Covariance Matrix Cleaning

Dynamical Covariances

Christian Bongiorno | Allocation de Portefeuille | 02-03-2023

HC Filtering - Previous Assignment

```
>>> Z = average(y)
>>> Z
array([[ 0.        ,  1.        ,  1.        ,  2.        ],
       [ 3.        ,  4.        ,  1.        ,  2.        ],
       [ 6.        ,  7.        ,  1.        ,  2.        ],
       [ 9.        , 10.        ,  1.        ,  2.        ],
       [ 2.        , 12.        ,  1.20710678,  3.        ],
       [ 5.        , 13.        ,  1.20710678,  3.        ],
       [ 8.        , 14.        ,  1.20710678,  3.        ],
       [11.        , 15.        ,  1.20710678,  3.        ],
       [16.        , 17.        ,  3.39675184,  6.        ],
       [18.        , 19.        ,  3.39675184,  6.        ],
       [20.        , 21.        ,  4.09206523, 12.        ]])
```

each row is a new clade

index clade-1 index clade-2 height

HC Filtering - Previous Assignment

```python
def dist(R):
    N = R.shape[0]
    d = R[idx]
    out = fastcluster.average(d)
    rho = 1-out[:,2]
    #
    #Genealogy Set
    dend = [([i],[]) for i in range(N)]+[[] for _ in range(out.shape[0])]

    for i,(a,b) in enumerate(out[:,:2].astype(int)):
        dend[i+N] = dend[a][0]+dend[a][1],dend[b][0]+dend[b][1]

    return dend[N:],rho
```

dend is a dictionary that contains the pairs of leaf nodes that join into the clade (keys)

rho is the average correlation coefficient that characterizes the clade

```python
def AvLinkC(Dend,rho,R):

    N = R.shape[0]
    Rs = np.zeros((N,N))

    for (a,b),r in zip(Dend,rho):
        z = np.array(a).reshape(-1,1),np.array(b).reshape(1,-1)
        Rs[z] = r

    Rs = Rs+Rs.T

    np.fill_diagonal(Rs,1)
    return Rs
```

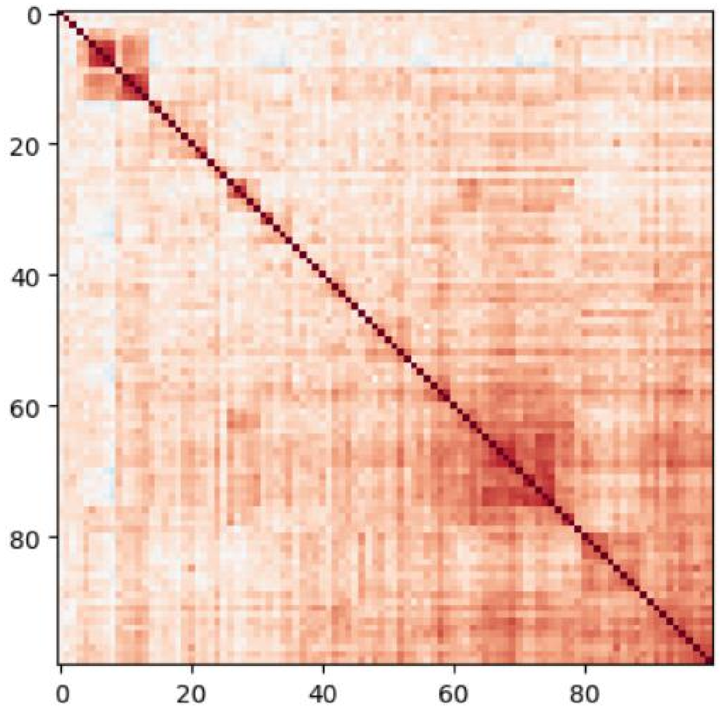
For all the clades, it creates the indices of the rectangular sub-matrix and associates the average coefficient.

Ordering or Rows/Columns- Previous Assignment

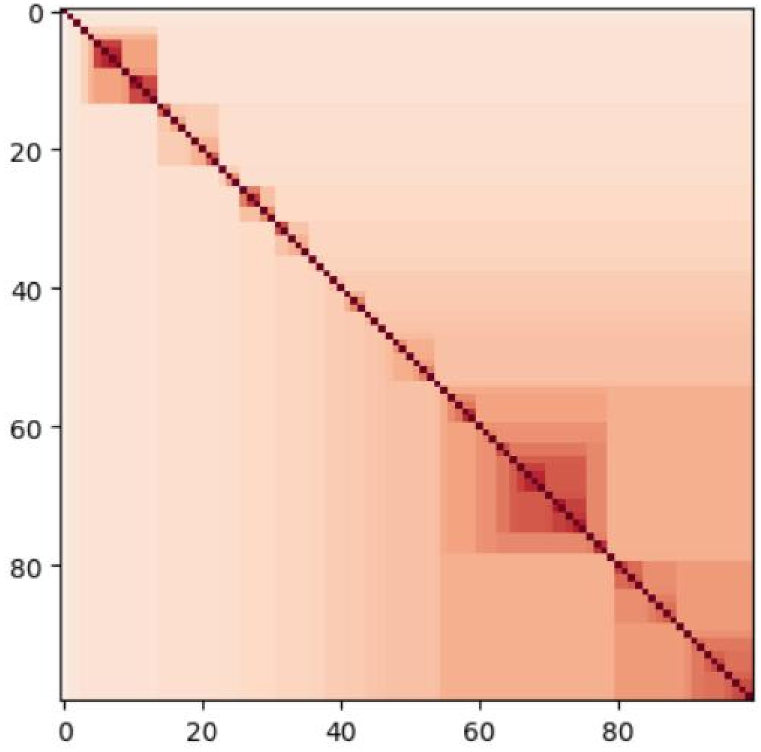
```
dend = average(D)

ordn = leaves_list(dend)
```

```
plt.imshow(R[ordn][:,ordn],vmin=-1,vmax=1,cmap=cm.RdBu_r);
```



