Index concentration and rich-get-richer phenomenon using Reinforced Urn Processes

Department of Economics at the **University of California, Berkeley**
Joint work with **Prof. Lisa Goldberg, Prof. Alex Shkolnik and Harrison Selwitz.**
Thesis supervisor: **Prof. Lisa Goldberg**
Presented by **: Nossaiba Kheiri**

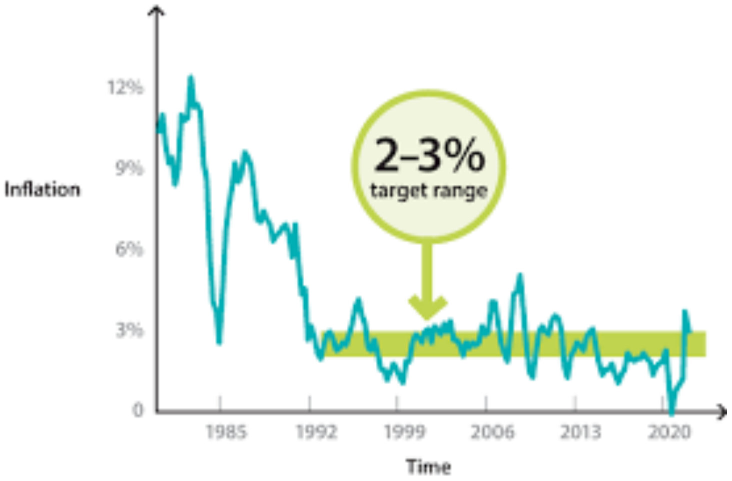
Contents

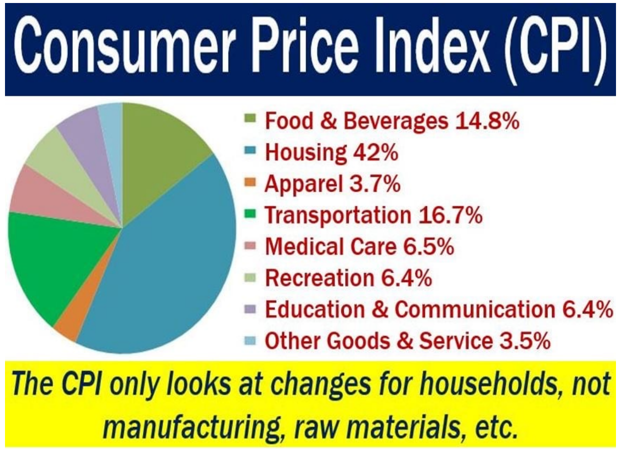
Motivation

In statistics and research design**, indexes** are useful **measures summarizing complex data** in a simple and **standardized** way. They allow for easy comparison over time or across different types.

Indexes are prominent in decision making.

They are also present to measure financial or economic data such as interest rates, inflation, Consumer Price Index (CPI), Human Development Index (HDI)…





Motivation

Financial indexes : measure the price performance of a portfolio “basket of securities”

They play two roles:

1. as standards for performance measurements to track the performance of the stock market over time.
2. as drivers of enormous investments. (investment tracking)

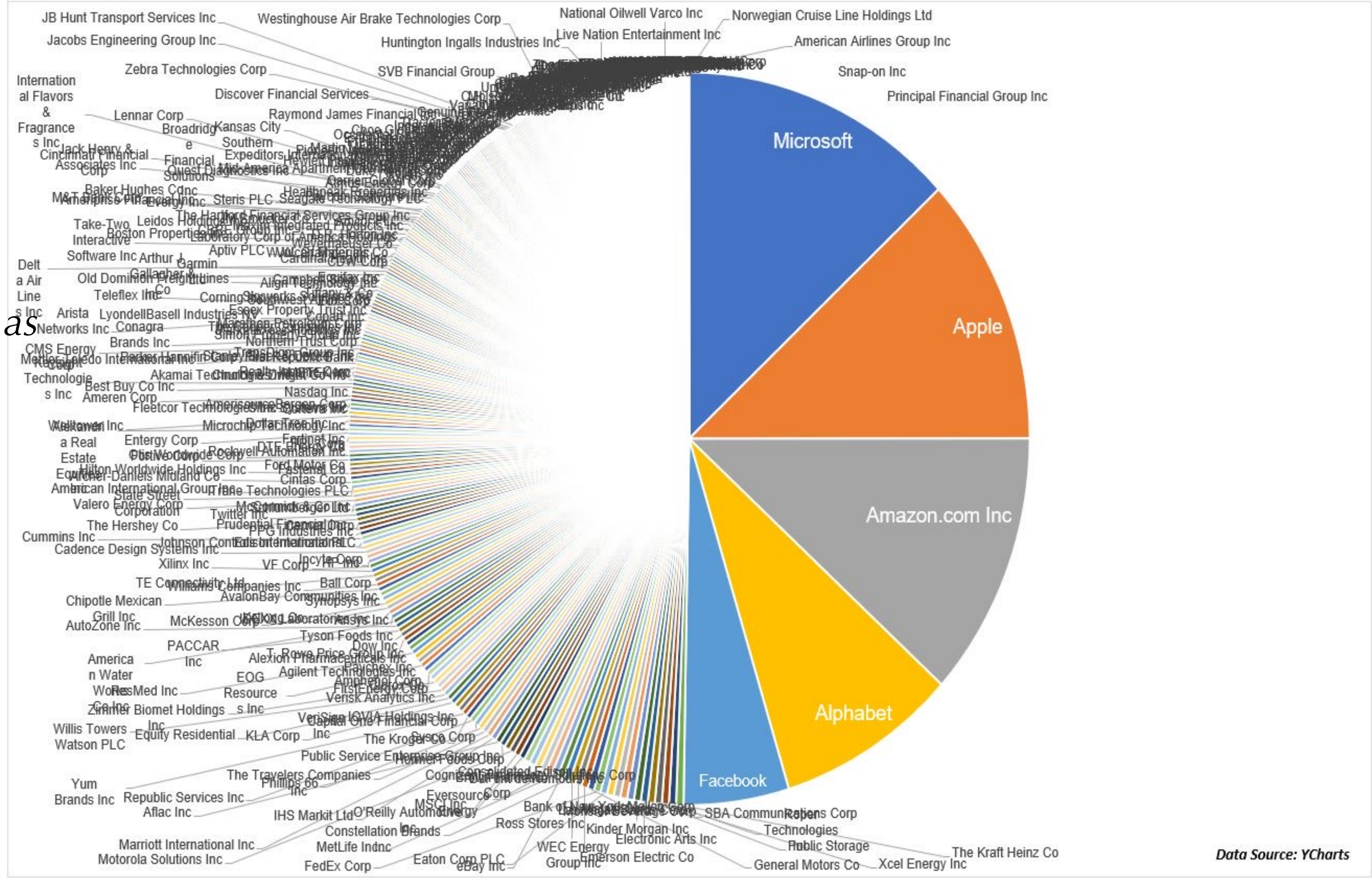
As such, they are tremendously **INFLUENTIAL**

