# Filtering the covariance matrix of non-stationary systems with constant eigenvalues

Christian Bongiorno[1], Damien Challet[1], Grégoire Loeper[2]

[1]CentraleSupélec, Université Paris Saclay

[2]BNP Paribas

15th April 2023

# Problem setting

$R = (r_{i,t})$ data matrix

- True correlation matrix $C$
- $N$ variables, $T$ timesteps
- Standard limit: fix $N$, $T \to \infty$

$$\hat{C} \to C$$

- Curse-of-dimensionality limit:

$$N, T \to \infty, \qquad N/T = q > 0$$

$$\hat{C} \sim P(\hat{C}), \text{ noisy}$$

# Why $q > 0$: non-stationarity

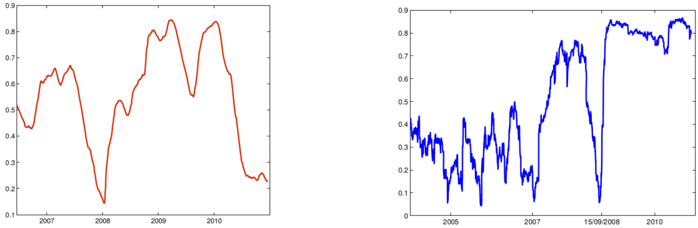



Figure 1: Left: EWMA estimator of average pairwise correlations of daily returns in EuroStoxx 50 index. Right: one year EMWA correlation between two ETF of the the S&P 500: SPDR XLE (energy) and SPDR XLK (technology)

[Cont and Wagalath (2013)]

# Noise: what to filter?

Spectral Decomposition of $C$

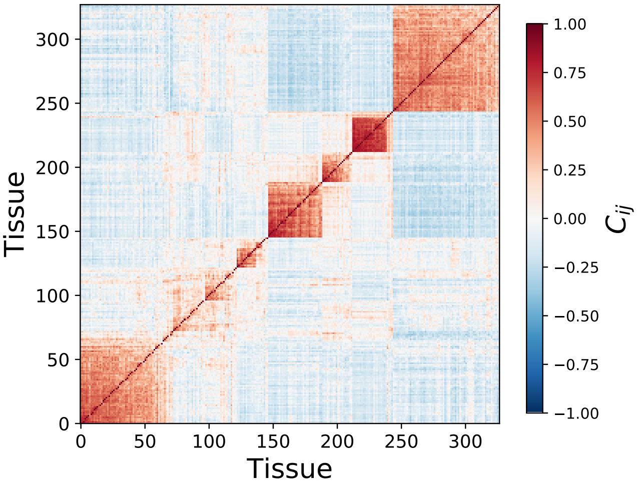
$$C = V^{\dagger} \Lambda V$$
$$\hat{C} = \hat{V}^{\dagger} \hat{\Lambda} \hat{V}$$

where

- $V$: eigenvectors matrix ($N \times N$)
- $\Lambda = \mathrm{diag}(\lambda_1, \ldots, \lambda_N)$: $N$ eigenvalues

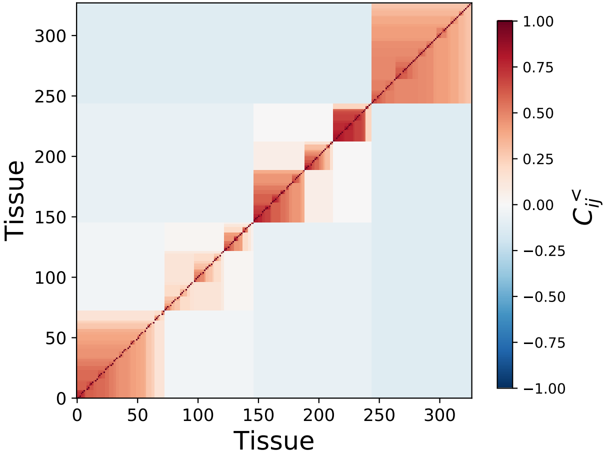
$$\lambda_1 \geq \cdots \geq \lambda_N \geq 0$$

