Rank-based models with listings and delistings

Theory and calibration

David Itkin
London School of Economics and Political Science (LSE)

April 7, 2025

LSE Statistics Research Showcase

Joint work in progress with Martin Larsson, Licheng Zhang (CMU)

Table of Contents

Long-term modelling of financial markets

THE LONDON SCHOOL OF ECONOMICS AND POLITICAL SCIENCE

- Many market participants are interested in long-time horizons
  - Pension funds, Trust funds, Endowments, etc.
- We will focus on equity markets (stocks).
- Modelling any noisy system over a long period of time is challenging.

  → but may be rewarding!

Long-term modelling of financial markets

THE LONDON SCHOOL OF ECONOMICS AND POLITICAL SCIENCE

- Many market participants are interested in long-time horizons
  - Pension funds, Trust funds, Endowments, etc.
- We will focus on equity markets (stocks).
- Modelling any noisy system over a long period of time is challenging.

  → but may be rewarding!

Questions:

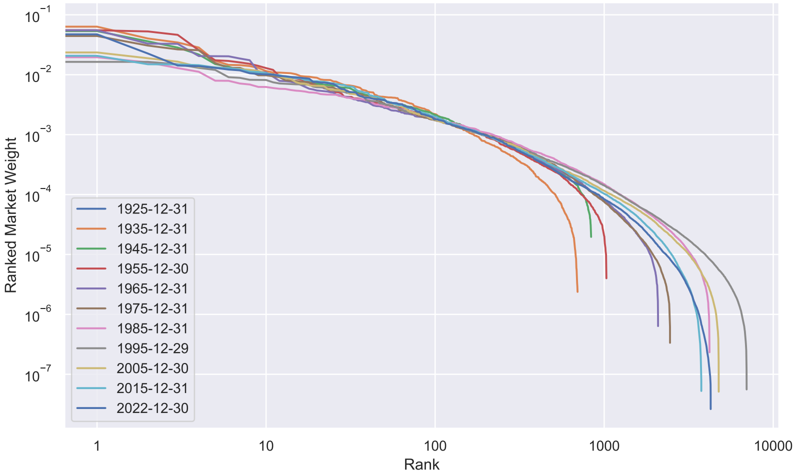
1. What features of equity markets persist over long-time horizons
2. Can we develop models capturing such features and procedures for statistical calibration?

Capital distribution curve

- Let $S_1, \dots, S_N$ denote the market capitalization processes of N companies.
- Set
$$\mu_i(t) = \frac{S_i(t)}{S_1(t) + \cdots + S_N(t)} \qquad i = 1, \dots, N$$
to be the market weights.
- The ranked market weights $\mu_{(1)} \geq \mu_{(2)} \geq \cdots \geq \mu_{(N)}$ are remarkably stable over time.

Capital distribution curve

Ranked market weights over time



Data Source: CRSP

First-order ranked based models

- Curve stability was first observed by Robert Fernholz who developed Stochastic Portfolio Theory (SPT) in 2002
  - SPT further developed together with Ioannis Karatzas, his students including Kardaras, Ruf and many others.

First-order ranked based models

THE LONDON SCHOOL OF ECONOMICS AND POLITICAL SCIENCE

- Curve stability was first observed by Robert Fernholz who developed Stochastic Portfolio Theory (SPT) in 2002
  - SPT further developed together with Ioannis Karatzas, his students including Kardaras, Ruf and many others.
- Classical models do not produce a stable capital distribution curve
- Fernholz proposed rank-based models, which are reduced-form models that can capture the empirical stability of the curve
  - Advantages for calibration due to continuity of data.

THE LONDON SCHOOL OF ECONOMICS AND POLITICAL SCIENCE

First-order ranked based models

- Curve stability was first observed by Robert Fernholz who developed Stochastic Portfolio Theory (SPT) in 2002
  - SPT further developed together with Ioannis Karatzas, his students including Kardaras, Ruf and many others.
- Classical models do not produce a stable capital distribution curve
- Fernholz proposed rank-based models, which are reduced-form models that can capture the empirical stability of the curve
  - Advantages for calibration due to continuity of data.
- Letting $X_i = \log S_i$ the model postulates dynamics

$$dX_i(t) = \gamma_{r_i(t)}dt + \sigma_{r_i(t)}dW_i(t)$$

where $\gamma_i \in \mathbb{R}$, $\sigma_i > 0$, $r_i(t)$ is the rank asset i occupies at time t and W is an N-dimensional Brownian Motion.

First-order ranked based models

THE LONDON SCHOOL OF ECONOMICS AND POLITICAL SCIENCE

- Curve stability was first observed by Robert Fernholz who developed Stochastic Portfolio Theory (SPT) in 2002
  - SPT further developed together with Ioannis Karatzas, his students including Kardaras, Ruf and many others.
- Classical models do not produce a stable capital distribution curve
- Fernholz proposed rank-based models, which are reduced-form models that can capture the empirical stability of the curve
  - Advantages for calibration due to continuity of data.
- Letting $X_i = \log S_i$ the model postulates dynamics

$$dX_i(t) = \gamma_{r_i(t)} dt + \sigma_{r_i(t)} dW_i(t)$$

  where $\gamma_i \in \mathbb{R}$, $\sigma_i > 0$, $r_i(t)$ is the rank asset i occupies at time t and W is an N-dimensional Brownian Motion.
- Under the stability condition

$$\bar{\gamma} := \frac{1}{N}\sum_{k=1}^{N}\gamma_k > \frac{1}{n}\sum_{k=1}^{n}\gamma_k, \quad n = 1, \dots, N-1,$$

  the market weights are ergodic representing stability of the curve.
- When $\gamma_N \gg \gamma_i$ for every other i we say the model is Atlas-like.