One of the widely common financial index construction is **Capitalization-weighting** **(**MSCI EAFE NASDAQ-100, Russell 2000, S&P 500…)

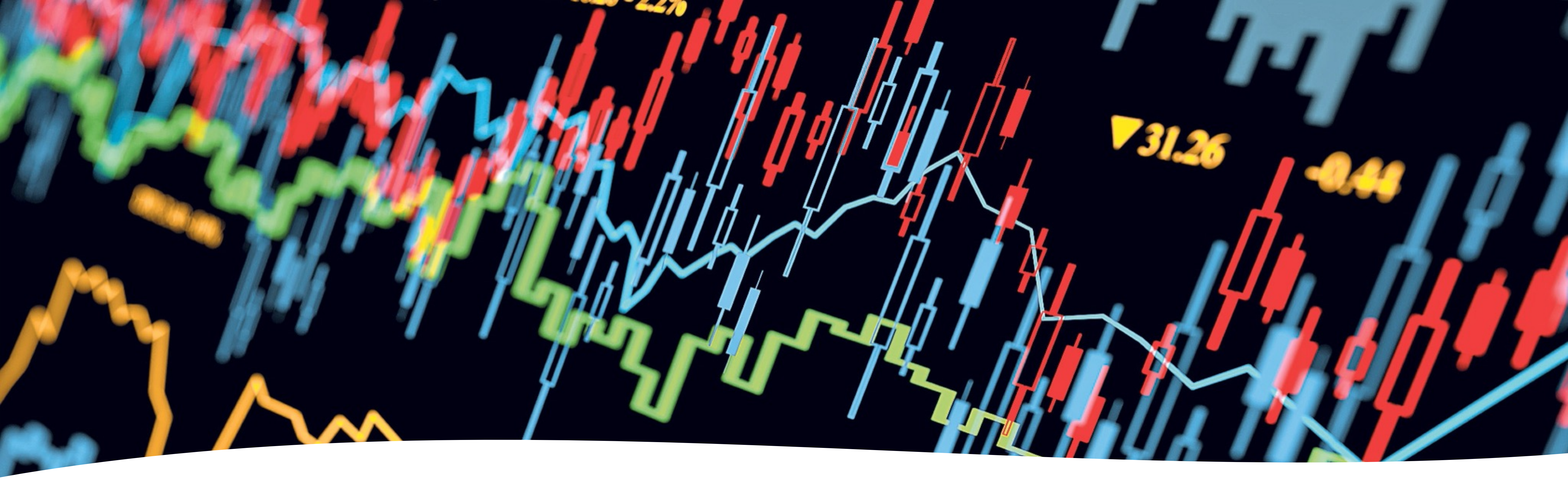
The concentration phenomenon

Your portfolio may not be as diversified as you think…

(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.



The five largest stocks have as big a weighting in the S&P 500 as the bottom 350 stocks in the index (2018)



Objective of the study

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?

This study attempts to understand the mechanisms behind index concentration using power laws as a measure of index concentration.

The approach

While there are many possible explanations for concentration of market-cap weighted indexes, we'll show that it can be explained by **random processes**.

Concentration may arise for **multiple reasons**, including differences in market capitalization, industry dominance, and investor sentiment..

We construct hypothetical indexes that match empirical market cap-weighted indexes using Random processes : **Reinforced urn processes** that mimic the concentration in actual indexes

Urn ≡ A portfolio constructed from the universe of stocks

Source: StockMarketEye

Main findings

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.

We present two mechanisms that might explain the phenomenon observed for market cap-weighted indexes.

These mechanisms come with **very intuitive** narratives

- **Born rich, get richer…**
- **Get richer with help along the way.**

We extend the The **Barabasi Albert model** using a Polya urn approach produces hypothetical indexes that match empirical market cap-weighted indexes in important ways…

Reinforced urn processes in Economics

Economic theory : **The lock-in effect** as a plausible explanation of the extreme imbalances observed in reality or constant popularity of few firms.(Eurich,Burtsher 2014).

The economist **W. Brian Arthur** proposed urn models to model **firms' concentration in a market**. While using simple urn models, his conclusions resonated with the economics community.(cited by 12701),