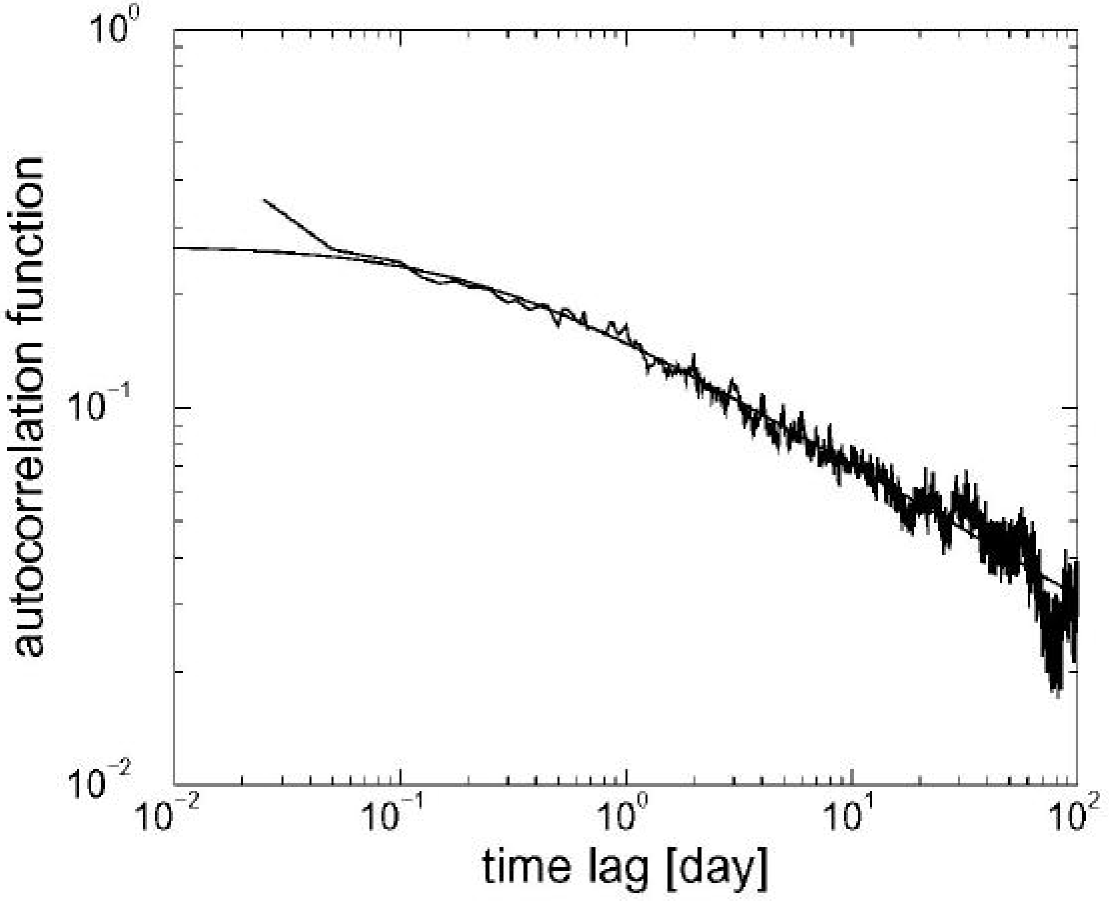

```
plt.imshow(Rf[ordn][:,ordn],vmin=-1,vmax=1,cmap=cm.RdBu_r);
```



Topics of Today

- We discuss the most popular method to model univariate variances
- We introduce the dynamical modeling of the covariance/correlation matrix
- I will present the DCC method with filtering
- I will show new results on the dynamical evolution of correlation matrices

Power-law Decay of the Volatility ACF



GARCH model

*GARCH stands for **G**eneralized **A**uto-**R**egressive **C**onditional **H**eteroskedasticity model*

$$r_t = \mu_t + \varepsilon_t$$
$$\varepsilon_t = \sigma_t z_t$$
$$\sigma_t^2 = \alpha_0 + \sum_{i=1}^{q} \alpha_i \varepsilon_{t-i}^2 + \sum_{i=1}^{p} \beta_i \sigma_{t-i}^2$$

The unconditional variance for a GARCH(1,1) is simply

$$\sigma^2 = \frac{\alpha_0}{1 - \alpha_1 - \beta_1}$$

This stationarity condition of the unconditional variance holds if

$$0 < \alpha_1 + \beta_1 < 1$$

Note that the ML estimation for the GARCH is very complex; you must check for correct convergence and changes the non-linear optimizer in case of a failure. This is particularly relevant when the time-series is not sufficiently large.

GARCH(1,1)

Most econometricians believe that the GARCH(1,1) is the most appropriate model for describing volatility.

The main idea is that the most recent information should be more important in the forecast.

GARCH(1,1):

$$r_t = \mu_t + \varepsilon_t$$
$$\varepsilon_t = \sigma_t z_t$$
$$\sigma_t^2 = \alpha_0 + \alpha \varepsilon_{t-1}^2 + \beta \, \sigma_{t-1}^2$$

In practical application, the drift can be set to zero or come from an ARMA model; both cases have a very limited empirical motivation.

DCC with Shrinkage

Large Dynamic Covariance Matrices

Robert F. Engle
Department of Finance, New York University, New York, NY 10012 (*rengle@stern.nyu.edu*)

Olivier Ledoit
Department of Economics, University of Zurich, CH-8032 Zurich, Switzerland (*olivier.ledoit@econ.uzh.ch*);
AlphaCrest Capital Management, New York, NY 10036 (*olivier.ledoit@alphacrestcapital.com*)

Michael Wolf
Department of Economics, University of Zurich, CH-8032 Zurich, Switzerland (*michael.wolf@econ.uzh.ch*)

Second moments of asset returns are important for risk management and portfolio selection. The problem of estimating second moments can be approached from two angles: time series and the cross-section. In time series, the key is to account for conditional heteroscedasticity; a favored model is Dynamic Conditional Correlation (DCC), derived from the ARCH/GARCH family started by Engle (1982). In the cross-section, the key is to correct in-sample biases of sample covariance matrix eigenvalues; a favored model is nonlinear shrinkage, derived from Random Matrix Theory (RMT). The present article marries these two strands of literature to deliver improved estimation of large dynamic covariance matrices. Supplementary material for this article is available online.

KEY WORDS: Composite likelihood; Dynamic conditional correlation; GARCH; Markowitz portfolio selection; Nonlinear shrinkage.

DCC with Shrinkage

The idea is to model a conditional correlation matrix $\mathbf{Q}_t$, not directly observed on data, and must be fitted with a complex likelihood maximization. Such equation as an unconditional (long-term expectations) correlation matrix **C**. The unconditional correlation **C** can be the sample correlation shrinked with CV (Non-linear shrinkage).

$$Q_t = (1 - \alpha - \beta)C + \alpha\, \mathbf{s}_{t-1}\mathbf{s}'_{t-1} + \beta\, Q_{t-1},$$

$\mathbf{s}_t$ are the devolatized returns, which are the demean returns divided by the conditional volatility of a GARCH(1,1).

$\mathbf{Q}_t$ is sometimes called a pseudo-correlation matrix since the diagonal is not exactly one. This can be corrected, and the conditional covariance Ht can be obtained by multiplying by the last conditional volatility of a GARCH(1,1).

$$R_t := \mathrm{Diag}(\mathrm{Q_t})^{-1/2}\, \mathrm{Q_t}\, \mathrm{Diag}(\mathrm{Q_t})^{-1/2}$$

$$H_t := D_t R_t D_t,$$

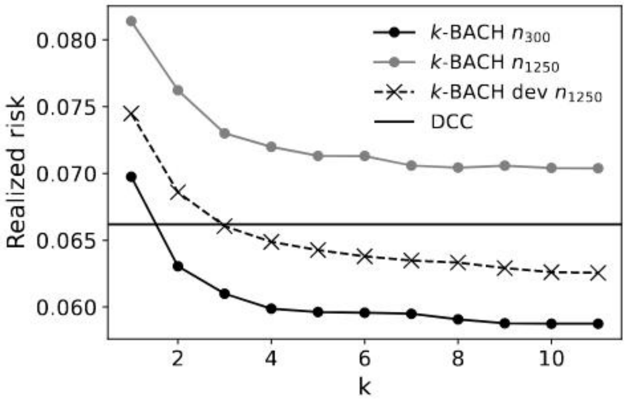
Problems with GARCH and DCC

Although the methods can be written in an elegant mathematical form, there is insufficient evidence that the actual time series work like that.