# Filter $\hat{V}$ and $\hat{\Lambda}$: ansatz

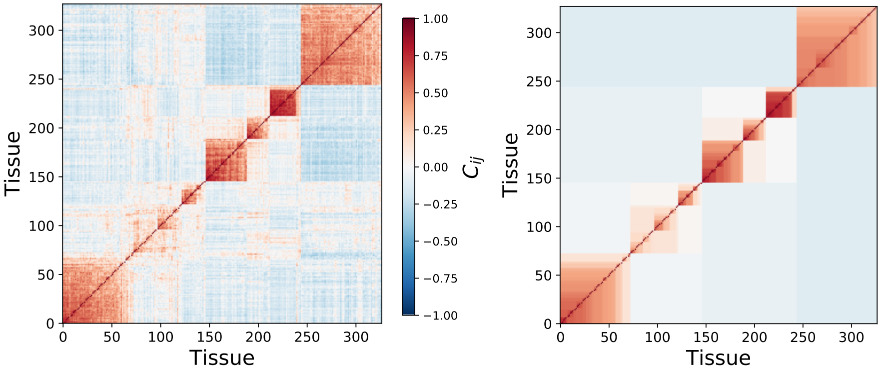


# Hierarchical ansatz

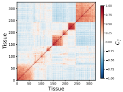
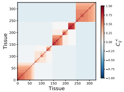
Hierarchical Clustering Average Linkage (Tumminello, Lillo, Mantegna 2007)

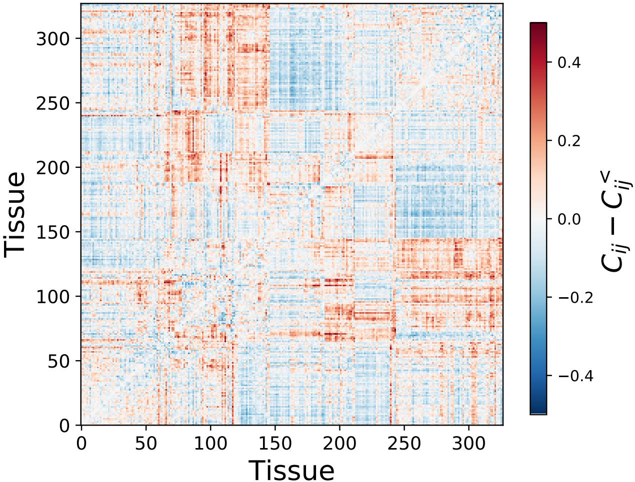


# HCAL: gene expression



## Are dependencies strictly hierarchical?

Residuals:  $-$  $=$



# Problem: hierarchical clustering is fragile

Bongiorno, Micciché, Mantegna (2022) [link]

- Build bootstraps of data matrix
- Compute hierarchical clustering
- The tree structure is not robust

# Solution: break and average HCAL

Bongiorno and Challet [2021] [2023]

1. Bootstrap times of $R$ $B$ times

$$R \to R^{(b)} \quad b = 1, \cdots, B$$

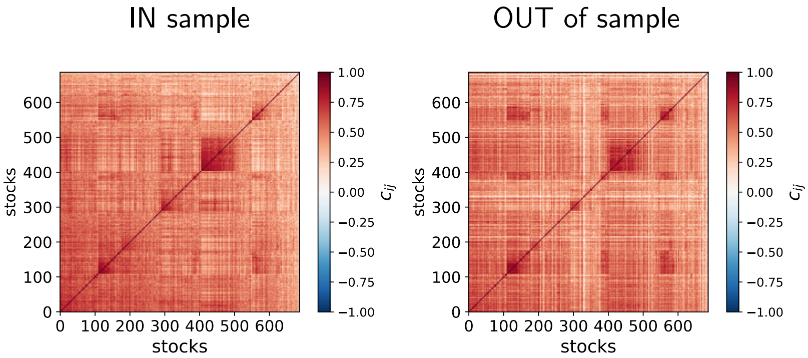
2. Apply HCAL to each bootstrap

$$C^{(b)} \to C^{(b)<}$$

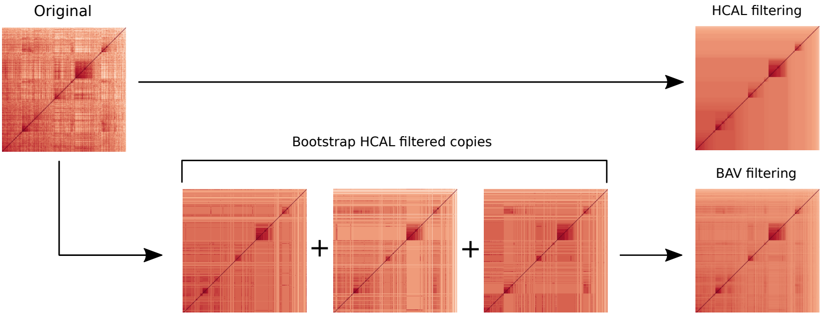
3. Average

$$C^{BAHC} = \frac{1}{B} \sum_{b=1}^{B} C^{(b)<}$$

# Example: US equity markets



# BAHC=⟨HCAL⟩




- Good eigenvectors
- Sub-optimal eigenvalues

# What is a good covariance estimator?

Filtered matrix $\Xi$ minimises some distance with true matrix

- $C$ known, one-shot context

$$||\Xi - C||$$

- $C$ unknown, in-sample + out-of-sample windows

$$||\Xi - E_{out}||$$

Example: option basket, portfolio optimization (finance, etc)

