the market capitalization distribution as predicted by **Theorem 2.3 (Pemantle, 2007)**.

One way to get a reinforcement matrix starting from the final distribution vector of market shares of each stock in the index π_{data} is the following construction which ensures the positivity of the matrix A since π_{data} is also positive (a 'market share' vector):

$$A = \pi_{data} \otimes \pi_{data} = \pi_{data} \pi_{data}^T \tag{2}$$

One can check that the row vector π_{data}^T is a left eigenvector of A. Indeed,
$\pi_{data}^T \cdot A = \pi_{data}^T \cdot \pi_{data} \cdot \pi_{data}^T = \langle \pi_{data}, \pi_{data} \rangle \cdot \pi_{data}^T = \|\pi_{data}\|^2 \cdot \pi_{data}^T$.

Using the construction above, we simulate the Pòlya urn model with uniform initial allocation in 10000 iterations ($n = 10000$) to replicate the limiting behavior described in the theorem above. Using the regression approach of equation 1, we examine whether power laws capture the concentration for the $S\&P500$.

Figure 1 shows the weight (market share or market cap weight) of each security present in the $S\&P500$ as of February 2023 taken from both the Pòlya urn simulation and the actual market shares (market-cap weights) of the $S\&P500$ data.We rank the stocks from 1 (stock with the highest market share) to roughly more 500 [5] (stock with the lowest market cap weight included in the $S\&P500$ index).

As figure 1 shows, the market share concentrates in the stocks with the lowest ranks. As the rank gets higher The Pólya urn algorithm with the "reverse engineered" reinforcement matrix gives a distribution of market shares close to the actual $S\&P500$.

Figure 2 shows the log-rank log size (log weight) regression of both the simulated and the actual market shares as described in equation 1. The power law fit coefficient (1.33) is taken from the analysis in (Goldberg et al., 2022) ($R^2 = 0.99$).There is a linear

[5]the number of stocks in the $S\&P500$ vary around 500 because of entries and exits of firms.

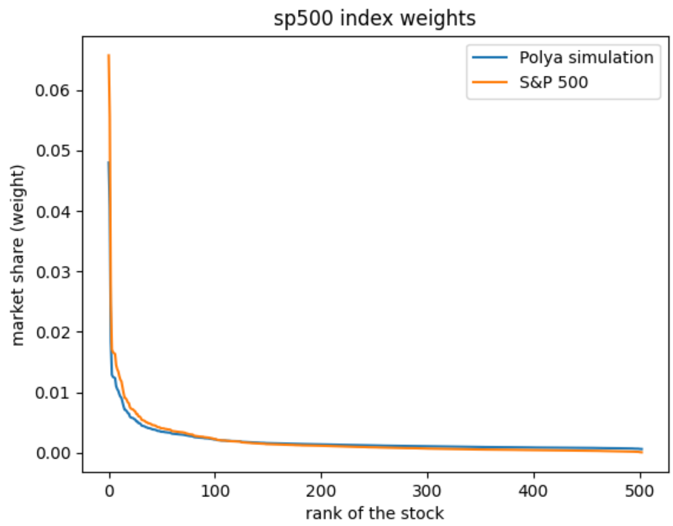

FIGURE 1: $S\&P500$ index market share predicted by the Pòlya model and actual market shares as of February 2023

fit to the elements with medium to small rank (closer to the tail) with high values of size (weight) in which the index is concentrated. Overall, we observe a concave relationship between log rank-log weight. Figure 3, represents the evolution of the sum of square residuals (SSR) of the Pòlya urn model market cap weight compared to the empirical $S\&P500$ data.For the tail values, the Pòlya urn simulation represents a slightly larger SSR compared to the values of higher rank stocks. However, the model overall has a high R^2 which indicates a satisfactory explanatory power of the model.

In practice, we do not have a priori the final distribution vector of market shares of each stock in the index π_{data}.

One way to approach this limitation would be to continuously update the reinforcement matrix by using previous realizations of the vector of market shares of each stock in the index π_{data} chosen while trying to match the current situation of the economy (recession, expansion...).