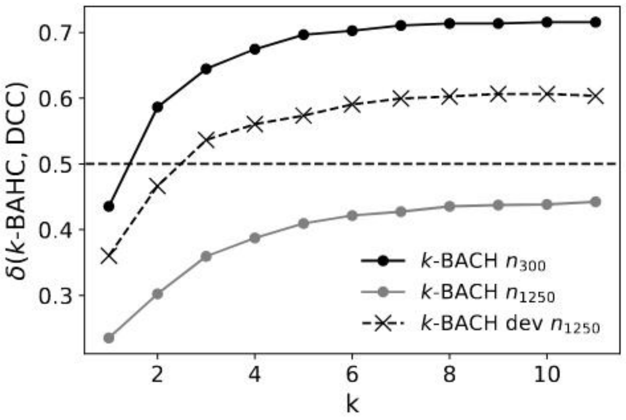
The calibration is very challenging:

- It is time costly;
- it requires long calibration windows (typically **5 years**);
- the parameters are strongly dependent on small perturbations of the data;
- sometimes the parameter optimization does not converge.

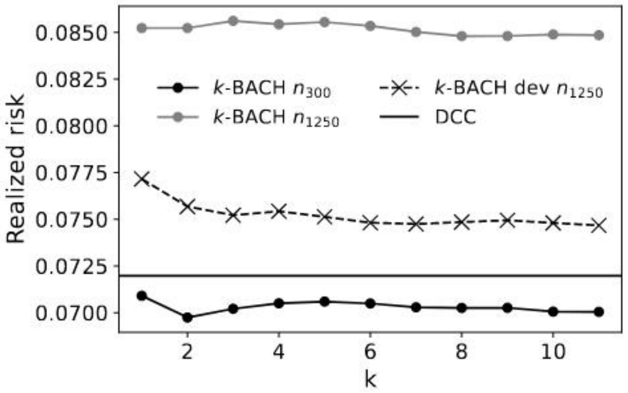
Limitations of a Long Time-Horizon



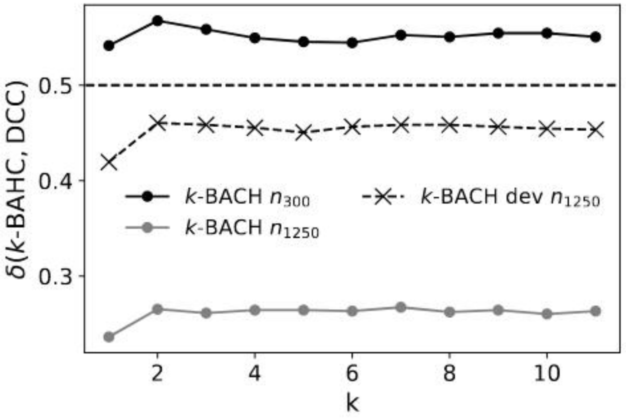
(a) Long-short realized risk.



(b) Long-short fraction of times k-BAHC performs.



(c) Long-only realized risk.



(d) Long-only fraction of times k-BAHC performs.

<u>Most of the stocks have a short-life time.</u>

Using a long time window implies reducing the number of stocks you consider in your universe.

Fewer stocks = Fewer Diversification Opportunities

In the plot, we compared the reduction of realized risk using the largest amount of stocks available using:

DCC on 1250 days.
k-BACH on 1250 days
k-BACH on 1250 days + return devolatization
k-BACH on 300 days (more stocks considered)

Bongiorno, C., & Challet, D. (2022). Reactive global minimum variance portfolios with k-BAHC covariance cleaning. The European Journal of Finance, 28(13-15), 1344-1360

Exponential Weighting

Another approach that can be used to give more importance to recent observations is exponential weighting. Although it is not grounded in econometrics models, it performs pretty well empirically. It requires short calibration windows and does not have convergence problems.

$$E := \frac{1-\beta}{1-\beta^T} \sum_{t=1}^{T} \beta^{T-t} x_t x_t'.$$

The exponentially weighted covariance matrix requires only the estimation of the parameter β.

$$W_{t,t} = T\frac{1-\beta}{1-\beta^T}\beta^{T-t},$$

It can be rearranged in a weighting matrix that allows to computes a weighted returns matrix.

$$\tilde{X} := W^{1/2}X.$$

On the weighted return matrix, one can apply CV or anything covariance cleaning approaches.

Tan, V. W., & Zohren, S. (2020). Large Non-Stationary Noisy Covariance Matrices: A Cross-Validation Approach. arXiv preprint arXiv:2012.05757.

New Routes on Dynamical-Covariance Matrices

Bongiorno, C., Challet, D., & Loeper, G. (2023). Filtering time-dependent covariance matrices using time-independent eigenvalues. Journal of Statistical Mechanics: Theory and Experiment, 2023(2), 023402.

A Time-Independent Eigenvalue Set (?!)

For $N = 25, T = 252$ days, you can use these numbers as eigenvalue correction

```
array([0.5  , 0.562, 0.595, 0.619, 0.638, 0.652, 0.665, 0.677, 0.688,
       0.7  , 0.709, 0.72 , 0.729, 0.739, 0.748, 0.76 , 0.772, 0.784,
       0.801, 0.818, 0.846, 0.887, 0.995, 1.283, 7.114])
```

In our experiments, they outperform the optimal methods, for every time period, for every selection of stocks, and every country!

A Short Recall on the Oracle Estimator

More recently, the focus was on the optimal eigenvalue set Λ_O, known as the oracle.

Λ_O minimizes the Frobenius norm with the out-of-sample covariance/correlation

$$||\Xi(\Lambda_O) - \Sigma_{\text{test}}||_F$$

The oracle can be obtained analytically

$$\Lambda_O = \text{diag}(V_{\text{train}}^{\dagger}\Sigma_{\text{test}}V_{\text{train}})$$

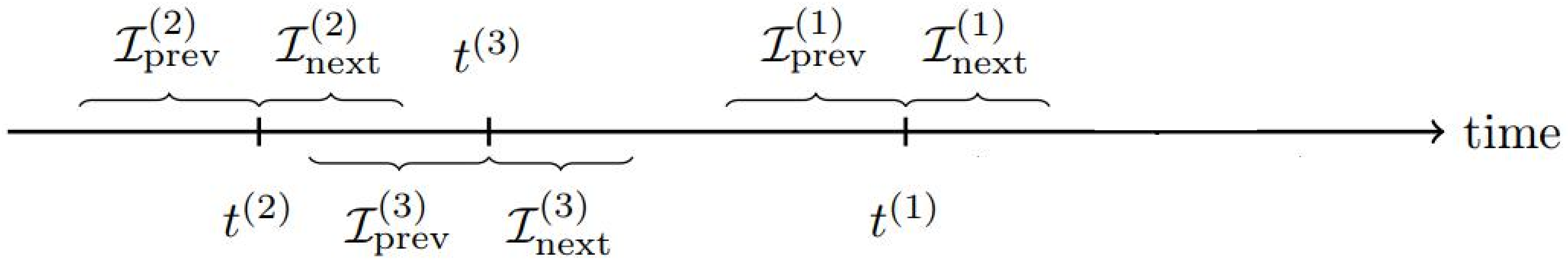
Of course must be regarded as the optimal target, but not practically useful since it requires knowledge of the future.

$$\Xi(\Lambda_O) = V_{\text{train}}\Lambda_O V_{\text{train}}^{\dagger}$$

Time-Preserving Time Intervals

We can create a list of train/test calibration windows; in these cases, we call them prev/next, which preserves the temporal order of the data.