1. What is a good distance?
2. How to build an optimal estimator?

# A formal approach to correlation/covariance cleaning

- True correlation $C$

- Estimated correlation $E = \hat{C}$

- Wishart (1928) [link]

$$P(E|C) = \frac{T^{NT/2}}{2^{NT/2}\Gamma_N(T/2)} \frac{\det(E)^{\frac{T-N-1}{2}}}{\det(C)^{T/2}} e^{-\frac{T}{2}\mathrm{Tr} C^{-1} E}$$

- Special case: $C = I_N$

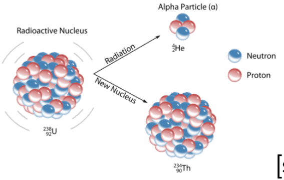
$$P(E|C) \propto \det(E)^{\frac{T-N-1}{2}} e^{-\frac{T}{2}\mathrm{Tr} E}$$

$\rightarrow$ eigenvalues only

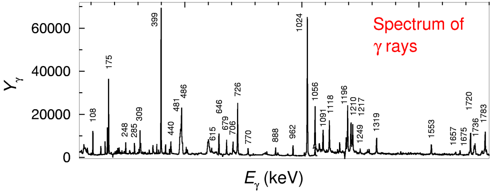
- Reciprocal question

$$P(C|E)?$$

# Random matrix theory (RMT)




[source]

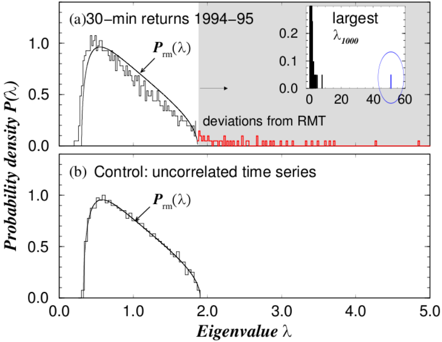



[source]

- Energy levels: eigenvalues of the energy operator
- Wigner:
  - Energy separation $\delta E$ is so complex is looks random.
  - $E[P(\delta E)]$,and $E[P(E)]$: universal for a given matrix family

# Empirical spectrum of correlation matrices

Laloux et al. (1999), Plerou et al. (2002)




- 98% of eigenvalues from random bulk: RMT
- Large eigenvalues not from finite $N$

# Rotationally Invariant Estimators

- Hyp: no information about $C$, thus no privileged structure of eigenvectors

  $\rightarrow$ keep eigenvectors $\hat{V}$
  $\rightarrow$ tweak the eigenvalues

- RI estimator

$$\Xi = \sum_i \xi(\lambda_i) v_i^\dagger v_i$$

# RIE: linear shrinkage

Ansatz

$$C(\alpha) = (1 - \alpha^*)E + \alpha^* F$$

where $F$ is a target matrix

Ledoit and Wolf (2004) give

$$\alpha^* = \arg\min_{\alpha} ||C - C(\alpha)||^2$$

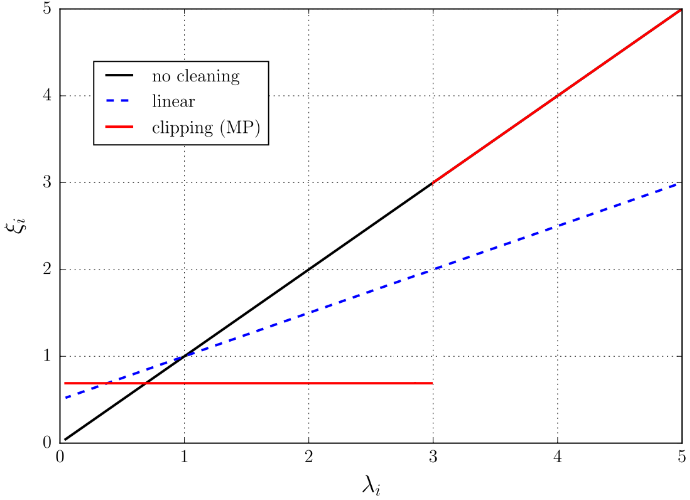
Special case:

$$F = I_N$$

Eigenvalues:

$$\xi(\lambda_i) = (1 - \alpha)\lambda_i + \alpha$$

# Empirical, clipped and shrunk eigenvalues



# Optimal RIE

- Oracle eigenvalues: use true correlation matrix

$$\Lambda^{Oracle} = \mathrm{diag}\left(\hat{V}^{\dagger} C \hat{V}\right)$$

- Optimal RIE: replace $\hat{\Lambda}$ by $\Lambda^{Oracle}$
- Eigenvalues of $C$ from those of $E$?