Parameter calibration

THE LONDON SCHOOL OF ECONOMICS AND POLITICAL SCIENCE

- Ranked log-caps: $X_{(1)} \geq X_{(2)} \geq \cdots \geq X_{(N)}$.
- Ranked dynamics:

$$dX_{(k)}(t) = \gamma_k dt + \sigma_k d\widetilde{W}_k(t) + \frac{1}{2} d\Lambda_k(t) - \frac{1}{2} d\Lambda_{k-1}(t),$$

  where Λ_k is the local time at zero of $X_{(k)} - X_{(k+1)}$ and $\Lambda_0 = \Lambda_N = 0$.
- This term Λ_k activates when two particles collide $X_{(k)} = X_{(k+1)}$ and ensures the ordering $X_{(k)} \geq X_{(k+1)}$ persists

Parameter calibration

- Ranked log-caps: $X_{(1)} \geq X_{(2)} \geq \cdots \geq X_{(N)}$.
- Ranked dynamics:

$$dX_{(k)}(t) = \gamma_k dt + \sigma_k d\widetilde{W}_k(t) + \frac{1}{2} d\Lambda_k(t) - \frac{1}{2} d\Lambda_{k-1}(t),$$

where Λ_k is the local time at zero of $X_{(k)} - X_{(k+1)}$ and $\Lambda_0 = \Lambda_N = 0$.

- This term Λ_k activates when two particles collide $X_{(k)} = X_{(k+1)}$ and ensures the ordering $X_{(k)} \geq X_{(k+1)}$ persists
- Volatility calibration: $\sigma_k^2 = \frac{d[X_{(k)}](t)}{dt}$,
  $\rightarrow$ standard estimation using quadratic variation.

Parameter calibration

- Ranked log-caps: $X_{(1)} \geq X_{(2)} \geq \cdots \geq X_{(N)}$.
- Ranked dynamics:

$$dX_{(k)}(t) = \gamma_k dt + \sigma_k d\widetilde{W}_k(t) + \frac{1}{2}d\Lambda_k(t) - \frac{1}{2}d\Lambda_{k-1}(t),$$

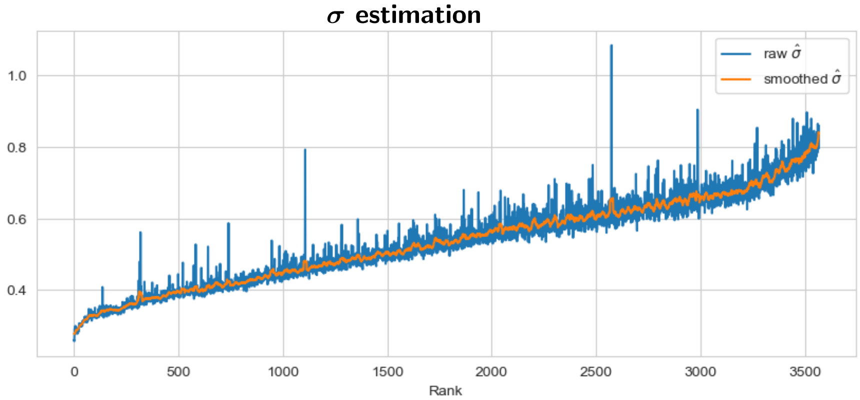
where Λ_k is the local time at zero of $X_{(k)} - X_{(k+1)}$ and $\Lambda_0 = \Lambda_N = 0$.

- This term Λ_k activates when two particles collide $X_{(k)} = X_{(k+1)}$ and ensures the ordering $X_{(k)} \geq X_{(k+1)}$ persists
- Volatility calibration: $\sigma_k^2 = \frac{d[X_{(k)}](t)}{dt}$,
  - → standard estimation using quadratic variation.
- Drift calibration: $\gamma_k - \bar{\gamma} = \frac{1}{2}\lim_{T\to\infty}(\frac{1}{T}\Lambda_{k-1}(T) - \frac{1}{T}\Lambda_k(T))$.
  - → Requires estimating collision rates $\lambda_k = \frac{1}{2}\lim_{T\to\infty}\frac{1}{T}\Lambda_k(T)$.
  - → Efficient method using a so-called "Master formula" for portfolio generation available and developed in Fernholz (2002)

Calibrated Parameters

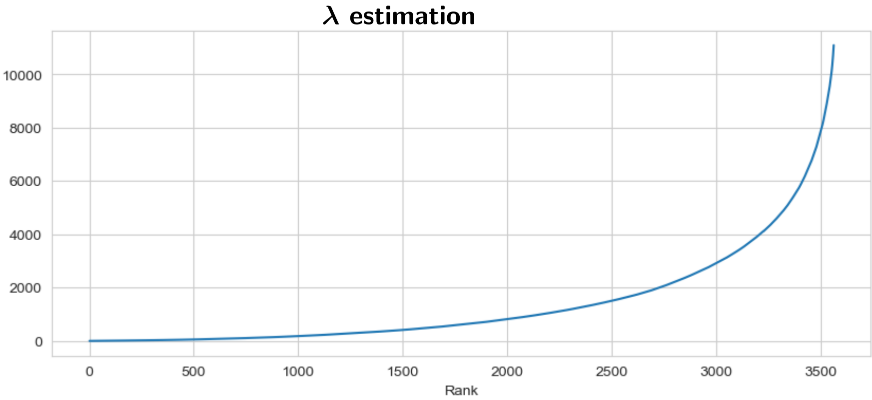
THE LONDON SCHOOL OF ECONOMICS AND POLITICAL SCIENCE

We use 51 years of CRSP data: 1973-2024 with $N = 3565$



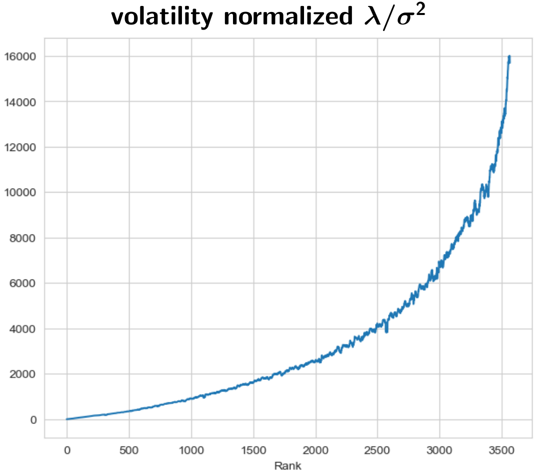
Calibrated Parameters

We use 51 years of CRSP data: 1973-2024 with $N = 3565$



Calibrated Parameters

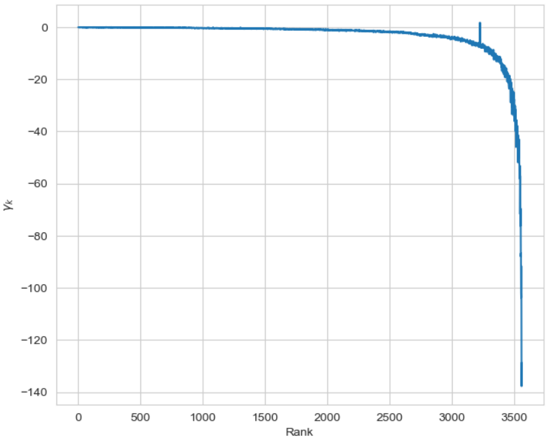
We use 51 years of CRSP data: 1973-2024 with $N = 3565$