Differently from the CV train/test split, this time, prev/next might be substantially different (not only sampling noise) if the data carries some temporal dependence.



Overlap Matrix and Time-Invariance

The Oracle eigenvalues can be also written as

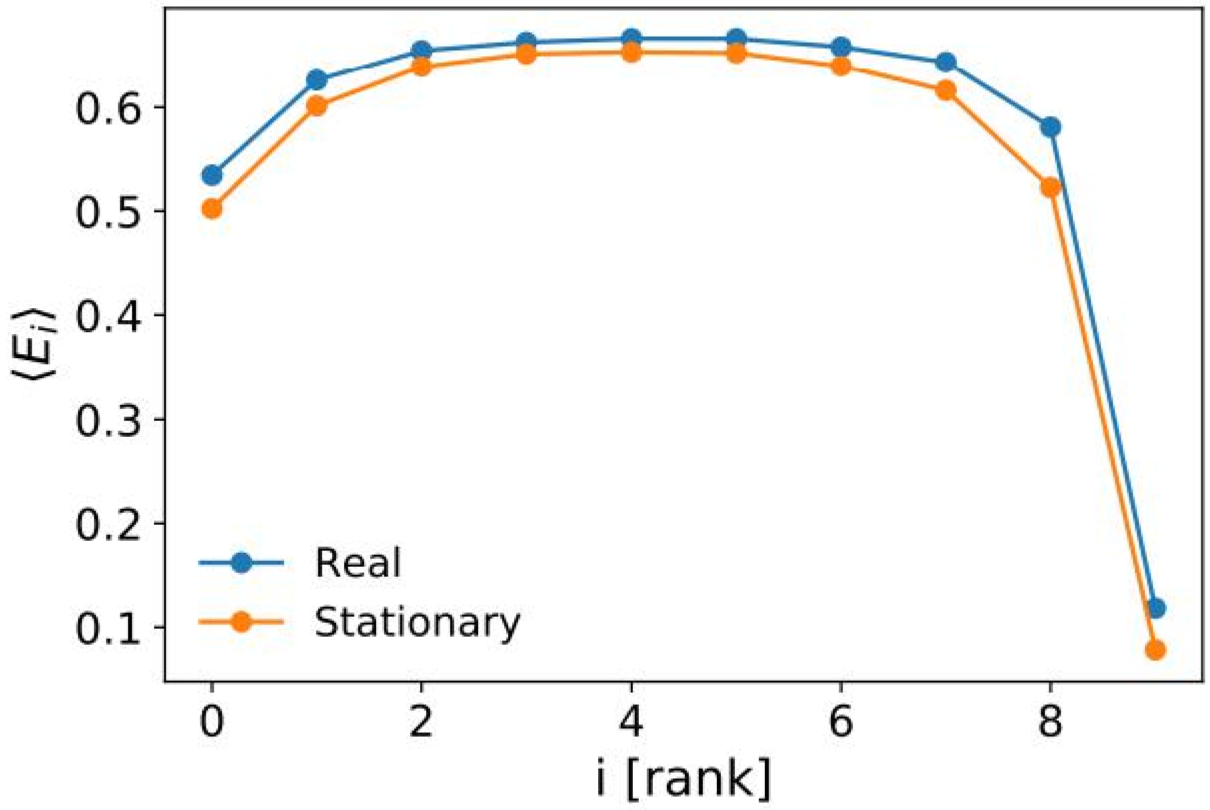
$$\lambda_O = \left(V_{\text{prev}}^{\dagger} V_{\text{next}}\right)^{\circ 2} \lambda_{\text{next}} = H^{\circ 2} \lambda_{\text{next}}$$

H is a rotation matrix from the past's eigenvector basis into the future's eigenvector basis.

$$E_i = -\sum_{j=1}^{n} h_{ij}^2 \log_n h_{ij}^2$$

$H^{\circ 2}$ is a bi-stochastic matrix, i.e., the sum by row and column is one.

Overlap Matrix and Time-Invariance



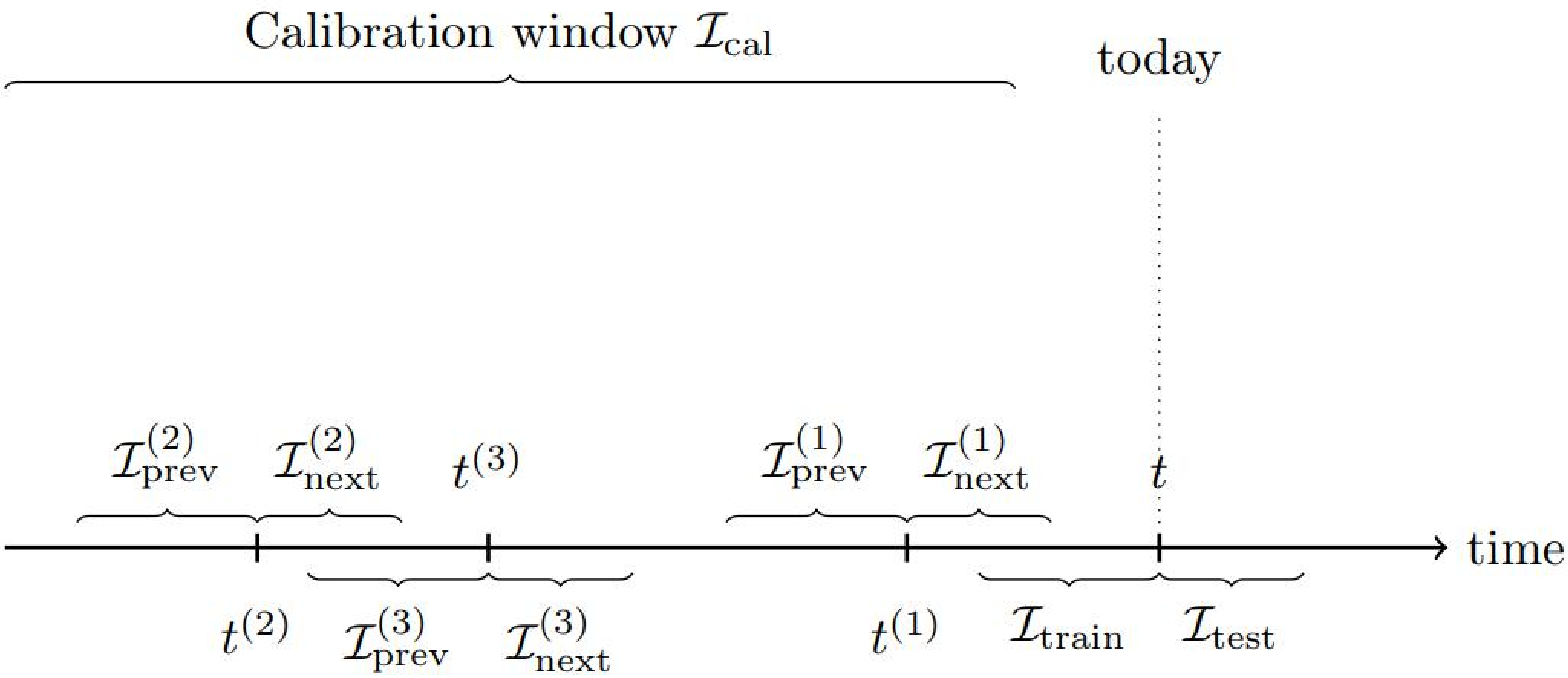
The stationary version is obtained with a temporal shuffling of the data.