# Link between $E$ and $C$

Marčenko-Pastur equation:

$$z\mathfrak{g}_E(z) = Z(z)\mathfrak{g}_C[Z(z)]$$
$$Z(z) = \frac{z}{1 - q + qz\mathfrak{g}_E(z)}$$
$$z \in \mathbb{C}$$

where

$$\mathfrak{g}_M(z) = \frac{1}{N}\mathrm{Tr}\sum_i \frac{v_i' v_i}{z - \lambda_i}$$

is the Stieltjes transform of $M$ and $(\{v\}, \{\lambda\})$ eigenvectors and eigenvalues of matrix $M$

# Optimal *stationary* RIE: NLS

Define

$$\hat{\xi}(\lambda) = \frac{\lambda}{|1 - q + q\lambda \lim_{\eta \to 0^+} \mathfrak{g}_E(\lambda - i\eta)|^2}$$

Note that $\hat{\xi}$ independent from $C$

Miracle:

$$\lim_{N/T=q, N, T\to\infty,} \hat{\xi} \to \xi^{oracle}$$

if

1. $C$ is constant
2. finite 4-th/12-th moment
3. infinite systems

# Some limits

- Small $\lambda$

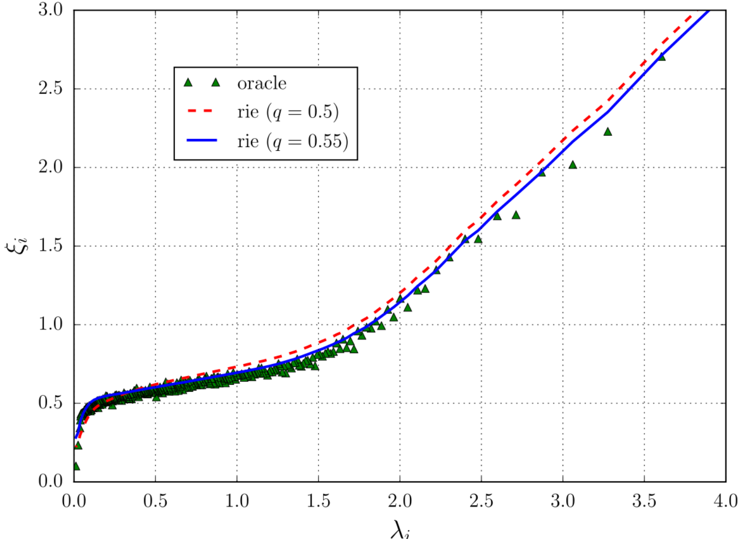
$$\hat{\xi}(\lambda) = \frac{1}{(1-q)^2} + O(\lambda^2) > 0$$

- Large $\lambda$

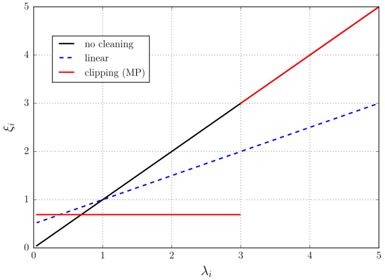
$$\hat{\xi}(\lambda) \simeq \lambda - 2q + O(1/\lambda) : \quad \text{affine}$$

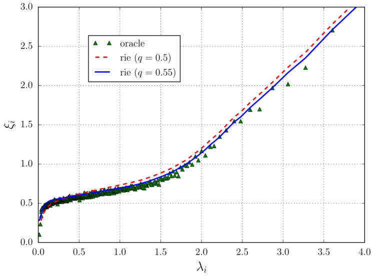
# Graphically

From Bun *et al.* (2017) [link]



# Graphically



# What about non-stationary systems?

E.g. financial markets:

$$\Lambda^{Oracle} = \text{diag}\left(V_{in}^{\dagger} C_{out} V_{in}\right)$$

DL says (Bongiorno, Challet, Loeper [2023]):