The paper models agents choosing between technologies competing for adoption

The framework can be extended to accommodate sequential-choice problems

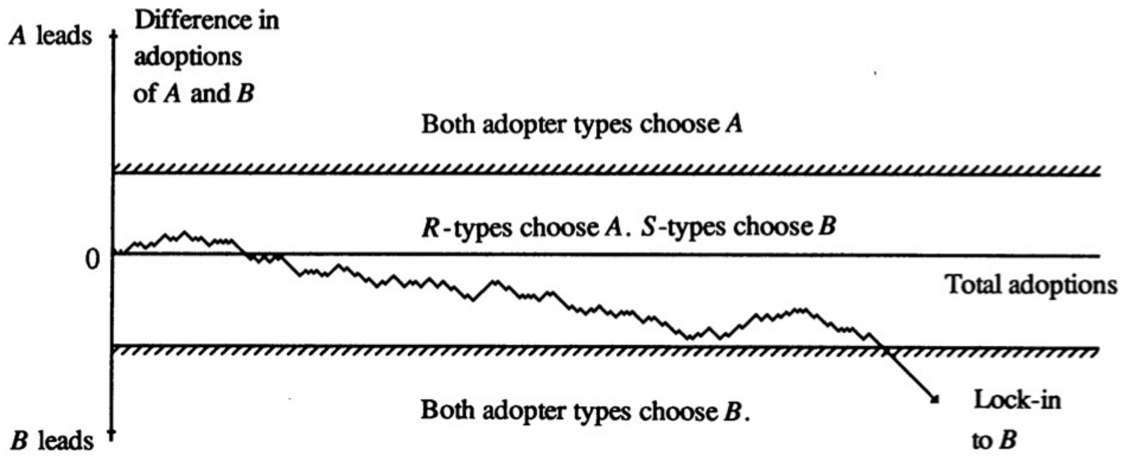


Fig. 1. Increasing returns adoption: a random walk with absorbing barriers

Source: Arthur, W. Brian. "Competing Technologies, Increasing Returns, and Lock-In by Historical Events." The Economic Journal.

Measures of concentration

Many measures of index concentration : **Herfindahl-Hirschman Index (HHI),Gini coefficient, power law coefficient**... We use power laws as a measure of concentration.

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

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

- **Power laws** (extension of Zipf's law) have general applications in linguistics, city populations, the number of citations of scientific publications, firms' concentration…

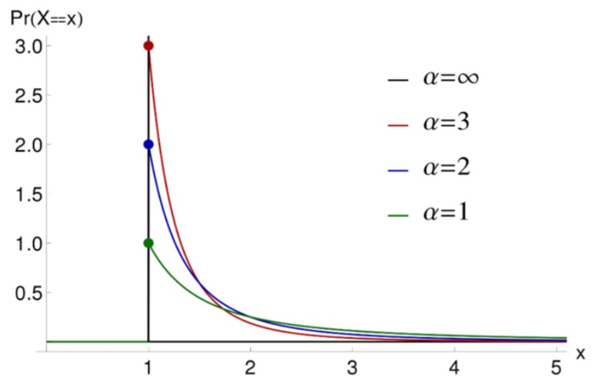
$$S_i \sim \zeta i^{-\alpha}$$

$$ln(i) = ln(\zeta) - \alpha \cdot ln(S_i) \qquad (1)$$

α is the power law coefficient which can be used as a **measure of index concentration.**
Si market share of stock i

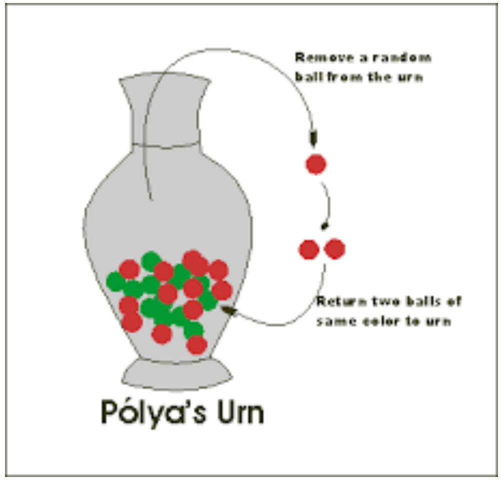
It is the probability distribution function (PDF) associated with the CDF given by Pareto's Law.

Figure 1: Pareto Distribution (various alpha)



Pòlya urn to model index concentration

- **Assumptions:** all the types (stocks, securities) enter the market at roughly the same time n=0
- Non-negative reinforcement for balls of different type $A_{ij} \geq 0$ and for A_{jj} **≥-1**



Initially: contains k≥2 types of balls (colors, stocks...)

Initial allocation of balls $R_{j(0)}$ for 1≤j ≤k

At each iteration: $R_{j(n)}$ the number of balls of type j at time n

1. a ball of type i is sampled from the urn at random
2. Replaced along with A_{ij} balls of type j

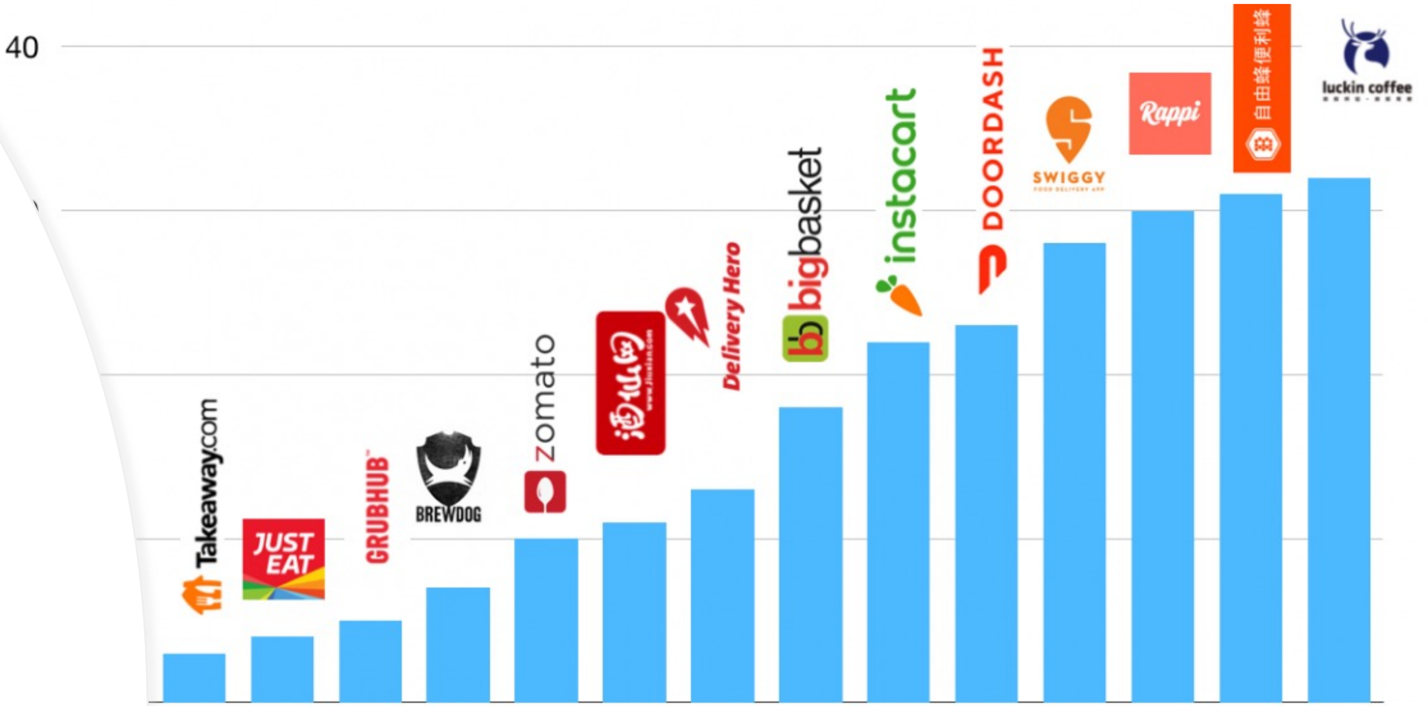
Variable of interest: X_n market share vector at time n i.e. $X_{nj} = \frac{R_{j(n)}}{\sum_{j=1}^{k} R_{j(n)}}$ and $R_{j(n+1)} = R_{j(n)} + A_{ij}$