This is an example of a 10x10 matrix.

Lower entropy on the overlap matrix row indicates a lower one-to-one overlap of the prev/next eigenvectors.

In summary, the eigenvectors varymore than what we should expect only from sampling noise

Calibration Set Up



Average Oracle Definition

We propose an eigenvalues shrinkage that encodes the average temporal evolution from prev to next

The average oracle estimator is defined from

$$\Lambda_O^{(b)} = \mathrm{diag}({V_{\mathrm{prev}}^{(b)}}^{\dagger} \Sigma_{\mathrm{next}}^{(b)} V_{\mathrm{prev}}^{(b)}).$$

$$\Lambda_{AO} := \frac{1}{B} \sum_{b=1}^{B} \Lambda_O^{(b)}$$

$$\Xi(\Lambda_{AO}) = V_{\mathrm{train}} \Lambda_{AO} V_{\mathrm{train}}^{\dagger}$$

The estimator must be adjusted with our best predictor of the univariate variance to be a covariance matrix

$$(\Sigma_{AO})_{ij} = \sqrt{(\Sigma_{\mathrm{train}})_{ii}}\, \Xi(\Lambda_{AO})_{ij} \sqrt{(\Sigma_{\mathrm{train}})_{jj}}$$

A Simple Benchmark

The evolution of the eigenvector might be described by a time-varying rotation matric H_t (note that Vt is not directly measurable)

$$V_{t+1} = V_t H_t$$

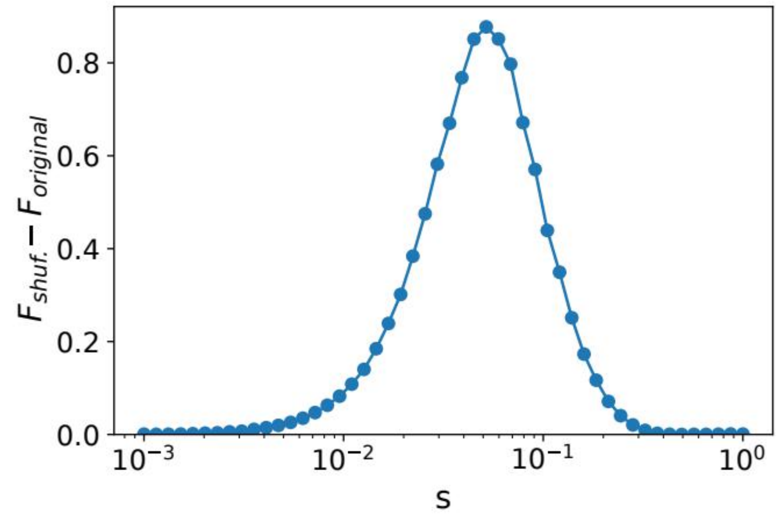
H_t can be decomposed in n(n-1)/2 Euler plane rotations

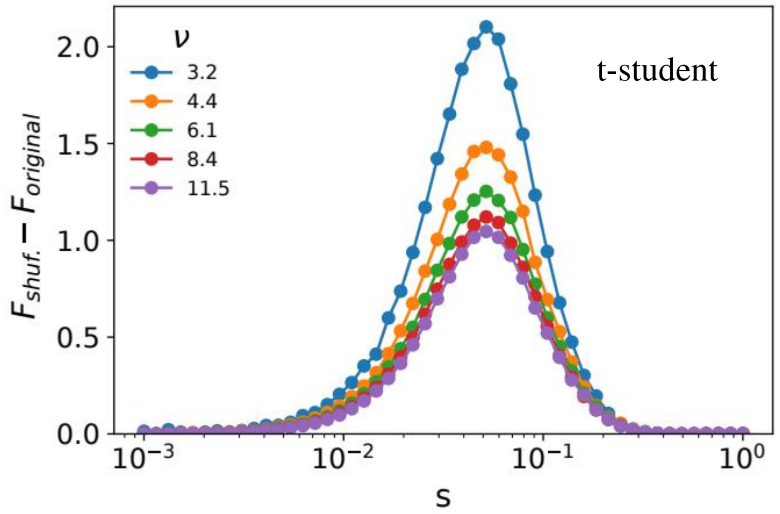
$$H_t = H(\alpha_{1,t}) H(\alpha_{2_t}) \cdots H(\alpha_{n(n-1)/2,t})$$

We randomly sample each rotation angle from a normal with standard deviation s.

$$\alpha_{i,t} \sim \mathcal{N}(0, s^2)$$

By shuffling the original data, we can measure the impact of the temporal dependence in the Frobenious norm for different rotation speed.



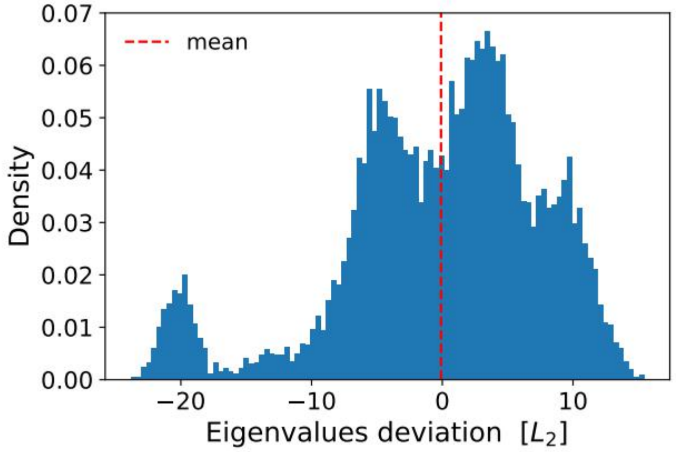


Stability of the Eigenvalues

$$D_{\bullet}^{(1)} = \sum_{i=1}^{n} |\lambda_{i,\bullet} - \lambda_{i,\text{next}}|$$

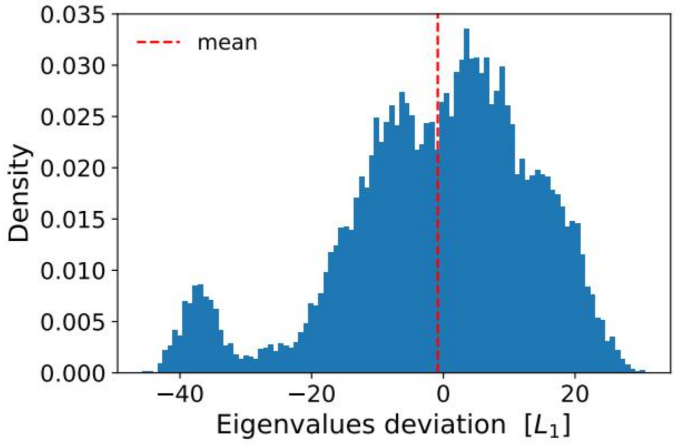
$$D_{\bullet}^{(2)} = \sqrt{\sum_{i=1}^{n} (\lambda_{i,\bullet} - \lambda_{i,\text{next}})^2}$$

The predictability of the current eigenvalues is not systematically better on the whole spectrum than an average set.