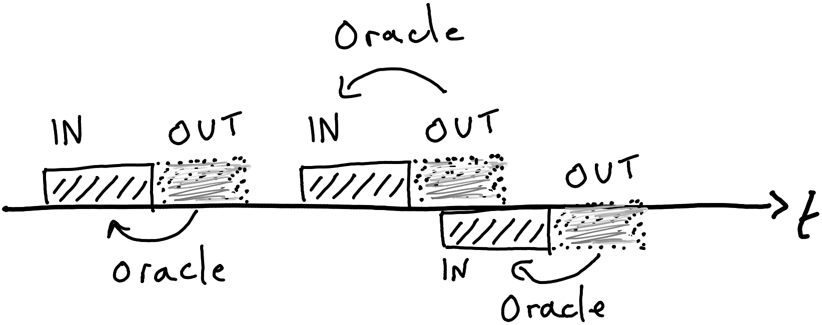
- Fix $N$ and $T$
- In many subperiods, average

$$\text{diag}\left(V_{in}^{\dagger} C_{out} V_{in}\right)$$

$\rightarrow$ Average Oracle

# How to compute the AO

Average over many IN-OUT time intervals



# How to compute the AO: calibration window

Average over many IN-OUT time intervals

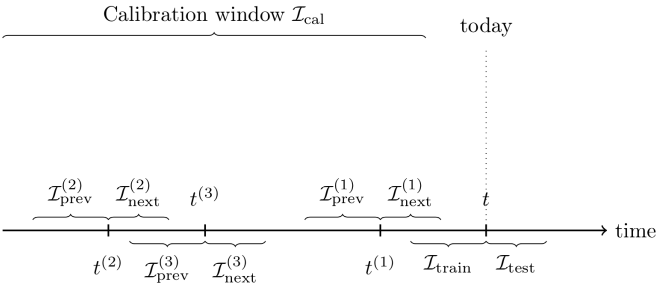



Figure 1: Average Oracle eigenvalues computation: Oracle eigenvalues are computed in many sub-intervals of a long calibration window and then averaged rank-wise. These eigenvalues can then be used outside of the calibration window.

# Average Oracle

- Fix $N$, $\delta t_{in}$, $\delta t_{out}$
- In calibration window, repeat $B \gg 1$ times:
  - random choice of periods
  - (random choice of assets)
  - compute $\Lambda_b^{Oracle}$, $\quad b = 1, \cdots, B$
- Average Oracle: rank-wise

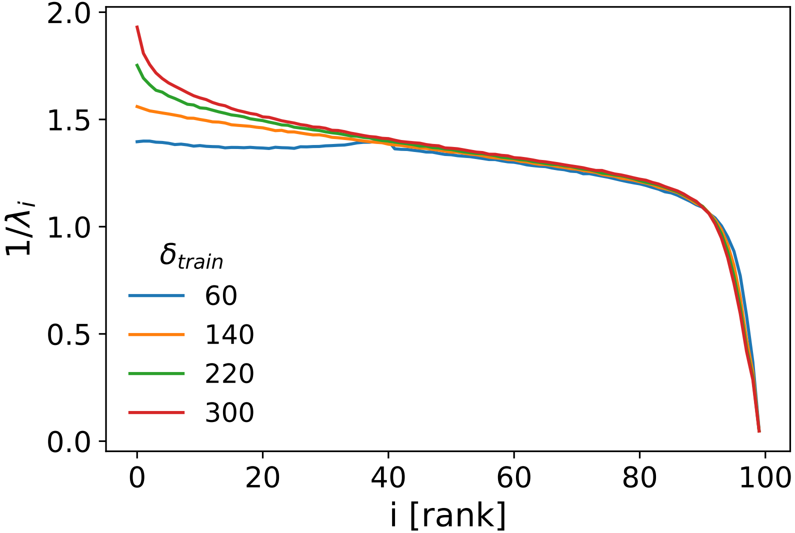
$$\lambda_k^{AO} = \frac{1}{B} \sum_{b=1}^{B} \lambda_k^{Oracle}$$

# Average Oracle

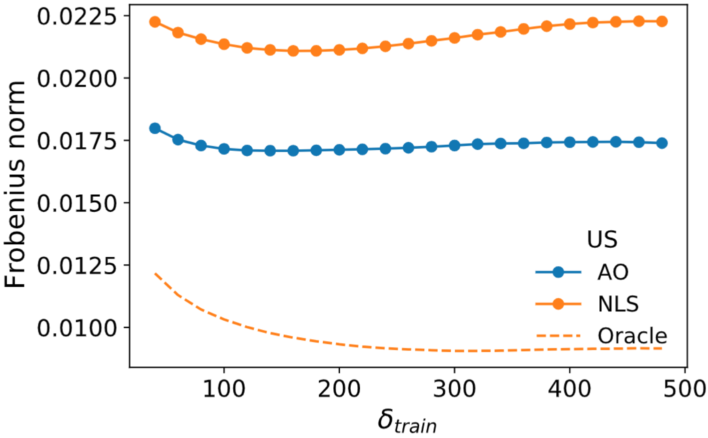
$$\Lambda^{AO} = E\left[\text{diag}\left(V_{in}^{\dagger} C_{out} V_{in}\right)\right]$$