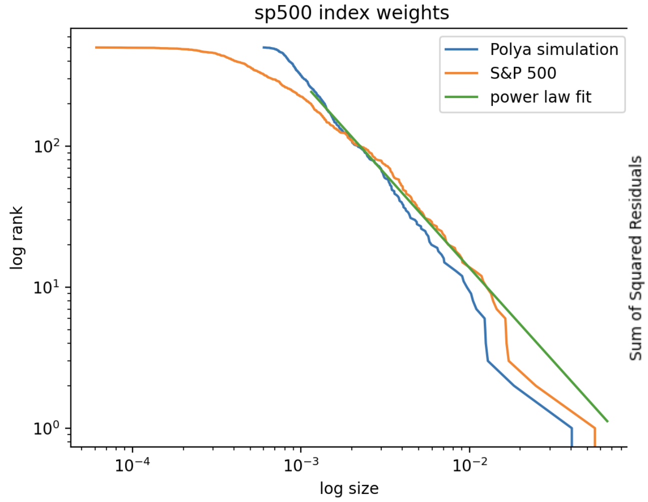

FIGURE 2: Log rank-Log size regression of the $S\&P500$ of both the simulated and the actual market shares

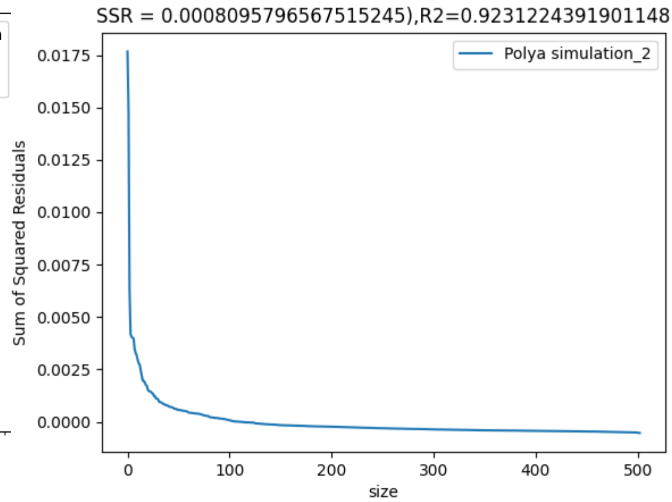

FIGURE 3: Evolution of the residuals for the Pòlya urn 2 model

Rich-get richer: Born rich, gets richer...

(Pemantle, 2007) showed that for a Pòlya urn with initial allocation $(R_{1(0)}, R_{2(0)})$ and identity reinforcement (i.e At each iteration, a ball is chosen randomly, returned to the urn along with one extra ball of the color drawn), **limiting distribution of X_n** is $\beta(R_{1(0)}, R_{2(0)})$ almost surely.[Theorem 2.1] In particular, when $R_{1(0)} = R_{2(0)} = 1$ variable X_n converges to the uniform distribution on [0,1]. (Blackwell and MacQueen, 1973) generalises this theorem to k-types Pòlya urn with initial allocation $R_1(0), \ldots, R_k(0)$.

The study proves that the limiting distribution is a **Dirichlet distribution**[6] with parameters $(R_1(0), \ldots, R_k(0))$, where the Dirichlet distribution with parameters $(R_1(0), \ldots, R_k(0))$ is defined to be the measure on the $(k-1)$-simplex with density:

$$\frac{\Gamma\left(R_1(0)+\cdots+R_k(0)\right)}{\Gamma\left(R_1(0)\right)\cdots\Gamma\left(R_d(0)\right)} \prod_{j=1}^{k} x_j^{R_j(0)-1} k x_1 \cdots k x_{k-1}$$

The independence between the types is guaranteed by construction (of the Pòlya urn). That is the joint distribution of $X = (X_1(n), \ldots, X_k(n))$ as $n \to \infty$ is Dirichlet as defined above. Such a result gives a theoretical characterization of the limiting

[6] $Dir(\alpha)$ a family of continuous multivariate probability distributions with real positive parameter α. It is a generalization of the beta distribution for the multi-variable case

behavior of the weights (market shares) in a market-cap weighted index (distribution, moments...), which we can view as another way to view the dynamics giving rise to index concentration through random processes using reinforced urns.