(a)

$D^{(2)}_{\langle\text{next}\rangle}$-$D^{(2)}_{\text{prev}}$



(b)

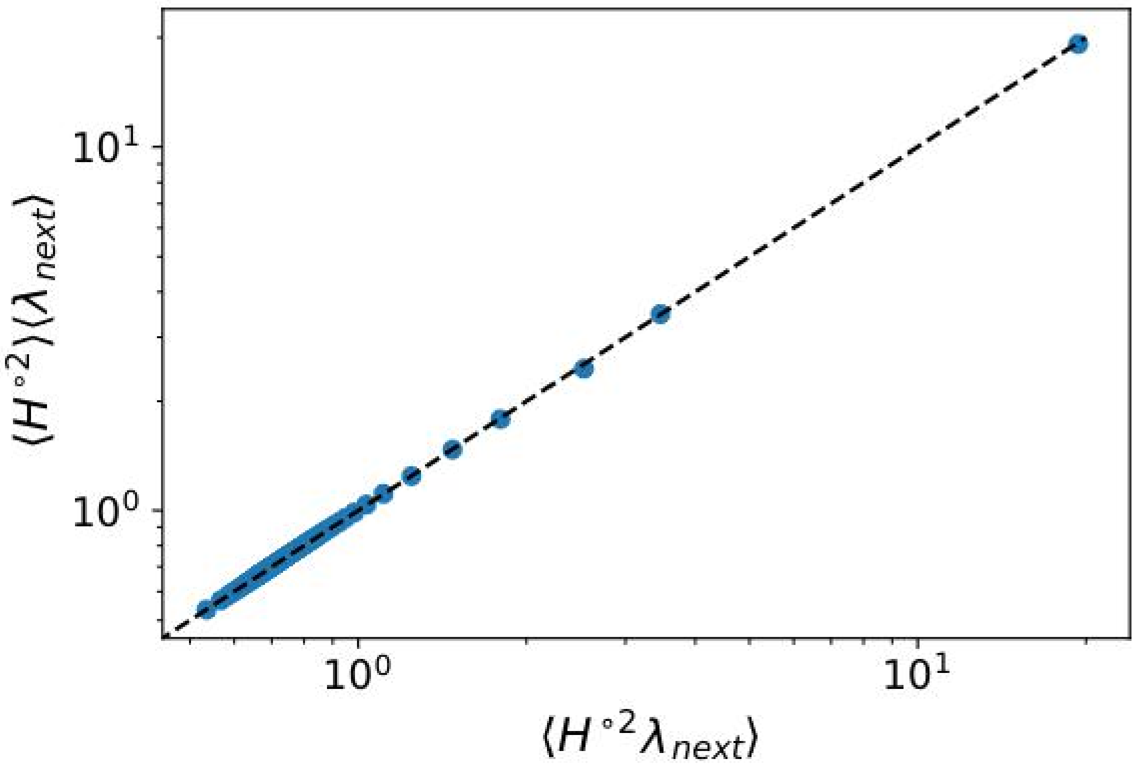
$D^{(1)}_{\langle\text{next}\rangle}$-$D^{(1)}_{\text{prev}}$

Decomposition of the Average Oracle Eigenvalues

$$\langle H^{\circ 2}\lambda_{\mathrm{next}}\rangle \simeq \langle H^{\circ 2}\rangle\langle\lambda_{\mathrm{next}}\rangle.$$

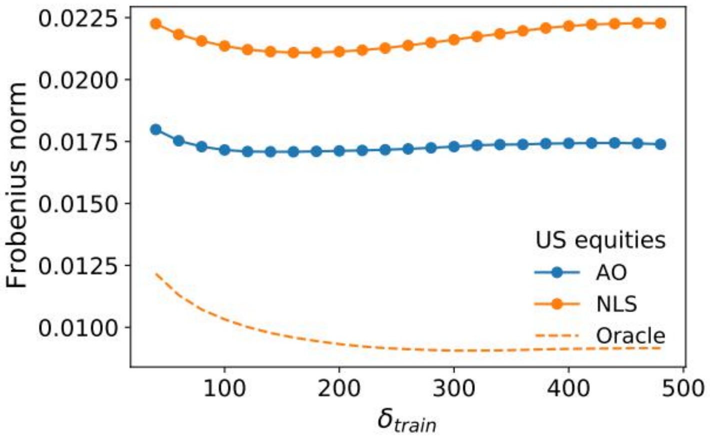
Average Oracle

Average Overlap Matrix
Stability component

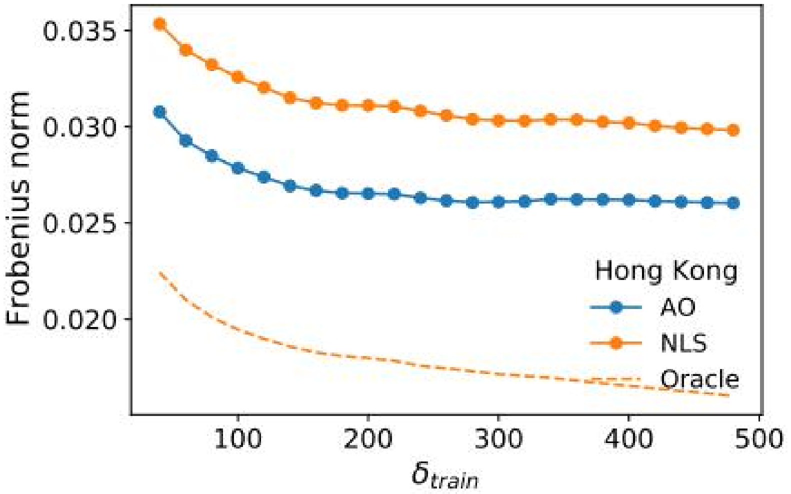


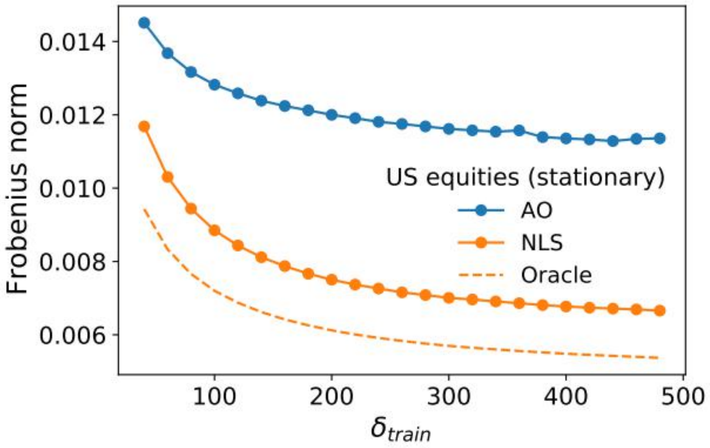
Observed statistical independence.