Calibrated Parameters

We use 51 years of CRSP data: 1973-2024 with $N = 3565$

γ estimation



Not plotted: $\gamma_d \approx 11000$

Puzzles

THE LONDON SCHOOL OF ECONOMICS AND POLITICAL SCIENCE

- Why are λ's increasing more and more rapidly?

Puzzles

THE LONDON SCHOOL OF ECONOMICS AND POLITICAL SCIENCE

- Why are λ's increasing more and more rapidly?
  - λ_k represents volatility-weighted rate of collisions,
  - the estimates imply rate is monotonically increasing across all ranks.
    - Persists even when accounting for volatility.
  - But aren't there fewer stocks at the tail end?

Puzzles

- Why are λ's increasing more and more rapidly?
  - λ_k represents volatility-weighted rate of collisions,
  - the estimates imply rate is monotonically increasing across all ranks.
    - Persists even when accounting for volatility.
  - But aren't there fewer stocks at the tail end?
- Why does the smallest stock have such a high drift?

Puzzles

- Why are λ's increasing more and more rapidly?
  - λ_k represents volatility-weighted rate of collisions,
  - the estimates imply rate is monotonically increasing across all ranks.
    - Persists even when accounting for volatility.
  - But aren't there fewer stocks at the tail end?
- Why does the smallest stock have such a high drift?
  - Directly linked to one-sided collision at final rank.
  - Artefact of the model due to fixed number of stocks.
  - Is there a simple ad-hoc way to modify the estimates?
  - NO! Removing the extreme estimate would violate stability condition of capital distribution curve.

Puzzles

THE LONDON SCHOOL OF ECONOMICS AND POLITICAL SCIENCE

- Why are λ's increasing more and more rapidly?
  - λ_k represents volatility-weighted rate of collisions,
  - the estimates imply rate is monotonically increasing across all ranks.
    - Persists even when accounting for volatility.
  - But aren't there fewer stocks at the tail end?
- Why does the smallest stock have such a high drift?
  - Directly linked to one-sided collision at final rank.
  - Artefact of the model due to fixed number of stocks.
  - Is there a simple ad-hoc way to modify the estimates?
  - NO! Removing the extreme estimate would violate stability condition of capital distribution curve.

Issues seem to be linked with the smallest stock/fixed universe.

Puzzles

- Why are λ's increasing more and more rapidly?
  - λ_k represents volatility-weighted rate of collisions,
  - the estimates imply rate is monotonically increasing across all ranks.
    - Persists even when accounting for volatility.
  - But aren't there fewer stocks at the tail end?
- Why does the smallest stock have such a high drift?
  - Directly linked to one-sided collision at final rank.
  - Artefact of the model due to fixed number of stocks.
  - Is there a simple ad-hoc way to modify the estimates?
  - NO! Removing the extreme estimate would violate stability condition of capital distribution curve.

Issues seem to be linked with the smallest stock/fixed universe.

- We try to address this by allowing for listings and delistings.
  - How prevalent are they?
- Campbell & Wong (2024) identified these as important drivers for capital distribution curve stability.

Model Setup

THE LONDON SCHOOL OF ECONOMICS AND POLITICAL SCIENCE

- We model a variable equity universe with a focus on developing estimators for empirical calibration,
  - Some recent literature on equity models with variable assets: Sarantsev & Karatzas (2016), Bayraktar, Kim & Tilva (2024)
  - Open markets are another related approach: Fernholz (2018), Karatzas & Kim (2020), Itkin & Larsson (2024).

Model Setup

THE LONDON SCHOOL OF ECONOMICS AND POLITICAL SCIENCE

- We model a variable equity universe with a focus on developing estimators for empirical calibration,
  - Some recent literature on equity models with variable assets: Sarantsev & Karatzas (2016), Bayraktar, Kim & Tilva (2024)
  - Open markets are another related approach: Fernholz (2018), Karatzas & Kim (2020), Itkin & Larsson (2024).
- Countable collection of log caps $(X_i; i \in \mathbb{N})$ with lifetime $[\beta_i, \delta_i]$,
  - β_i is listing time, δ_i is delisting time.

Model Setup

THE LONDON SCHOOL OF ECONOMICS AND POLITICAL SCIENCE

- We model a variable equity universe with a focus on developing estimators for empirical calibration,
  - Some recent literature on equity models with variable assets: Sarantsev & Karatzas (2016), Bayraktar, Kim & Tilva (2024)
  - Open markets are another related approach: Fernholz (2018), Karatzas & Kim (2020), Itkin & Larsson (2024).
- Countable collection of log caps $(X_i; i \in \mathbb{N})$ with lifetime $[\beta_i, \delta_i]$,
  - β_i is listing time, δ_i is delisting time.
- We study the first-order dynamics

$$dX_i(t) = \gamma_{r_i(t)} dt + \sigma_{r_i(t)} dW_i(t), \quad \beta_i \leq t \leq \delta_i,$$

where $r_i(t)$ is the rank of asset i at time t among the listed assets.