The initial allocation determines the limiting distribution. For the equal initial distribution case, we get a uniform distribution of the limiting distribution. As the number of types grows, the distribution becomes more skewed to the type with higher initial allocation.

A ball of the same type as the drawn ball is added in each step. As soon as one type dominates, this type is likely to dominate forever, it will grow stronger and stronger. The beginning of the process is critical to its long-term evolution. This result aligns with our expectations and intuitive understanding of the problem. Indeed, types starting with a larger initial allocation have a higher likelihood of generating greater output over time, they are 'intended' to succeed by birth. In contrast, types having a low initial allocation which might get absorbed in the long term with a lower share of the market. - "**Born rich,gets richer**" narrative.

Simulation

We now present the simulation of this model. Suppose that the initial allocation is randomly drawn from the universe of possible initial allocations.

We simulate multiple paths (states of the world) and plot the average (mean) market share along each type (in our case index). One may also try to use the median instead of the mean however it might be problematic if there is a high asymmetry in the initial allocation which is highly plausible in reality. Although the simulation poorly fits the $S\&P500$ data, it gives insight into the effect of initial allocation on changing the curvature of the simulation result to better-fit data. One major drawback of this simulation is that the initial allocation was drawn randomly. We might approach this limitation by Optimal transport or computational methods to retrieve the parameters of a Dirichlet distribution, such as maximizing the log-likelihood with gradient ascent, Newton-Raphson, and Fixed point iteration. This remark might be exploited in further studies. Also, the problem of constructing a distribution from the empirical data should be taken into consideration.

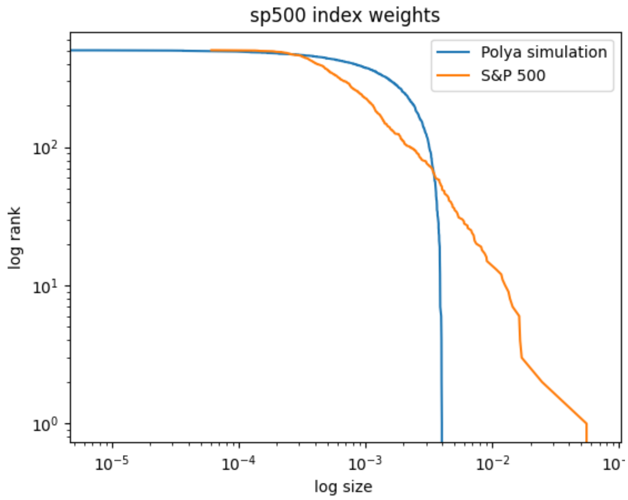

FIGURE 4: Log rank-Log size regression of the $S\&P500$ of both the simulated and the actual market shares

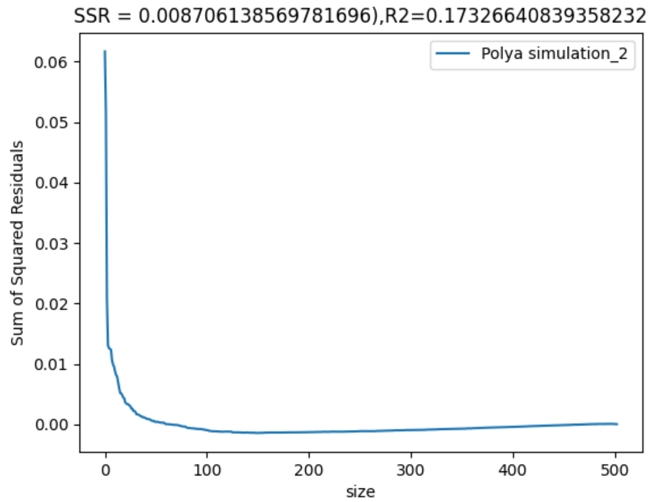

FIGURE 5: Evolution of the residuals for the Pòlya urn model

0.4 Barabási–Albert to model index concentration

In this section, we present the **Barabási–Albert model** which is an algorithm to generate power laws (scale-free networks) using preferential attachment[7]. We present an analogy of the Barabási-Albert model to the approach taken in this study to model concentration using Polya urns.[8]One interesting feature of using this model to study index concentration is that it provides an alternative explanatory mechanism without having prior information about the initial allocation (as in section 2) or to artificially 'reverse engineer' the reinforcement matrix (as in section 1).