Frobenious Norm Distance

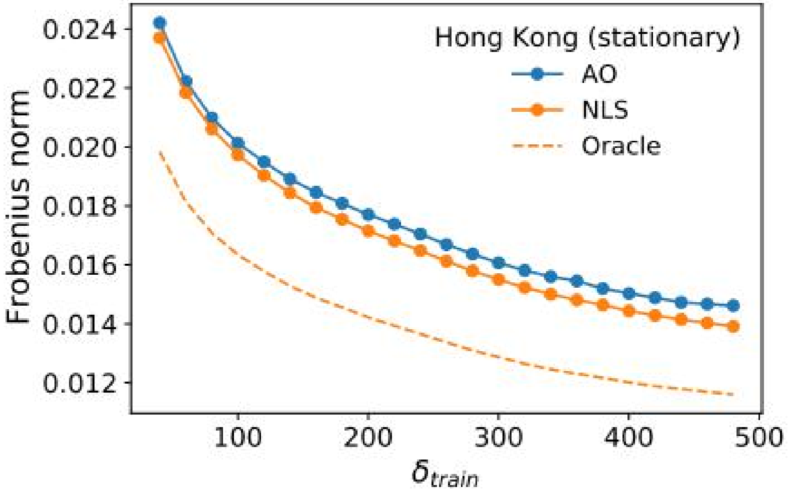


(a)



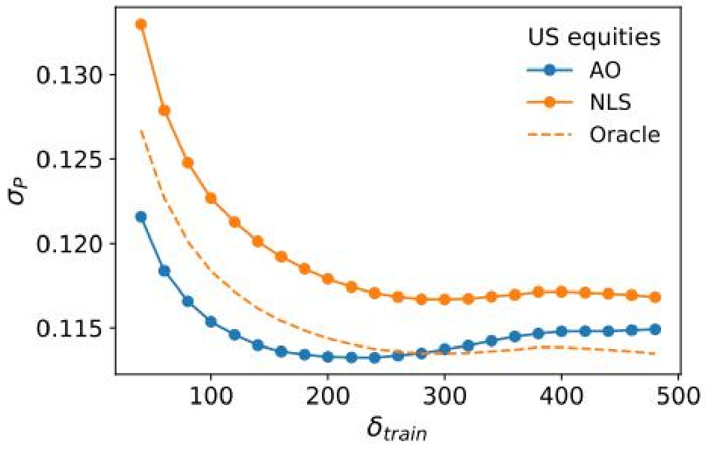


(b)

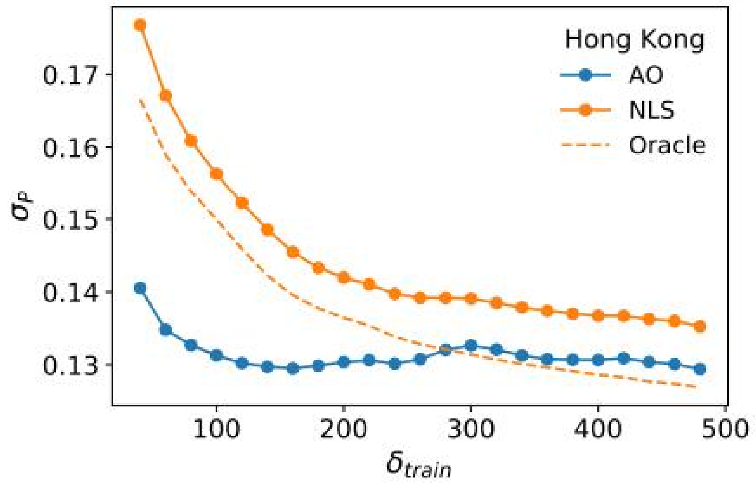


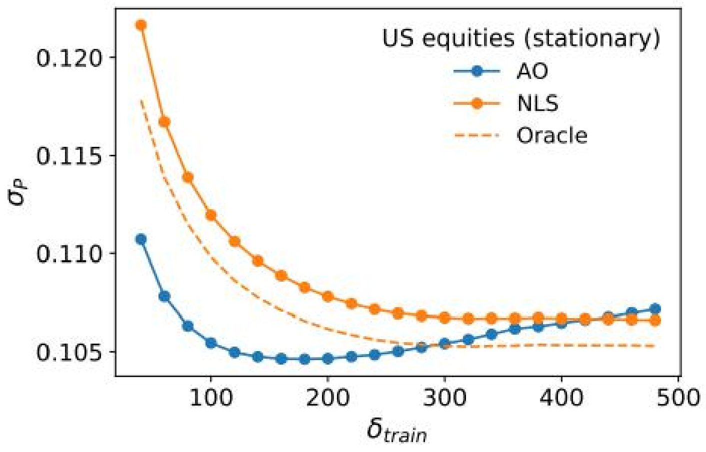
Note that the eigenvalues tested on Hong Kong were calibrated on US equities

Volatility of Global Minium Variance Portfolio

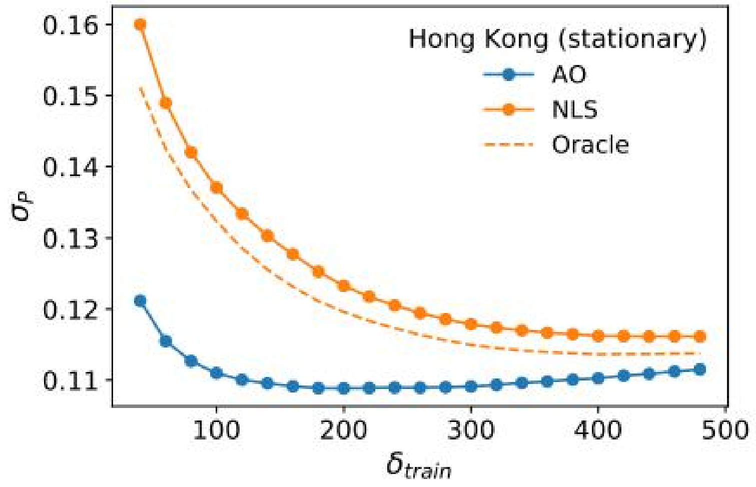


(a)





(b)



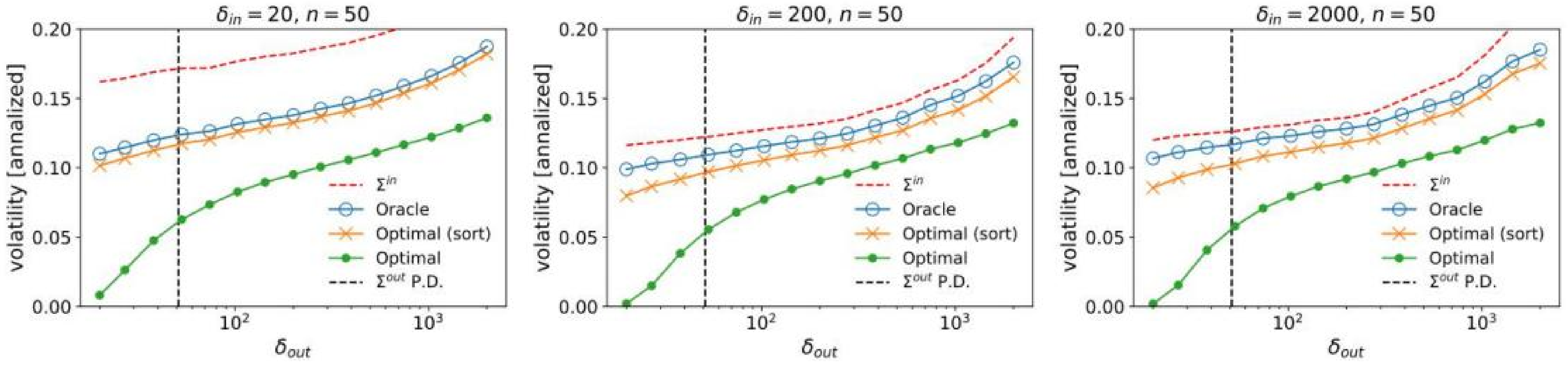
It might seem surprising that the GMV portfolio of **AO also beat the Oracle**.