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

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

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

- **Initial parameters** : strictly positive reinforcement matrix Aij > 0
- **The limiting distribution of the market share vector Xn :**determined by the unique positive left eigenvector of the reinforcement matrix (Pemantle, 2007) /Perron-Frobenius Theorem.
- The model is **deterministic**, regardless of the initial distribution of balls.
- **Intuition:** Firms grow with help along the way (positive reinforcement), this help outweighs the effect of the initial allocation. (Venture capitals,unicorns)



Simulation

Reverse engineering the reinforcement matrix :

- construct the reinforcement matrix 'artificially' given data from the current distribution of the stocks in the index.
- Normalized unique positive left eigenvector is the market capitalization distribution as predicted by Theorem 2.3 (Pemantle, 2007).

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

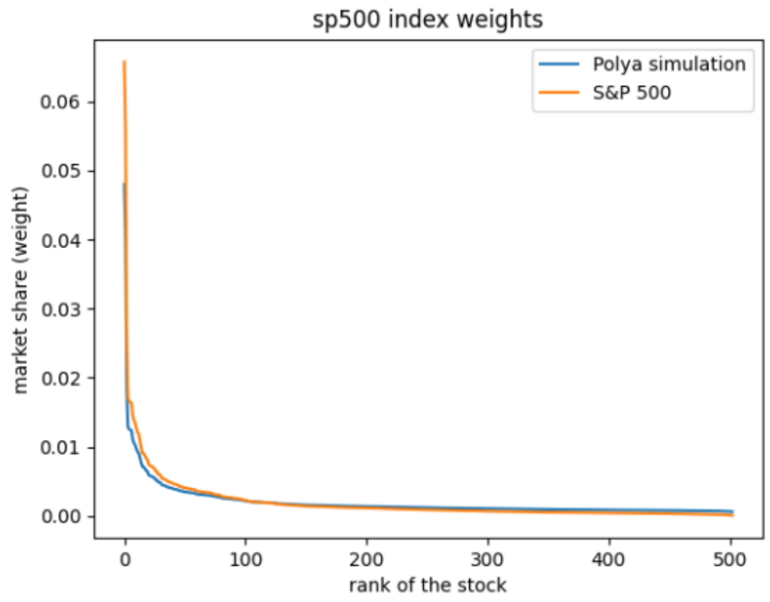


FIGURE 1: $S\&P500$ index market share predicted by the Pòlya model and actual market shares as of February 2023

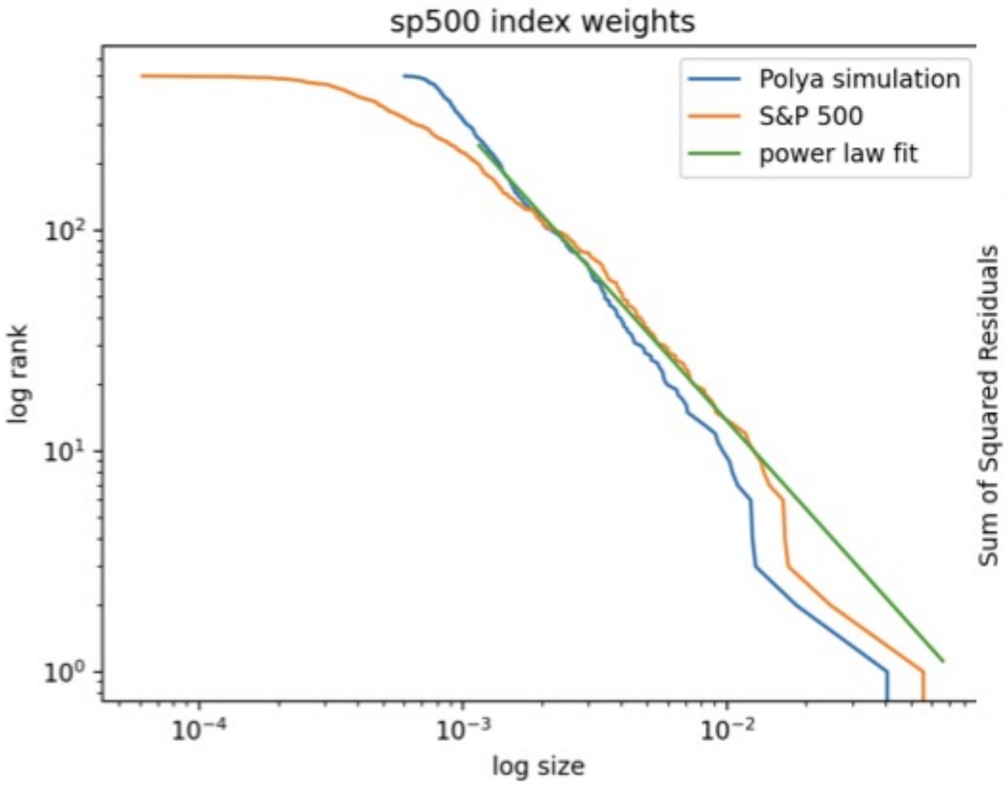


FIGURE 2: Log rank-Log size regression of the $S\&P500$ of both the simulated and the actual market shares