Barabási–Albert model

The Barabási-Albert model is widely studied in **network science** explaining how complex networks could emerge from simple growth rules.

(Barabási and Albert-László, 1999) was first proposed by Albert-László Barabási and Réka Albert to study web networks. The model is now a cornerstone of network theory, giving insight into networks with applications in social networks, disease spread, information diffusion, internet ...

[7]nodes prefer to link to highly connected nodes.

[8]The analogy noticed and proposed by the thesis advisors Prof.Alexander Shkolnik and Prof.Lisa Goldberg.

At the time $k = 0$, we start with a **seed network** consisting of a fixed number of nodes n. Then, new nodes are added to the network one by one. Each new node picks exactly one of the existing nodes, with probability proportional to their **degree** (popularity) which is measured by the number of nodes they are connected to.

The **rich-get-richer phenomenon** also emerges in this network where the highly connected nodes tend to continue to acquire more connections and become hubs due to the preferential attachment property ultimately leading to the **power law distribution**. Another remarkable phenomenon that appears is what is known in the literature (Newman, 2009) as the **first-mover advantage**. Nodes that are added earlier have higher degrees. The expected degree of a node i is the square root of the time. That is as time increases the number of nodes linked to node i (its degree) k, decreases.

$$k_i(t) = (\frac{t}{t_i})^{0.5} \tag{3}$$

(node i joins the network at time t_i) (Barabási and Albert-László, 1999) The probability p_i that the new node is connected to node i is: $p_i = \frac{k_i}{\sum_j k_j}$, where k_i is the degree of node i. The degree distribution resulting from the Barabási-Albert model is scale free[9], it is a power law of the form:

$$P(k) \sim k^{-3} \tag{4}$$

(Barabási and Albert-László, 1999)

The Barabási–Albert Pòlya approach

We now present a modified version of the Pòlya urn that is surprisingly mathematically equivalent to the Barabási–Albert process. We add one more ball with a new color at each iteration along with a reinforcement of the ball drawn by one ball (identity reinforcement). Resulting in $k + 2$ types at time k and $2(1 + k)$ marbles in total.

Each node in the network represents a type, in our case a new stock that incepts but for which survival depends on the number of iterations until the new type is selected. The number of nodes (number of edges) connected to each node, representing the

[9]system following a power law at least asymptotically

popularity of the node, can be seen as equivalent to the number of balls of each type. Newly added nodes are equivalent to a new stock added after a draw.

Similarly to the Barabási network for which the distribution of degrees is a power law, the number of balls of a particular type follows a power law distribution.

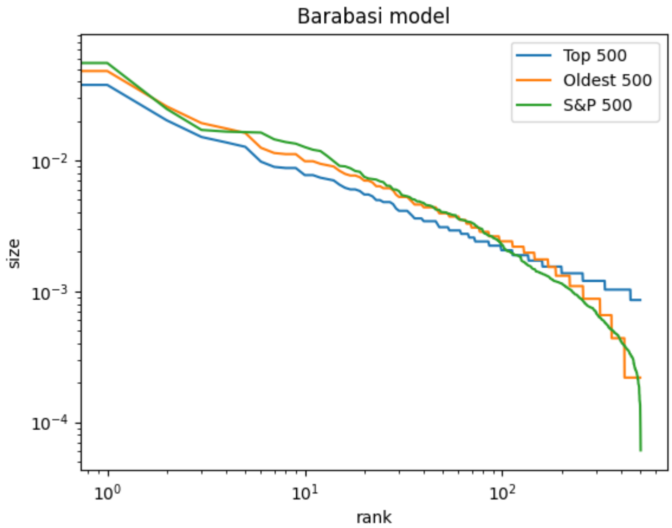

FIGURE 6: Log rank-Log size regression of the $S\&P500$ of both the simulated and the actual market shares for the Barabási-Albert model

Figure 6 shows the log-rank log size (log weight) regression of both the simulated using the modified Polya urn model and the actual market shares as described in equation 1. We observe the **first-mover advantage** phenomenon described in equation 3 for the plot of the 500 oldest balls (nodes). However when taking the Top 500 nodes, the log size-log rank is linear which is predicted for a power law distribution, but it does not capture the **concavity** present in reality for the $S\&P500$ index.