- Average influence of future over past eigenvalues
- Average influence time-evolution of $C$
- $\hat{\Lambda}$ is irrelevant to eigenvalue cleaning (zero-th order approximation)

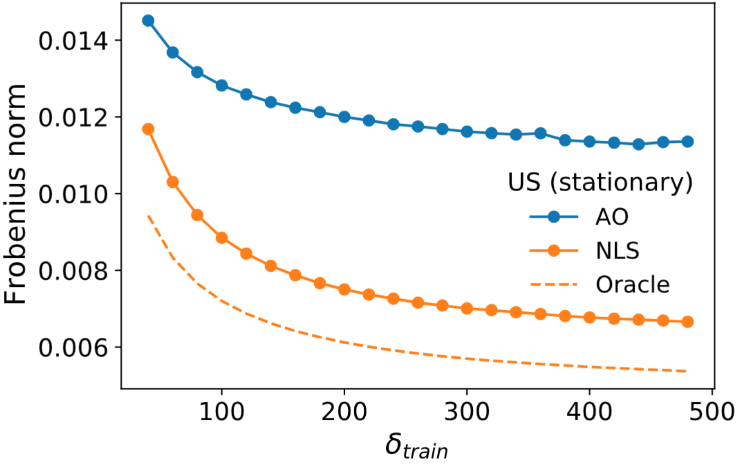
# Average Oracle eigenvalues



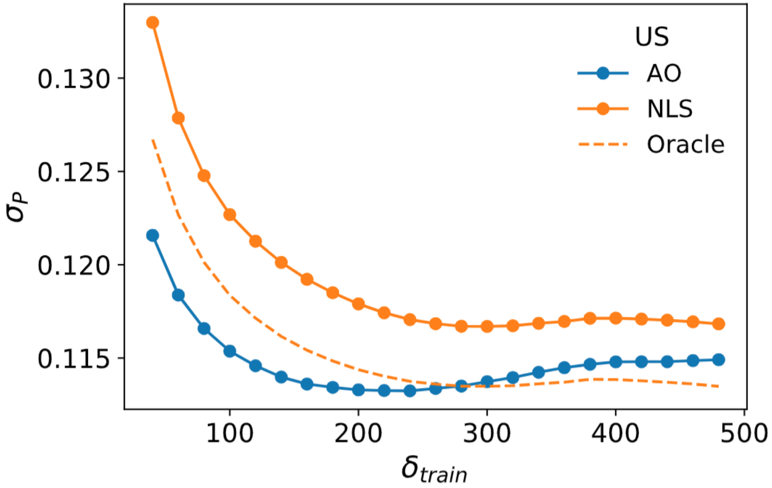
# US equity data $N = 100$



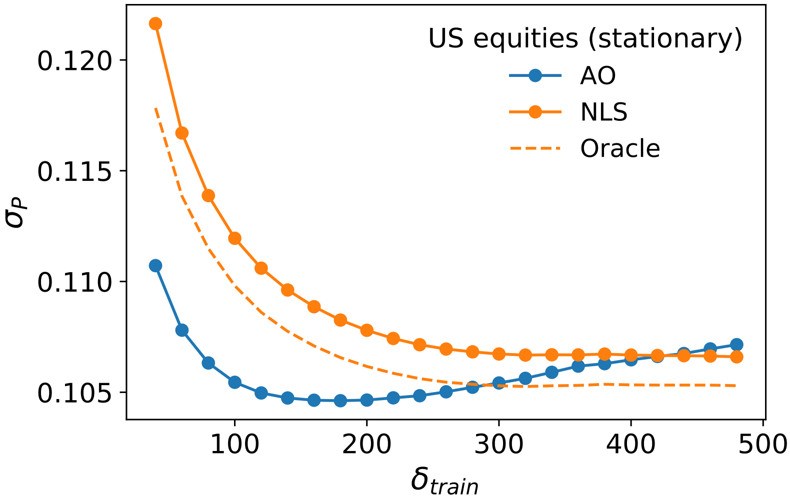
# Results: stationarized (shuffled) returns $N = 100$



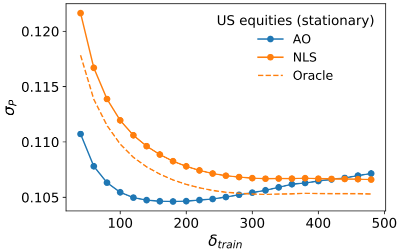
# Results: Global Minimum Variance portfolios $N = 100$