Model Setup

THE LONDON SCHOOL OF ECONOMICS AND POLITICAL SCIENCE

- We model a variable equity universe with a focus on developing estimators for empirical calibration,
  - Some recent literature on equity models with variable assets: Sarantsev & Karatzas (2016), Bayraktar, Kim & Tilva (2024)
  - Open markets are another related approach: Fernholz (2018), Karatzas & Kim (2020), Itkin & Larsson (2024).
- Countable collection of log caps $(X_i; i \in \mathbb{N})$ with lifetime $[\beta_i, \delta_i]$,
  - β_i is listing time, δ_i is delisting time.
- We study the first-order dynamics

$$dX_i(t) = \gamma_{r_i(t)} dt + \sigma_{r_i(t)} dW_i(t), \quad \beta_i \leq t \leq \delta_i,$$

  where $r_i(t)$ is the rank of asset i at time t among the listed assets.
- We set $I(t) = \{i : t \in [\beta_i, \delta_i]\}$ and $N(t) = |I(t)|$.
- We assume nondegeneracy: $N(t) \geq 1$ and finite activity: $\sum_{t \leq T} \Delta N(t) < \infty$ for every $T > 0$.

Ranked Dynamics

THE LONDON SCHOOL OF ECONOMICS AND POLITICAL SCIENCE

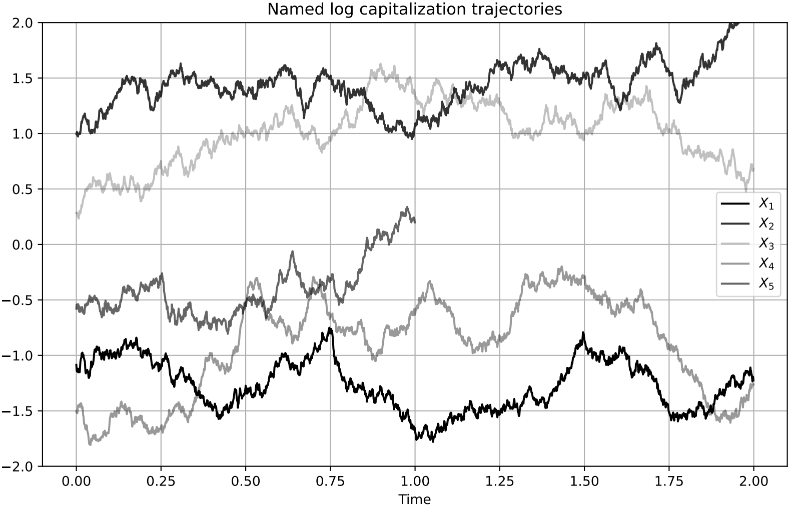
- Despite X_i being continuous on its lifetime $X_{()}$ experiences jumps at listing and delisting times.
- Moreover, the market capitalization $\Sigma(t) = \sum_{i \in I(t)} S_i(t)$, the market weights

$$\mu_i(t) = \frac{S_i(t)}{\Sigma(t)} 1_{\{i \in I(t)\}},$$

and the ranked market weights $\mu_{()}(t)$ experience jumps as well.

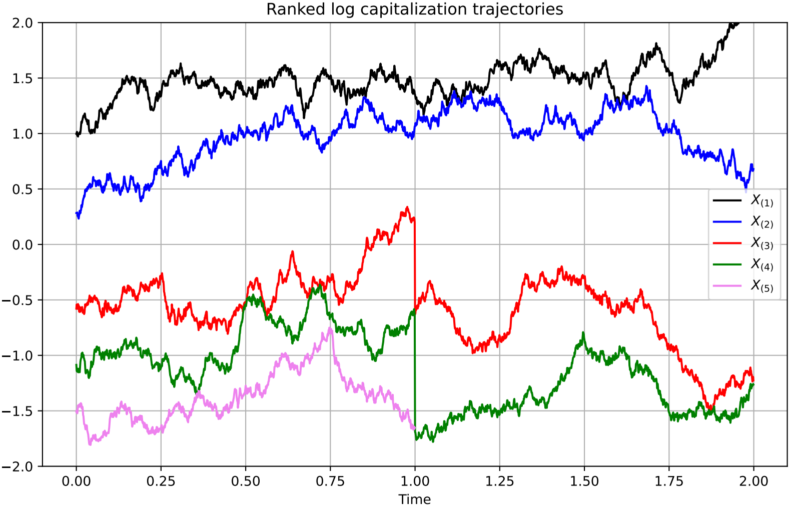
Illustration

THE LONDON SCHOOL OF ECONOMICS AND POLITICAL SCIENCE



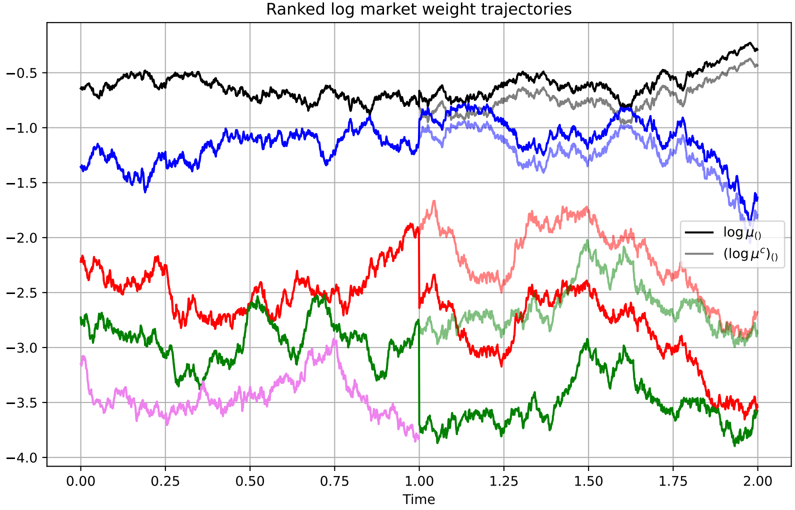
Illustration

THE LONDON SCHOOL OF ECONOMICS AND POLITICAL SCIENCE



Illustration

THE LONDON SCHOOL OF ECONOMICS AND POLITICAL SCIENCE



Ranked Dynamics

THE LONDON SCHOOL OF ECONOMICS AND POLITICAL SCIENCE

- Despite X_i being continuous on its lifetime $X_{()}$ experiences jumps at listing and delisting times.
- Moreover, the market capitalization $\Sigma(t) = \sum_{i \in I(t)} S_i(t)$, the market weights

$$\mu_i(t) = \frac{S_i(t)}{\Sigma(t)} 1_{\{i \in I(t)\}},$$

  and the ranked market weights $\mu_{()}(t)$ experience jumps as well.
- Jumps have a larger effect on the smaller ranks.
- Care is needed dealing with portfolios and the collision estimator.