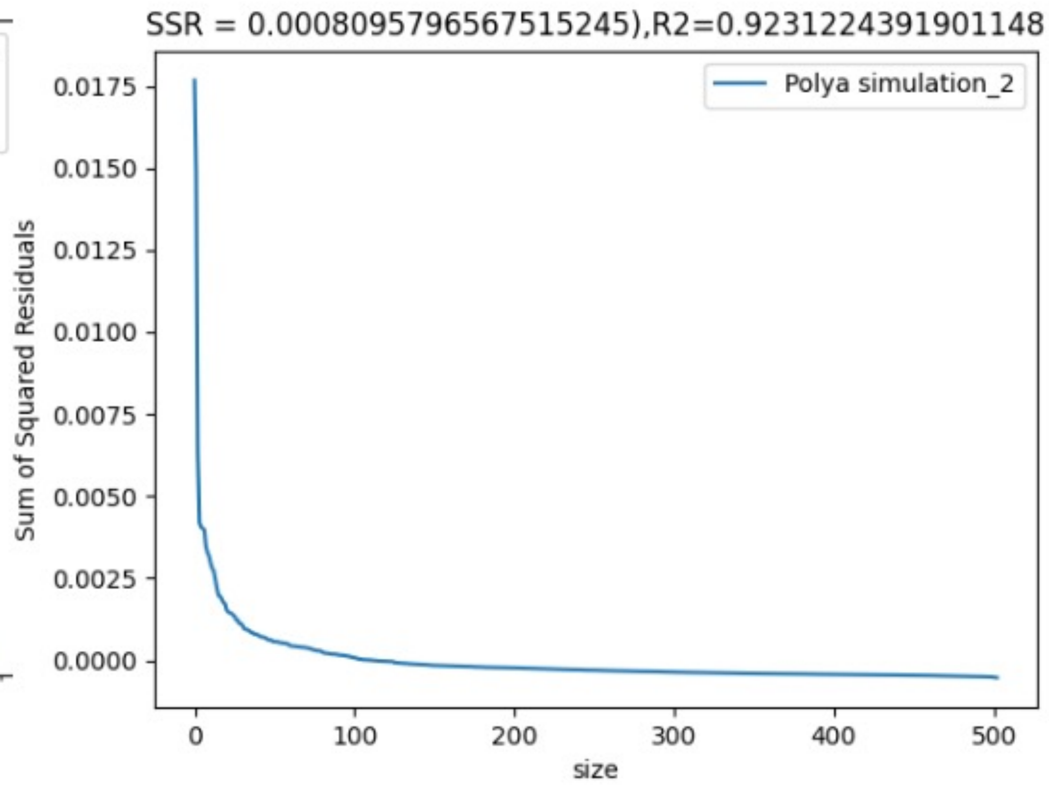


FIGURE 3: Evolution of the residuals for the Pòlya urn 2 model

$$ln(i) = ln(\zeta) - \alpha \cdot ln(S_i) \quad (1)$$

Limitations

- **In practice,** we do not have a priori the final distribution vector of market shares of each stock in the index
- Approach: Continuously update the reinforcement matrix by using previous realizations of the vector of market shares of each stock in the index data chosen while trying to match the current situation of the economy (recession, expansion...).

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

- **Parameters:** with initial allocation (R1(0), . . . , Rk(0)) and identity reinforcement
- **The limiting distribution** is a Dirichlet distribution with parameters (R1(0), . . . , Rk(0)) :a theoretical characterization of the limiting behavior of the weights (market shares) in a market-cap weighted index (distribution, moments...),
- The initial allocation determines the limiting distribution.
- **Intuition:** 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. **Intended' to succeed by birth**

Simulation

- 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).

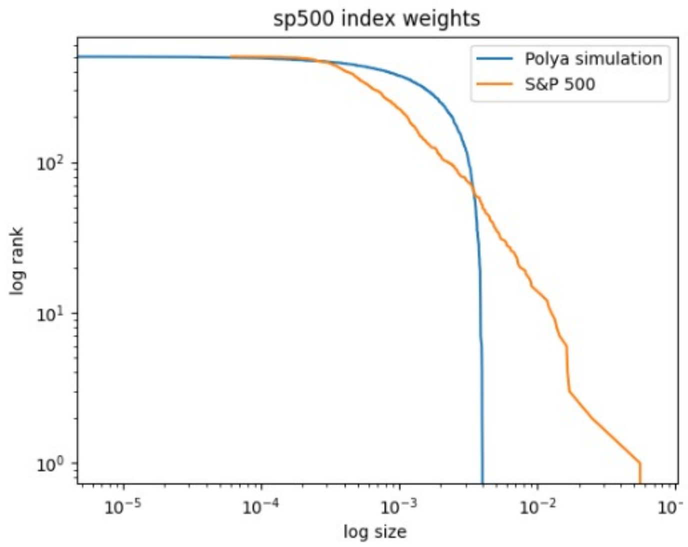


FIGURE 4: Log rank-Log size regression of the $S\&P500$ of both the simulated and the actual market shares