Barabási-Albert Failure (BAF) model

To address the previous issue, we present a model where we allow for eliminating the slowly growing stocks that do not grow rapidly. We introduce a new parameter in the **failure parameter (m)**. One might think of it as stocks that do not grow well enough to achieve a bad **reputation** and get kicked out eventually by venture capitalists. The model now records the number of iterations since the "new" type was added and it discards this type if the number of iterations is higher than m.

The results of the simulation in 8 suggest that when we introduce the failure effect, the graph of the simulation using the Barabási-Albert Failure (BAF) model gains more **curvature** [10] compared to 7 which presents a more linear fit to the power law. This mechanism of births and deaths of new indexes seems **consistent with economic reality**.
As suggested by the BAF model seems to fit well the data with a small sum of square residuals and a higher R-squared than the previous models. For the initial Barabasi model, the R squared went from 0.2% to 93.4% which means that the BAF has a high explanatory power of the data.

Although there exist more theoretical ways to generate empirically observed curvature like Atlas.., the model provides a promising approach that can be used in the context of modeling concentration using reinforced urn processes by adding a parameter to capture failure.

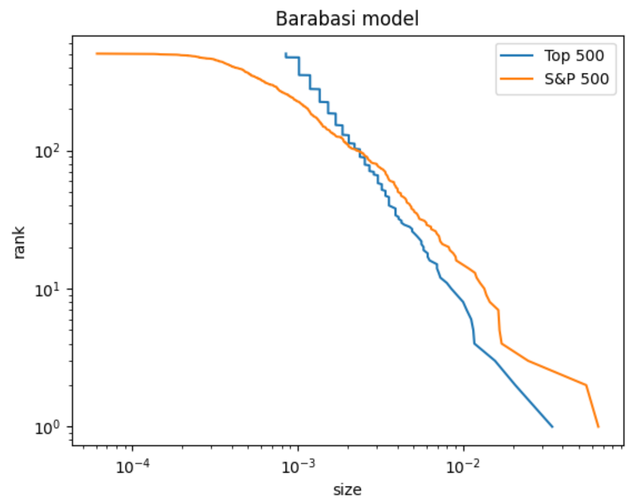

FIGURE 7: Log rank-Log size regression of the $S\&P500$ of both the simulated and the actual market shares for the Barabási-Albert model

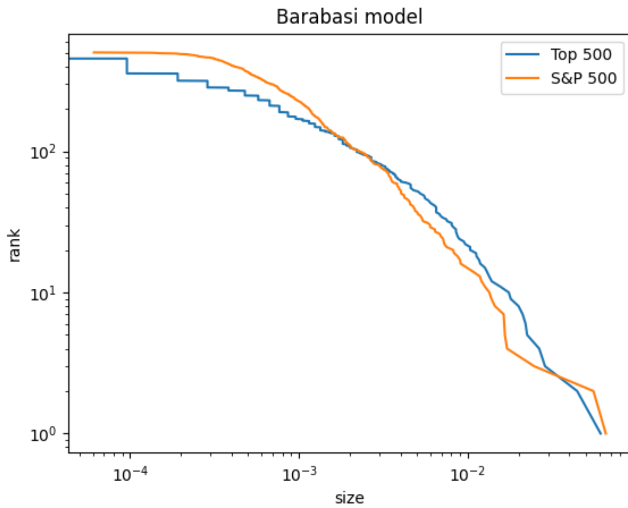

FIGURE 8: Log rank-Log size regression of the $S\&P500$ of both the simulated and the actual market shares for the Barabási-Albert model with failure **BAF**,m=100

[10] which is characteristic of the $S\&P500$ (linearity in the tails) as described in figure 2.