Ranked Dynamics

- Despite X_i being continuous on its lifetime $X_{()}$ experiences jumps at listing and delisting times.
- Moreover, the market capitalization $\Sigma(t) = \sum_{i \in I(t)} S_i(t)$, the market weights

$$\mu_i(t) = \frac{S_i(t)}{\Sigma(t)} 1_{\{i \in I(t)\}},$$

  and the ranked market weights $\mu_{()}(t)$ experience jumps as well.
- Jumps have a larger effect on the smaller ranks.
- Care is needed dealing with portfolios and the collision estimator.
- Indeed, the wealth $V^{\mathcal{M}}$ of an investor trading the market portfolio now satisfies

$$\log V^{\mathcal{M}}(t) = \log \Sigma^c(t),$$

  whereas classically $\log V^{\mathcal{M}}(t) = \log \Sigma(t)$.

Collision estimation with listings/delistings

- We developed a new "Master formula" for portfolio generation in this setting with listings and delistings.
- As in the classical setting the collision estimator can be derived by looking at the large-cap portfolio investing in the top k assets.
- In the fixed investment universe

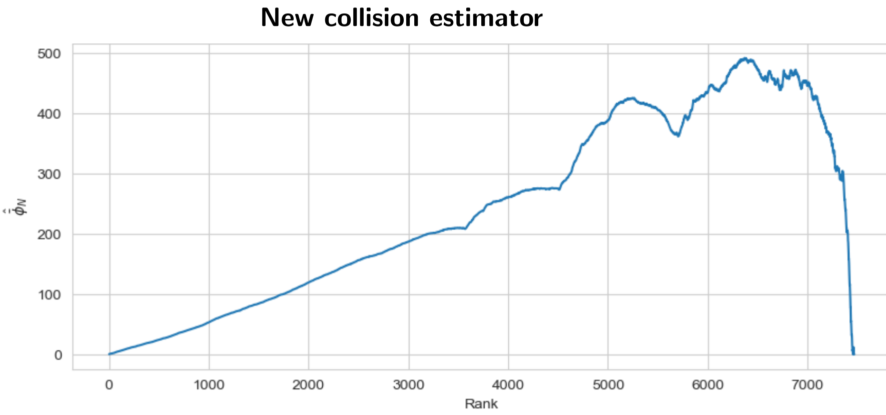
$$d\Lambda_k(t) = \frac{\mu_{(1)}(t) + \cdots + \mu_{(k)}(t)}{\mu_{(k)}(t)} d \log \left(\frac{V^{\mathcal{M}_k}}{V^{\mathcal{M}}} \times \frac{1}{\mu_{(1)} + \cdots + \mu_{(k)}} \right)(t)$$

- With listings/delistings (on the set $\{|I(t)| \geq k\}$):

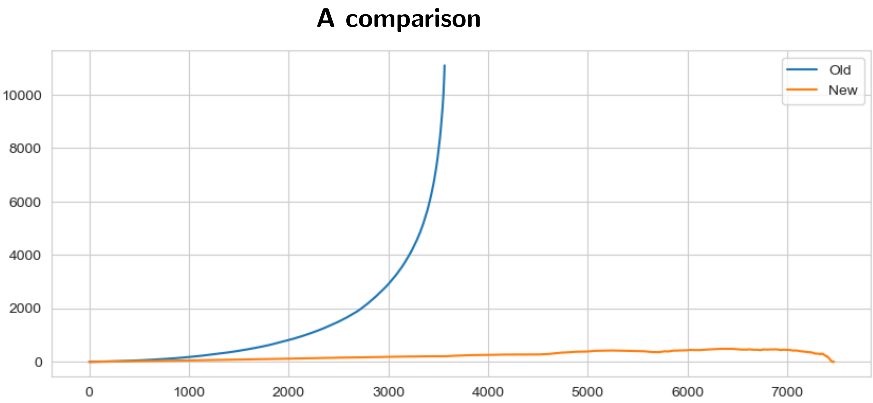
$$d\Lambda_k(t) = \frac{\tilde{\mu}_{(1)}(t) + \cdots + \tilde{\mu}_{(k)}(t)}{\tilde{\mu}_{(k)}(t)} d \log \left(\frac{V^{\mathcal{M}_k}}{V^{\mathcal{M}}} \times \frac{1}{\tilde{\mu}_{(1)} + \cdots + \tilde{\mu}_{(k)}} \right)(t),$$

where $\log \tilde{\mu}(t) = \log \mu^c(t)$.

Collision estimation results



Collision estimation results



Collision estimation with listings/delistings

- In the fixed investment universe

$$d\Lambda_k(t) = \frac{\mu_{(1)}(t) + \cdots + \mu_{(k)}(t)}{\mu_{(k)}(t)} d\log\left(\frac{V^{\mathcal{M}_k}}{V^{\mathcal{M}}} \times \frac{1}{\mu_{(1)} + \cdots + \mu_{(k)}}\right)(t)$$

- With listings/delistings (on the set $\{|I(t)| \geq k\}$):

$$d\Lambda_k(t) = \frac{\tilde{\mu}_{(1)}(t) + \cdots + \tilde{\mu}_{(k)}(t)}{\tilde{\mu}_{(k)}(t)} d\log\left(\frac{V^{\mathcal{M}_k}}{V^{\mathcal{M}}} \times \frac{1}{\tilde{\mu}_{(1)} + \cdots + \tilde{\mu}_{(k)}}\right)(t),$$

where $\log\tilde{\mu}(t) = \log\mu^c(t)$.

The original estimator registers rank change caused by a listing/delisting as a collision.

- Although the original estimator is consistent in the fixed asset model, when applied to real data, it produces bias.
- The bias propogates effecting the smallest stocks the most since a listing/delisting at rank k causes a jump for each $\mu_{(\ell)}$, $\ell > k$.

THE LONDON SCHOOL OF ECONOMICS AND POLITICAL SCIENCE

Particle Density

- What is a driver of collisions?
- Naively, we expect
  - More collisions, if highly volatile,
  - Fewer collisions if neighbours are positively correlated,
  - More collisions if particles tightly packed.