## Global Minimum Variance portfolios (stationarized data)
## $N = 100$



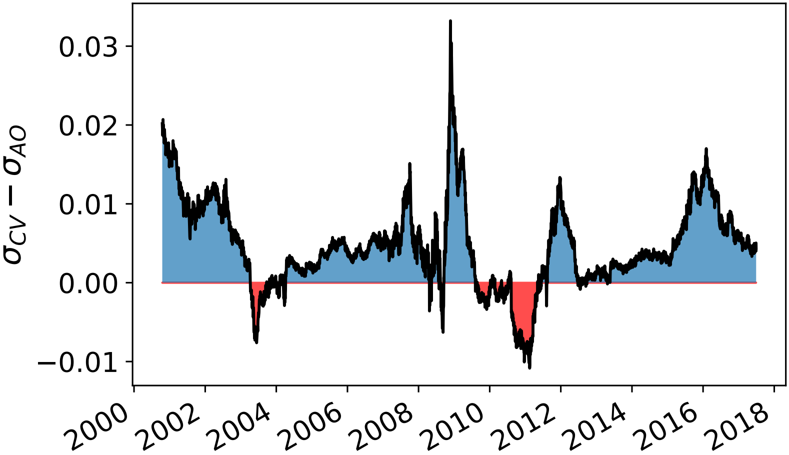
# Global Minimum Variance portfolios (stationarized data) $N = 100$




- AO better than *Oracle*
- GMV not equivalent to minimum of Frobenius norm

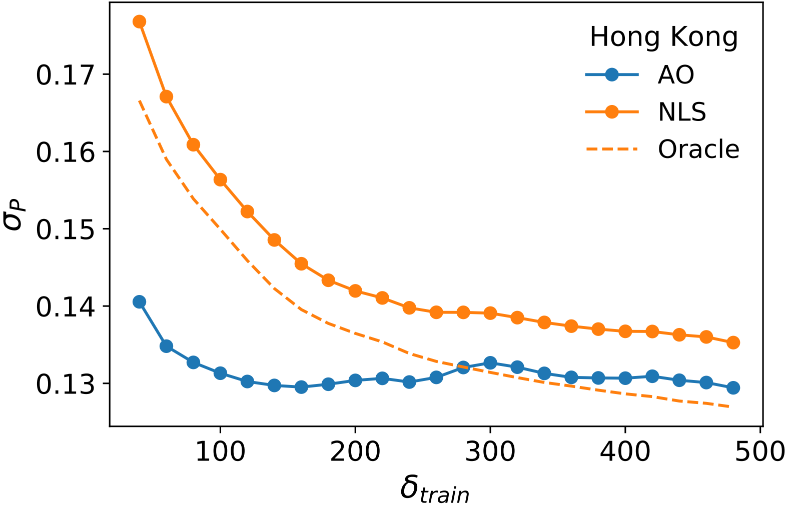
B&Ch (2022)
[arxiv 2112.07521]

# AO vs NLS: time evolution



# Transfer learning

- Calibrate AO on US data
- Apply it on HK data ($N = 100$ )

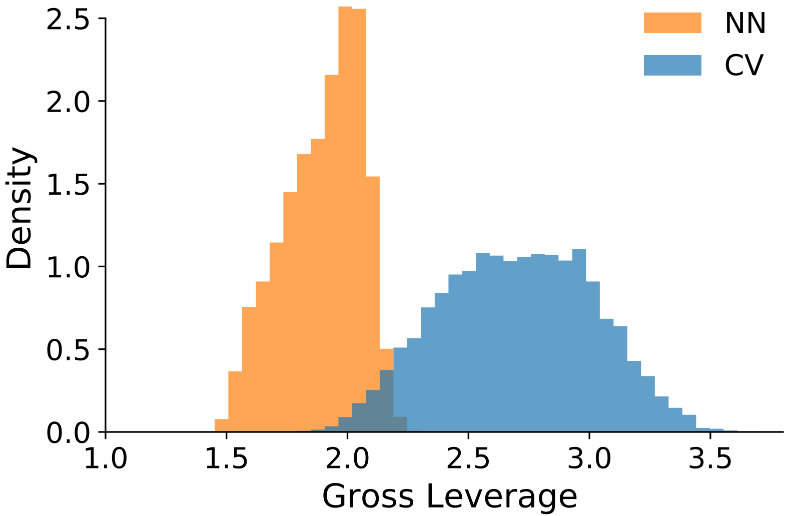


# Why AO works

1. Less noise
2. Non-stationarity
3. Heavy tails
4. Finite size systems

# 1. Less noise

$N = 100,\ t_{in} = 300$



# 2. Non-stationarity

- Decomposition

$$\Lambda_{Oracle} = \left(V_{\mathtt{train}}^{\dagger} V_{\mathtt{test}}\right)^{\circ 2} \Lambda_{\mathtt{test}} = O^{\circ 2} \Lambda_{\mathtt{test}}$$