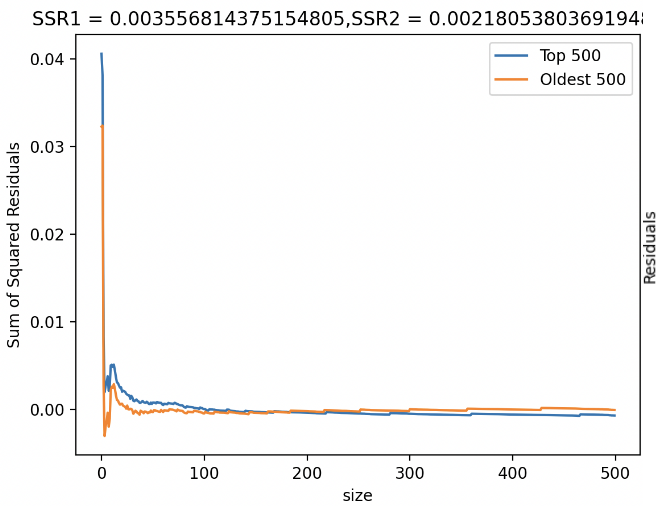

FIGURE 9: Evolution of the residuals of the Barabási-Albert model m=100 and R squared

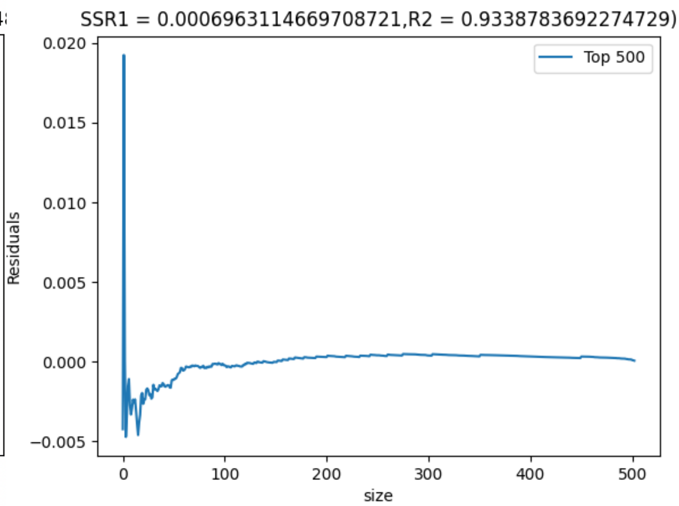

FIGURE 10: Evolution of the residuals of the **BAF**,m=100 and R squared

Conclusion

We use reinforced urn processes to understand the mechanisms giving rise to the dominance of a few stocks in financial indexes using an algorithmic approach. Random processes generate hypothetical indexes that mimic the concentration in actual indexes. However, real-world processes are more complex than the simple dynamics of these models.

Our study is based on (Goldberg et al., 2022) where reflecting geometric Brownian motion is used to model the observed concentration. Results from the Polya urn models give a characterization (distribution, moments...) of the limiting behavior of the weights (market shares) in a market-cap weighted index which makes use of well-established theorems describing the limiting behavior of such models. The results provide a characterization of the limiting behavior of the weights (market shares) in a market cap weighted index, which we can view as an alternative method of understanding the dynamics causing index concentration through random processes using reinforced urns. We present two narratives that might explain the **rich-get-richer** phenomenon observed for market cap weighted indexes: **Born rich, get richer, and get richer with help along the way.**

For the Barabasi-Albert model, we introduced a new parameter in the **failure parameter (m)**. One might think of it as stocks that do not grow well enough to achieve a bad **reputation** and get kicked out eventually by venture capitalists. The Barabasi-Albert model that we extend using a Polya urn approach produces hypothetical indexes that match empirical market cap-weighted indexes in important ways and allows to control the curvature using the introduced failure parameter to match the one intrinsic to the actual index.