THE LONDON SCHOOL OF ECONOMICS AND POLITICAL SCIENCE

Particle Density

- What is a driver of collisions?
- Naively, we expect
  - More collisions, if highly volatile,
  - Fewer collisions if neighbours are positively correlated,
  - More collisions if particles tightly packed.
- We define the neighbouring rank correlation

$$\rho_k = \lim_{T \to \infty} \frac{\frac{1}{T}[X_{(k)}, X_{(k+1)}](T)}{\sigma_k \sigma_{k+1}}$$

,

Particle Density

- What is a driver of collisions?
- Naively, we expect
  - More collisions, if highly volatile,
  - Fewer collisions if neighbours are positively correlated,
  - More collisions if particles tightly packed.
- We define the neighbouring rank correlation

$$\rho_k = \lim_{T \to \infty} \frac{\frac{1}{T}[X_{(k)}, X_{(k+1)}](T)}{\sigma_k \sigma_{k+1}}$$

,

- We define average particle density at rank k as

$$\phi_k = \lim_{T \to \infty} \frac{1}{T} \int_0^T \frac{2n-1}{X_{(k-n)}(t) - X_{(k+n)}(t)} dt.$$

for a hyperparameter n.

Particle Density

- What is a driver of collisions?
- Naively, we expect
  - More collisions, if highly volatile,
  - Fewer collisions if neighbours are positively correlated,
  - More collisions if particles tightly packed.
- We define the neighbouring rank correlation

$$\rho_k = \lim_{T\to\infty} \frac{\frac{1}{T}[X_{(k)}, X_{(k+1)}](T)}{\sigma_k \sigma_{k+1}}$$

,

- We define average particle density at rank k as

$$\phi_k = \lim_{T\to\infty} \frac{1}{T} \int_0^T \frac{2n-1}{X_{(k-n)}(t) - X_{(k+n)}(t)} dt.$$

for a hyperparameter n.

What is the relationship between σ_k, ρ_k, ϕ_k and how is this related to the collision rates λ_k?

Local Model

- Studying this in the global model is difficult, so build a local model.
- **Idea:** For a fixed rank k create a synthetic large particle model where a typical particle behaves like our rank k one.

Local Model

- Studying this in the global model is difficult, so build a local model.
- **Idea:** For a fixed rank k create a synthetic large particle model where a typical particle behaves like our rank k one.
- We choose the following model

$$dX_i^N(t) = -\frac{\sigma^2}{2} X_i^N(t)dt + \sigma dW_i(t), \quad i = -N, \dots, 0, \dots, N,$$

  with $d[W_i, W_j](t) = \rho dt$ for all $i \neq j$.
- This is an Ornstein–Uhlenbeck process with stationary measure $N(0, \Sigma_N)$ where $\Sigma_N = (1-\rho)I_N + \rho 1_{N \times N}$.

Local Model

- Studying this in the global model is difficult, so build a local model.
- **Idea:** For a fixed rank k create a synthetic large particle model where a typical particle behaves like our rank k one.
- We choose the following model

$$dX_i^N(t) = -\frac{\sigma^2}{2}X_i^N(t)dt + \sigma dW_i(t), \quad i = -N, \dots, 0, \dots, N,$$

with $d[W_i, W_j](t) = \rho dt$ for all $i \neq j$.

- This is an Ornstein–Uhlenbeck process with stationary measure $N(0, \Sigma_N)$ where $\Sigma_N = (1-\rho)I_N + \rho 1_{N\times N}$.
- To study rank k we would take $\sigma = \sigma_k$, $\rho = \rho_k$, start the process at stationarity and study the median particle $X_{(0)}$ as the typical one.
  - The analysis leads to the same conclusion for any fixed quantile.

Local Model

- Studying this in the global model is difficult, so build a local model.
- **Idea:** For a fixed rank k create a synthetic large particle model where a typical particle behaves like our rank k one.
- We choose the following model

$$dX_i^N(t) = -\frac{\sigma^2}{2} X_i^N(t)dt + \sigma dW_i(t), \quad i = -N, \ldots, 0, \ldots, N,$$

  with $d[W_i, W_j](t) = \rho dt$ for all $i \neq j$.
- This is an Ornstein–Uhlenbeck process with stationary measure $N(0, \Sigma_N)$ where $\Sigma_N = (1-\rho)I_N + \rho 1_{N\times N}$.
- To study rank k we would take $\sigma = \sigma_k$, $\rho = \rho_k$, start the process at stationarity and study the median particle $X_{(0)}$ as the typical one.
  - The analysis leads to the same conclusion for any fixed quantile.
- We are interested in understanding the relationship between σ, ρ, λ and ϕ as $N \to \infty$.

Some order statistics results

THE LONDON SCHOOL OF ECONOMICS AND POLITICAL SCIENCE

- We can write $X_i^N = \sqrt{\rho}Y + \sqrt{1-\rho}Z_i^N$ where Y and $(Z_i^N; i = -N, \ldots, N)$ are IID N(0,1) random variables.
- Hence the gaps satisfy $X_i^N - X_j^N = \sqrt{1-\rho}(Z_i^N - Z_j^N)$ so it is enough to study the IID case.

Some order statistics results

- We can write $X_i^N = \sqrt{\rho}Y + \sqrt{1-\rho}Z_i^N$ where Y and $(Z_i^N; i = -N, \ldots, N)$ are IID N(0,1) random variables.
- Hence the gaps satisfy $X_i^N - X_j^N = \sqrt{1-\rho}(Z_i^N - Z_j^N)$ so it is enough to study the IID case.