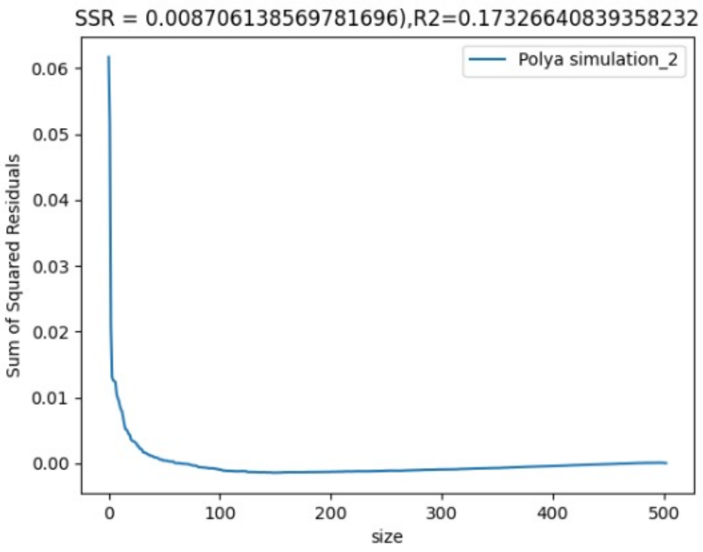
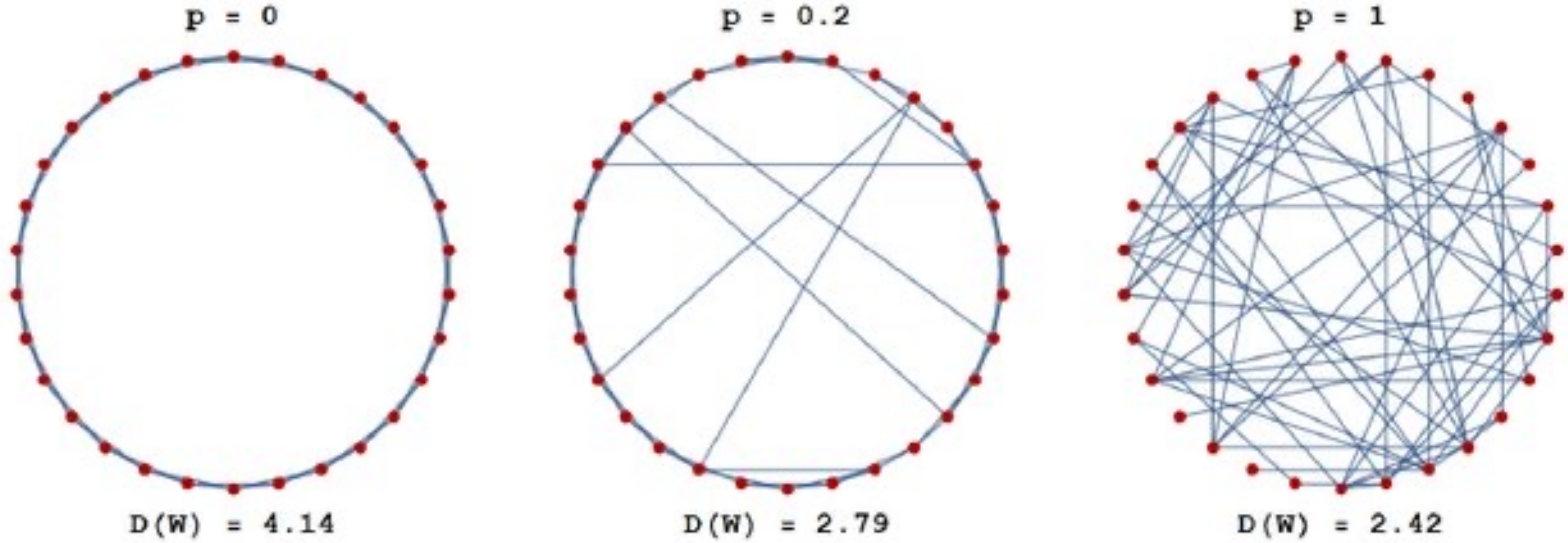


FIGURE 5: Evolution of the residuals for the Pòlya urn model

Limitations

- 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.
- The initial **allocation was drawn randomly**.
- Approach to 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.



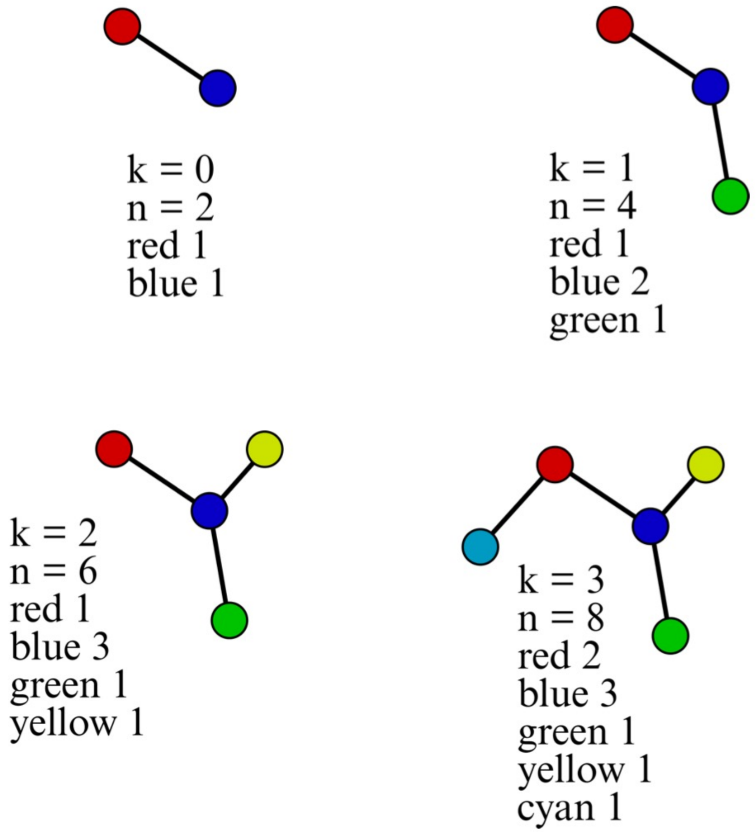
Source: **Network Biology cell dev.,Zenil et al.**

Barabási–Albert to model index concentration

- The Barabási-Albert model is widely studied in **network science** explaining how complex networks could emerge from simple growth rules. (initially to study web networks)
- The Barabási–Albert model which is an algorithm to **generate power laws** (scale-free networks) using preferential attachment
- **One interesting feature** :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).

The Barabási–Albert Pòlya approach

- 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,
- The number of nodes (number of edges) connected to each node, representing the 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.