  $O$ : rotation matrix from $V_{\text{train}}$ to $V_{\text{test}}$
  $O^{\circ 2} = (O_{ij}^2)$: overlaps between eigenvectors

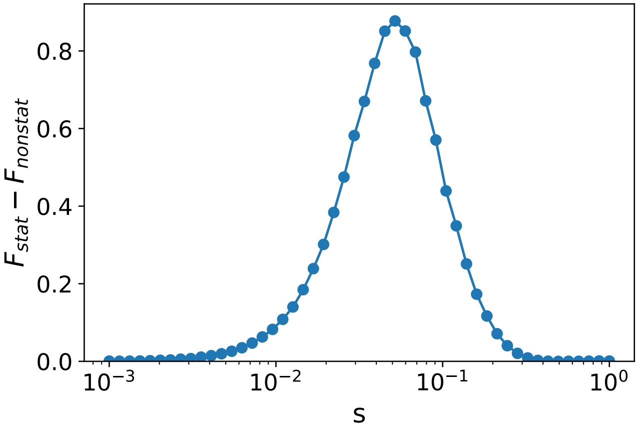
- fact

$$E(O^{\circ 2} \Lambda_{\mathtt{test}}) \simeq E(O^{\circ 2}) E(\Lambda_{\mathtt{test}})$$

- Influence of $O^{\circ 2}$ on Frobenius norm?

# Simple non-stationary model

- Fixed eigenvalues, random eigenvector rotations, amplitude $s$
- Compute $E(O^{\circ 2})$ from model and also from stationarized data from the model



# AO vs NLS

- Alternating world

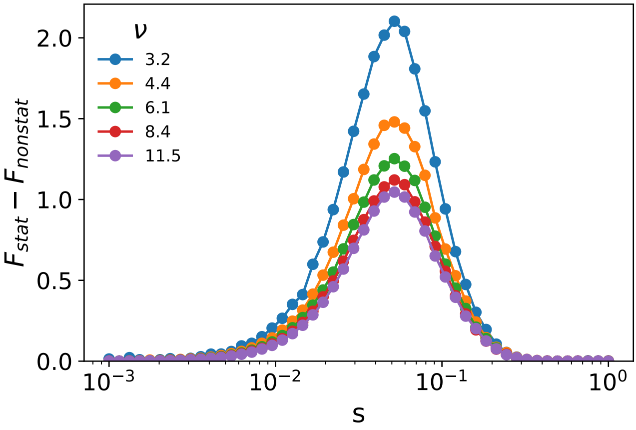
$$C(k) = C_0, C_1, C_0, C_1$$

- NLS $\rightsquigarrow$ wrong matrix
- AO $\rightsquigarrow$ average matrix
- General idea if $C(k+1) = C_1$

$$||\Xi^{NLS} - C_1|| \rightsquigarrow ||C_0 - C_1|| \geq ||\Xi^{AO} - C_1|| \rightsquigarrow ||\frac{C_0 + C_1}{2} - C_1||$$

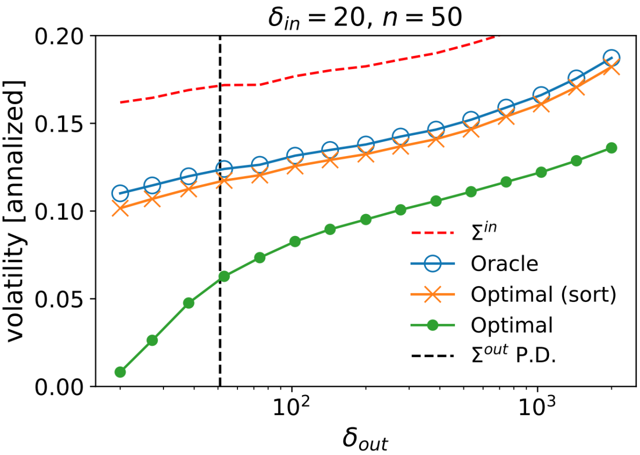
# 3. Role of heavy tails

Factor model with Student t-distributions, tail exponent $\nu$



# What is a good distance?

- Calculus: Frobenius, Kullback-Leibler
- Portfolio: not Frobenius (Bongiorno and Challet [2023])



# Conclusions

- Average Oracle recipe: just plug in the AO eigenvalues
- AO: better than "optimal" RIE for non-stationary systems
- AO: zero-th order correction of non-stationarity
- TODO:
  1. write a paper about AO vs DCC+NLS
  2. maths: AO with estimation noise