Proposition (I., Larsson, Zhang (2025+))

1. *Then density f^N on $[0,\infty)$ of the random variable $N(Z_{(0)}^N - Z_{(1)}^N)$ satisfies $\lim_{N\to\infty} f^N(0) = \sqrt{2/\pi}$.*
2. *The random variable $N(Z_{(-n)} - Z_{(n)})$ converges in distribution as $N \to \infty$ to a $\Gamma(2n, \sqrt{2/\pi})$ random variable.*
3. *The weak convergence in part two extends to the unbounded function $h(x) = 1/x$.*

Collision parameter in the local model

THE LONDON SCHOOL OF ECONOMICS AND POLITICAL SCIENCE

- We have

$$\frac{\lambda_0^N}{N} = \lim_{\epsilon \to 0} \mathbb{E}\left[\frac{1}{T}\int_0^T \frac{1}{N\epsilon} 1_{\{X_{(0)}^N(t) - X_{(1)}^N(t) < \epsilon\}} d[X_{(0)}^N - X_{(1)}^N](t)\right]$$

$$= 2\sigma^2\sqrt{1-\rho} f^N(0) \longrightarrow 2\sigma^2\sqrt{1-\rho} \times \sqrt{\frac{2}{\pi}},$$

Collision parameter in the local model

THE LONDON SCHOOL OF ECONOMICS AND POLITICAL SCIENCE

- We have

$$\frac{\lambda_0^N}{N} = \lim_{\epsilon \to 0} \mathbb{E}\left[\frac{1}{T}\int_0^T \frac{1}{N\epsilon} 1_{\{X_{(0)}^N(t) - X_{(1)}^N(t) < \epsilon\}} d[X_{(0)}^N - X_{(1)}^N](t)\right]$$

$$= 2\sigma^2\sqrt{1-\rho} f^N(0) \longrightarrow 2\sigma^2\sqrt{1-\rho} \times \sqrt{\frac{2}{\pi}},$$

- and

$$\frac{\phi_0^N}{N} = \mathbb{E}\left[\frac{2n-1}{N(X_{(-n)} - X_{(n)})}\right] \longrightarrow (2n-1)\int_0^\infty \frac{\lambda^{2n}}{\Gamma(2n)} x^{2n-2} e^{-\sqrt{2/\pi}x} dx$$

$$= \sqrt{\frac{2}{\pi}} \times \frac{1}{\sqrt{1-\rho}}.$$

Collision parameter in the local model

- We have

$$\frac{\lambda_0^N}{N} = \lim_{\epsilon \to 0} \mathbb{E}\left[\frac{1}{T}\int_0^T \frac{1}{N\epsilon} 1_{\{X_{(0)}^N(t) - X_{(1)}^N(t) < \epsilon\}} d[X_{(0)}^N - X_{(1)}^N](t)\right]$$

$$= 2\sigma^2\sqrt{1-\rho} f^N(0) \longrightarrow 2\sigma^2\sqrt{1-\rho} \times \sqrt{\frac{2}{\pi}},$$

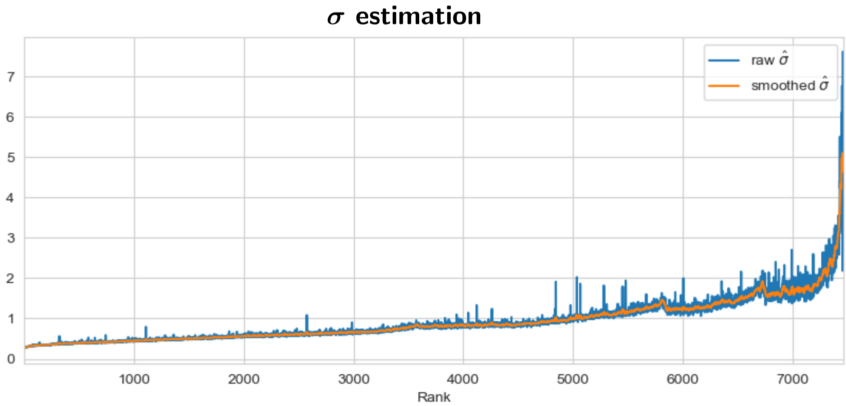
- and

$$\frac{\phi_0^N}{N} = \mathbb{E}\left[\frac{2n-1}{N(X_{(-n)} - X_{(n)})}\right] \longrightarrow (2n-1)\int_0^\infty \frac{\lambda^{2n}}{\Gamma(2n)} x^{2n-2} e^{-\sqrt{2/\pi}x} dx$$

$$= \sqrt{\frac{2}{\pi}} \times \frac{1}{\sqrt{1-\rho}}.$$

- Hence, $\phi_0^N \approx \frac{\lambda_0^N}{2\sigma^2(1-\rho)}$.
- Let see how the listing and delisting model performs with real data.

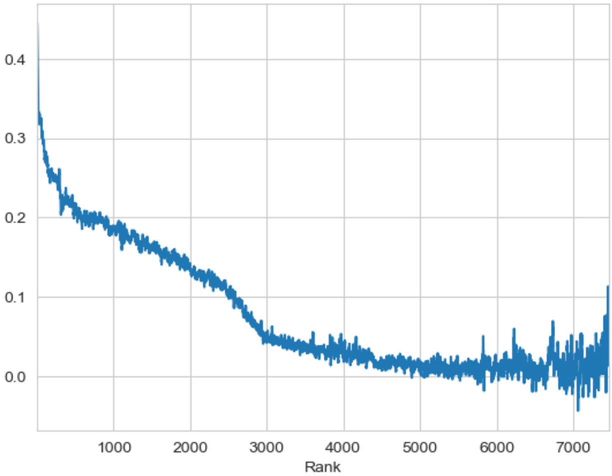
Collisions and Particle Density: Empirics



Collisions and Particle Density: Empirics

THE LONDON SCHOOL OF ECONOMICS AND POLITICAL SCIENCE

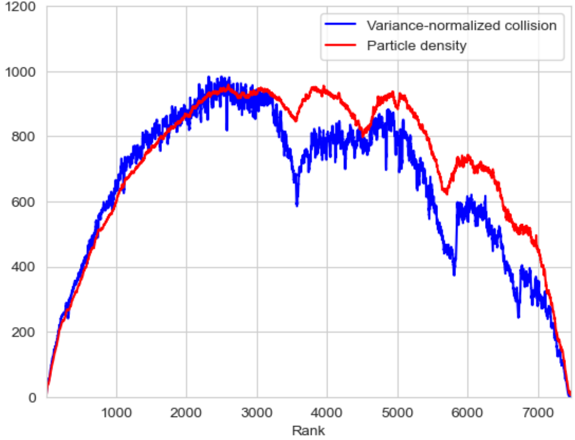
ρ-estimation



Collisions and Particle Density: Empirics

THE LONDON SCHOOL OF ECONOMICS AND POLITICAL SCIENCE

ϕ_k **vs** $\frac{\lambda_k}{2\sigma_k^2(1-\rho_k)}$ **with** $n = 5$.