Source : Alex Shkolnik

Simulation

- Similarily 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.
- Does not capture the concavity present in reality for the S&P500 index.

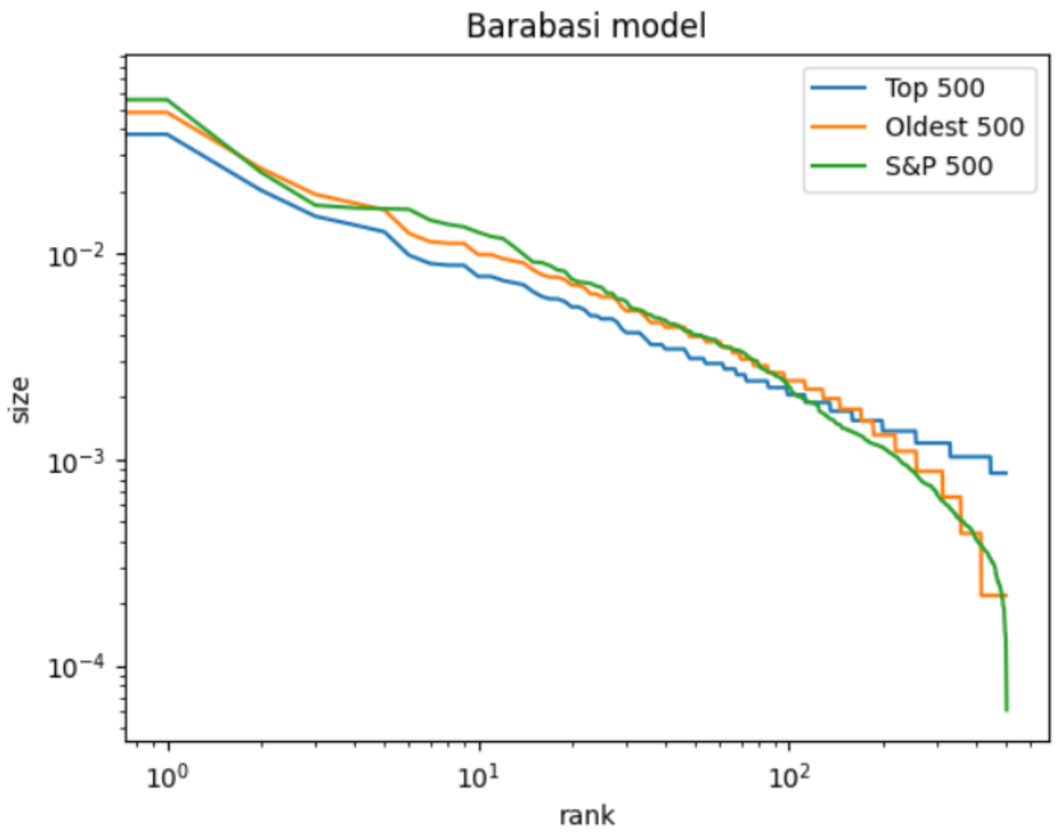


FIGURE 6: Log rank-Log size regression of the $S\&P500$ of both the simulated and the actual market shares for the Barabási-Albert model

Barabási-Albert Failure (BAF) model : Getting kicked out eventually by venture capitalists

- Eliminating the slowly growing stocks that do not grow rapidly.
- **New parameter** : the failure parameter (m).
- This mechanism of births and deaths of new indexes seems consistent with economic reality: 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.
- Gain in curvature
- 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.

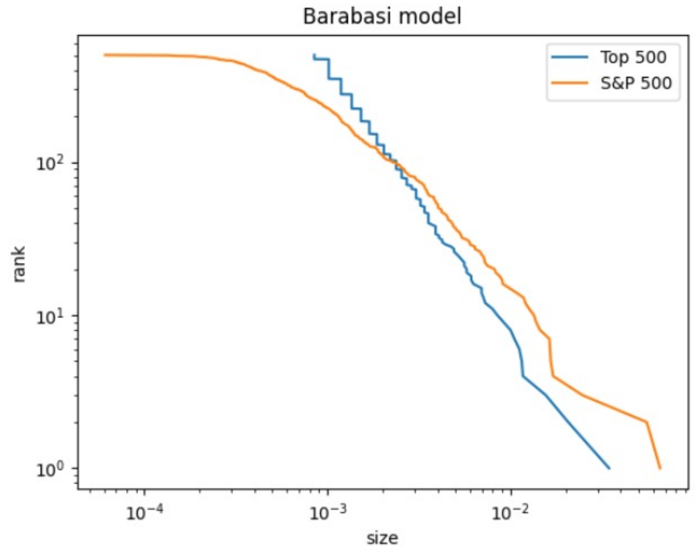


FIGURE 7: Log rank-Log size regression of the $S\&P500$ of both the simulated and the actual market shares for the Barabási-Albert model

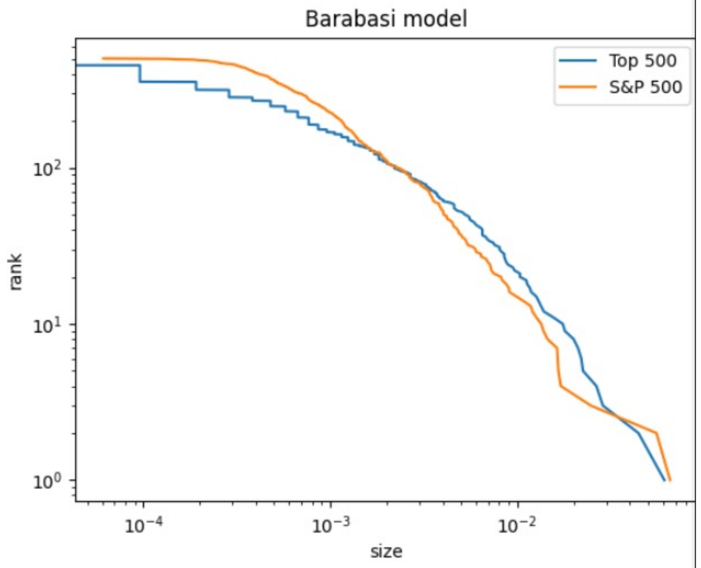


FIGURE 8: Log rank-Log size regression of the $S\&P500$ of both the simulated and the actual market shares for the Barabási-Albert model with failure **BAF**,m=100