This paper attempts to explain the concentration phenomenon using random processes that arise naturally in random networks used to model the internet, trade, and political persuasion... Moreover, the paper contributes to the literature by suggesting models that can be used to test different policy interventions and market structures, to see how they might affect concentration levels and dynamics. This approach can help policymakers and investors to better understand the risks and opportunities associated with concentration, and to design strategies that promote greater market efficiency and stability.

The results of this study may be useful for financial institutions and policymakers in their efforts to promote financial stability and mitigate systemic risks. Further research can be done to explore the potential of reinforced urn processes in other areas of finance and to improve the accuracy of these models. Overall, the findings of this thesis have important implications for the practice and theory of financial modeling and risk management. Furthermore, the models presented can be expanded to include time-dependent reinforcement.

Bibliography

Barabási and Albert-László (1999). “Network Science The Barabási-Albert model”. en. In.

Blackwell, David and James B. MacQueen (1973). “Ferguson Distributions Via Polya Urn Schemes”. In: *The Annals of Statistics* 1.2. Publisher: Institute of Mathematical Statistics, pp. 353–355. ISSN: 0090-5364. URL: https://www.jstor.org/stable/2958020 (visited on 03/16/2023).

Debru, Pierre (Nov. 2020). *SP 500 or SP 5? Is the current concentration sustainable long term?* en. URL: https://www.etfstream.com/articles/sp-500-or-sp-5-is-the-current-concentration-sustainable-long-term (visited on 03/25/2023).

Eurich, Markus and Michael Burtscher (n.d.). “The Business-to-Consumer Lock in Effect”. en. In: ().

Goldberg, Shkolnik et al. (2022). “Is Index Concentration an Inevitable Consequence of Market-Capitalization Weighting? Consortium for Data Analytics in Risk”. In: URL: https://cdar.berkeley.edu/publications/index-concentration-inevitable-consequence-market-capitalization-weighting (visited on 03/13/2023).

MacCluer, C. R. (Jan. 2000). “The Many Proofs and Applications of Perron’s Theorem”. en. In: *SIAM Review* 42.3, pp. 487–498. ISSN: 0036-1445, 1095-7200. DOI: 10.1137/S0036144599359449. URL: http://epubs.siam.org/doi/10.1137/S0036144599359449 (visited on 02/03/2023).

Multivariate Generalized Beta Distributions with Applications to Utility Assessment - David L. Libby, Melvin R. Novick, 1982 (n.d.). URL: https://journals.sagepub.com/doi/abs/10.3102/10769986007004271?journalCode=jeba (visited on 03/09/2023).

Newman, M. E. J. (2009). *The first-mover advantage in scientific publication - IOPscience*. URL: https://iopscience.iop.org/article/10.1209/0295-5075/86/68001 (visited on 03/18/2023).

Pemantle, Robin (2007). "A survey of random processes with reinforcement". In: *Probability Surveys* 4.none. Publisher: Institute of Mathematical Statistics and Bernoulli Society, pp. 1 –79. DOI: 10.1214/07-PS094. URL: https://doi.org/10.1214/07-PS094.

Perron–Frobenius theorem (Feb. 2023). en. Page Version ID: 1139285392. URL: https://en.wikipedia.org/w/index.php?title=Perron%E2%80%93Frobenius_theorem&oldid=1139285392 (visited on 03/09/2023).

"Power Laws and Rich-Get-Richer Phenomena" (n.d.). en. In: ().

Roedenbeck, Marc and Barnas Nothnagel (Jan. 2008). "Rethinking Lock-in and Locking: Adopters Facing Network Effects". In: *Journal of Artificial Societies and Social Simulation* 11.

Shah, Sameena et al. (June 2010). "Trail formation in ants. A generalized Polya urn process". In: *Swarm Intelligence* 4, pp. 145–171. DOI: 10.1007/s11721-010-0041-9.

Trippa, Lorenzo, Peter Müller, and Wesley Johnson (Mar. 2011). "The multivariate beta process and an extension of the Polya tree model". In: *Biometrika* 98.1, pp. 17–34. ISSN: 0006-3444. DOI: 10.1093/biomet/asq072. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3744636/ (visited on 03/09/2023).