A more granular look?

Conclusion and future work

THE LONDON SCHOOL OF ECONOMICS AND POLITICAL SCIENCE

Summary of results

- Introduced a rank-based model with listings and delistings,
- Derived a new master formula for portfolio generation
- Derived collision estimator accounting for listings/delistings, which corrects bias of previously used estimator when applied to real data,
- Studied local model and connected collisions to particle density.

To be done:

- Estimate listing and delisting rates,
- Pick a (Markovian) birth/death mechanism for global model,
- Conduct numerical and simulation experiments for global model.
- Theoretical analysis of global model?

Conclusion and future work

THE LONDON SCHOOL OF ECONOMICS AND POLITICAL SCIENCE

Summary of results

- Introduced a rank-based model with listings and delistings,
- Derived a new master formula for portfolio generation
- Derived collision estimator accounting for listings/delistings, which corrects bias of previously used estimator when applied to real data,
- Studied local model and connected collisions to particle density.

To be done:

- Estimate listing and delisting rates,
- Pick a (Markovian) birth/death mechanism for global model,
- Conduct numerical and simulation experiments for global model.
- Theoretical analysis of global model?

Thank you!

Master Formula for Functional Generation

- In this setting we say a portfolio π is functionally generated if

$$\log\left(\frac{V^{\pi}(T)}{V^{\mathcal{M}}(T)}\right) = \log\left(\frac{G(\tilde{\mu}(T))}{G(\tilde{\mu}(0))}\right) + \Gamma(T)$$

for some function $G : \bigcup_d \mathbb{R}^d \to \mathbb{R}$ and process of finite variation Γ.

- Here $\log \tilde{\mu}(t) = \log \mu^c(t)$ $\longrightarrow$ differs from the standard setting.

Theorem (I., Larsson, Zhang 2025+)

For a C^2 function F, the portfolio $\pi_i(t) = \sum_k \eta_k(t-) 1_{\{r_i(t-)=k\}}$ which invests

$$\eta_k(t) = \tilde{\mu}_{(k)}(t)\left(\partial_k \log F(\tilde{\mu}_{()}(t)) + \frac{1 - \sum_{\ell=1}^{N(t)} \tilde{\mu}_{(\ell)}(t)\partial_\ell \log F(\tilde{\mu}_{()}(t))}{\sum_{\ell=1}^{N(t)} \tilde{\mu}_{(\ell)}(t)}\right) 1_{\{k \le |I(t)|\}}$$

in the asset at rank k is functionally generated by $G(x) = F(x_{()})$ with

$$d\Gamma(t) = -\frac{1}{2}\sum_{k,\ell=1}^{N(t)} \frac{\partial_{k\ell} F(\tilde{\mu}(t))}{F(\tilde{\mu}(t))} d[\tilde{\mu}_{(k)}, \tilde{\mu}_{(\ell)}](t) - \frac{1}{2}\sum_{k=1}^{N(t)} (\eta_k(t) - \eta_{k+1}(t)) d\Lambda_k(t).$$

Collision rates estimation

THE LONDON SCHOOL OF ECONOMICS AND POLITICAL SCIENCE

- Applying this with the function $F(x) = x_{(1)} + \cdots + x_{(k \wedge d)}$ for $x \in \mathbb{R}^d$ yields that the large-cap portfolio of size k,

$$\pi_i(t) = \frac{\tilde{\mu}_i(t)}{\tilde{\mu}_{(1)}(t) + \cdots + \tilde{\mu}_{(k \wedge N(t))}(t)} 1_{\{r_i(t-) \leq k\}}$$
$$= \frac{\mu_i(t-)}{\mu_{(1)}(t-) + \cdots + \mu_{(k \wedge N(t-))}(t-)} 1_{\{r_i(t-) \leq k\}}$$

has wealth process

$$\log\left(\frac{V^{\mathcal{M}_k}(T)}{V^{\mathcal{M}}(T)}\right) = \log\left(\frac{\tilde{\mu}_{(1)}(T) + \cdots + \tilde{\mu}_{(k \wedge N(T))}(T)}{\tilde{\mu}_{(1)}(0) + \cdots + \tilde{\mu}_{(k \wedge N(0))}(0)}\right) + \frac{\tilde{\mu}_{(k \wedge N(t))}(t)}{\tilde{\mu}_{(1)}(t) + \cdots + \tilde{\mu}_{(k \wedge N(t))}(t)} d\Lambda_k(t),$$

with the convention that $d\Lambda_k(t) = 0$ on $\{|I(t)| < k\}$.

Discretized Estimators

- Old estimator for local time

$$\frac{1}{T}\sum_{i=0}^{M-1}\frac{S_{(1)}(t_i)+\cdots+S_{(k)}(t_i)}{S_{(k)}(t_i)}\log\left(\frac{S_{(1)}(t_i+1)+\cdots+S_{(k)}(t_i+1)}{S_{n_1(t_i)}(t_i+1)+\cdots+S_{n_k(t_i)}(t_i+1)}\right),$$

where $n_k(t)$ is the name occupying the k'th rank at time t.

- New estimator for local time:

$$\frac{1}{Tp(k)}\sum_{i=0}^{M-1}1_{\{|I(t_i)|\geq k\}}\frac{S_{(1)}(t_i)+\cdots+S_{(k)}(t_i)}{S_{(k)}(t_i)}$$
$$\times\log\left(\frac{S_{J_{(1)}^{t_i,t_i+1}}(t_i+1)+\cdots+S_{J_{(k)}^{t_i,t_i+1}}(t_i+1)}{S_{J_{(1)}^{t_i,t_i}}(t_i+1)+\cdots+S_{J_{(k)}^{t_i,t_i}}(t_i+1)}\right),$$

where

- $p(k)=\frac{1}{M}\sum_{i=0}^{M-1}1\{|I(t_i)>k|\}$,
- $J_{(\ell)}^{t,s}$ is the name of of ℓ'th largest market cap based on time s values and out of only the names that are listed at time t and $t+1$.