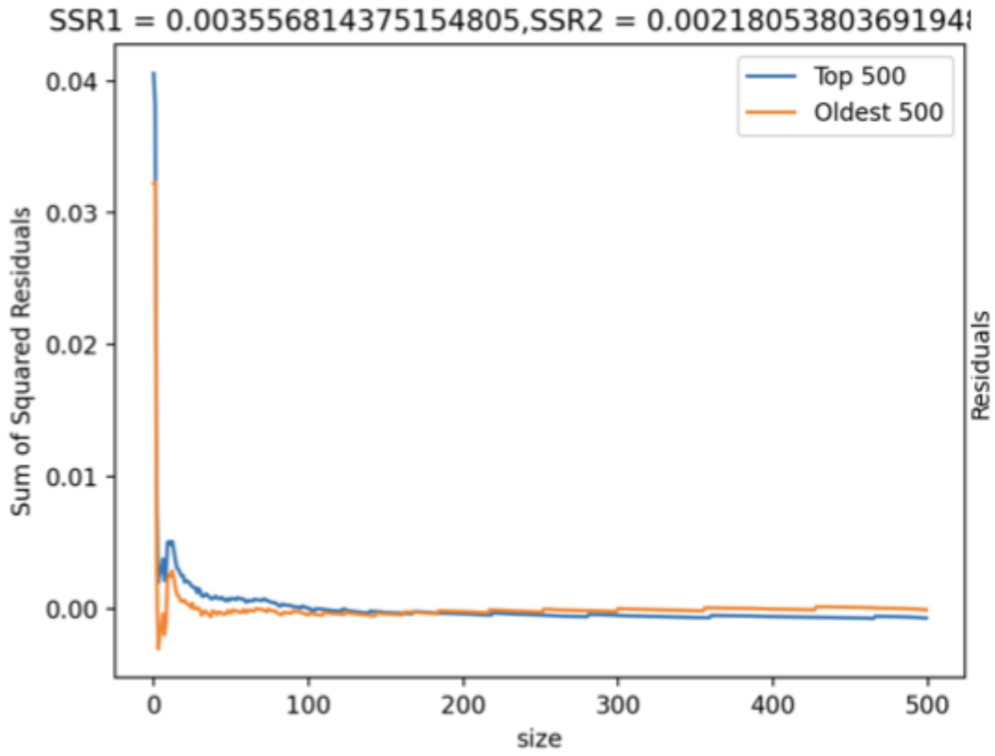


FIGURE 9: Evolution of the residuals of the Barabási-Albert model m=100 and R squared

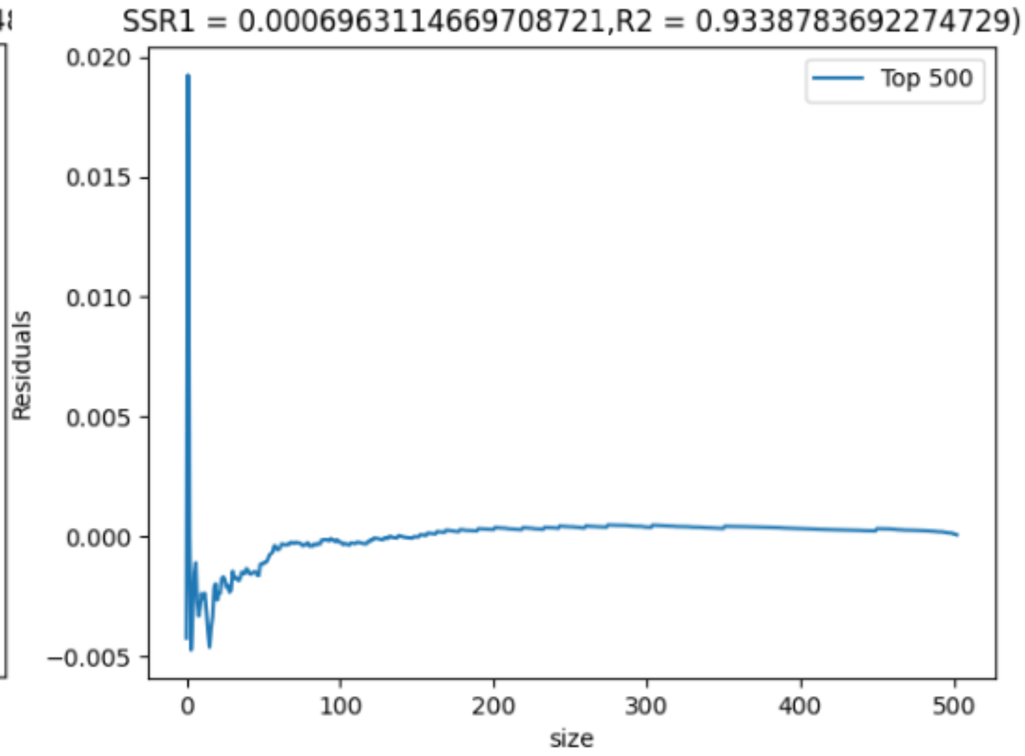


FIGURE 10: Evolution of the residuals of the **BAF**,m=100 and R squared

What is next ?

- **Policy/understanding of market implications**: The results of this study may be useful for financial institutions and policymakers in their efforts to promote diversification of portfolios to achieve financial stability and mitigate systemic risks (various approaches to slow concentration, albeit at the expense of higher turnover (stability)- like in (Goldberg et al., 2022))
- Interesting relation to other fields (graph theory, urn processes…)

A framework to study concentration by dynamically constructing the index using the parameters until convergence but there is more to explore/study:

- Better criterion of convergence
- Different stocks, different weighting
- Study the concentration-turnover paradigm (for Russel 1000, Russel 3000, Russel 2000 large capped, MSCI...) to track the turnover of the top and the bottom.
- Control curvature