However, minimizing the Frobenious norm is not exactly equal to minimizing the future risk.

The Oracle Estimator is Suboptimal for GMV Porfolios

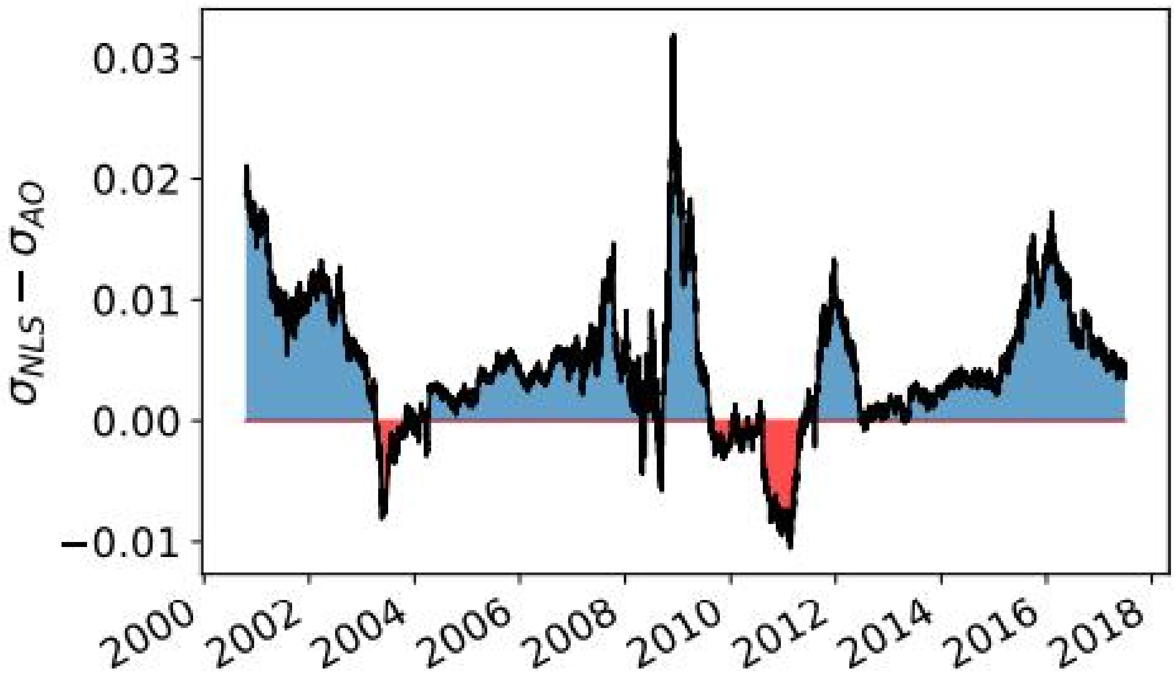
$$
\begin{array}{lll}
\min_{\mathbf{w},\Xi^{-1},\zeta} & \sum_{i=1}^{n} w_i \Sigma^{\text{out}}_{ij} w_j & \\
\text{subject to} & w_k = \sum_{j=1}^{n} \Xi^{-1}_{kj}; & k = 1, \cdots, n \\
& \sum_{k=1}^{n} w_k = 1 & \\
& \Xi^{-1}_{ij} = \sum_{k=1}^{n} \zeta_k v^{\text{in}}_{ik} v^{\text{in}}_{jk}; i, j = 1, \cdots, n & \\
& \zeta_k \geq 0; & k = 1, \cdots, n.
\end{array}
$$

$$\zeta_k := \frac{1}{\lambda^*_k}$$



Bongiorno, C and Challet. D "Non-linear shrinkage of the price return covariance matrix is far from optimal for portfolio optimisation." Finance Research Letters (2022):

Systematic Reduction of the Portfolio Volatility



Turnover

Now we can introduce the re-balance of the portfolios.

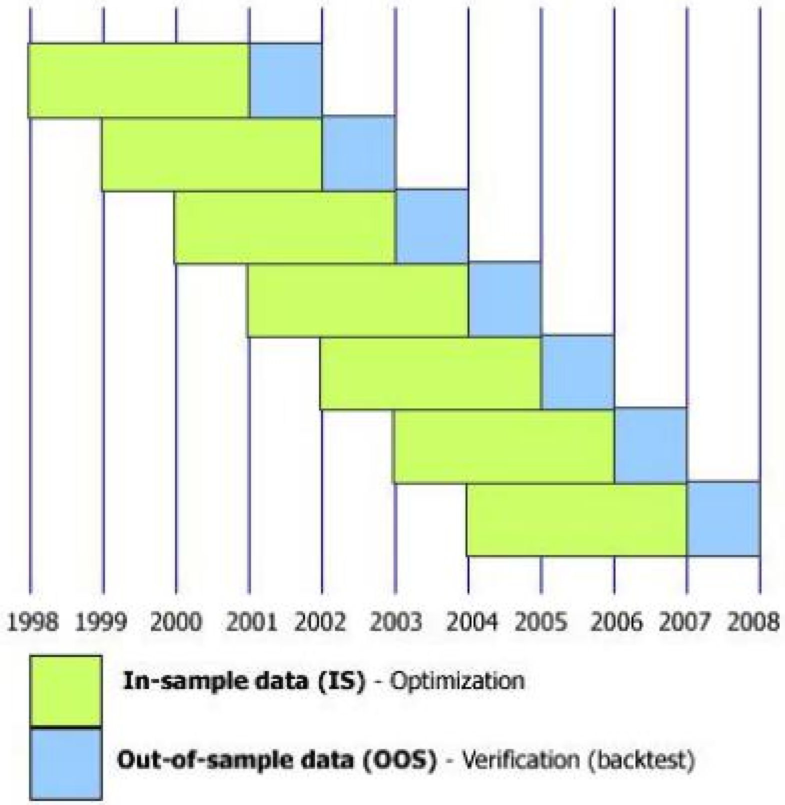
You should select and in-sample and out-of-sample size window size, find your portfolio based on in-sample data, and leave the composition fixed over the successive out-of-sample window.
Then find another portfolio and do it the same.

If the portfolio changes drastically you will pay higher transaction costs.

Let us consider to do $\Delta\tau$ re-balancing, that means to have $\Delta\tau$ out-of-sample windows of size Δt days

A natural metric for turnover[1] is:

$$\Upsilon = \frac{1}{\Delta\tau}\sum_{\tau=1}^{\Delta\tau-1}\sum_{i=1}^{N}\left|w_i^{(\tau)} - w_i^{(\tau+1)}\right|$$



[1] Reigneron, P. A., Nguyen, V., Ciliberti, S., Seager, P., & Bouchaud, J. P. (2019). The case for long-only agnostic allocation portfolios. arXiv:1906.05187.

A Comparison of the Methods Performances

	SR	MEAN	VOL	Turnover	GrossLev	N_{eff}
NotFilt1200	1.101	0.098	0.113	0.174	2.765	6.368
NotFilt240	0.745	0.071	0.125	0.974	4.151	3.290
AO1200	**1.229**	**0.112**	0.114	**0.068**	**1.702**	**13.981**
AO240	1.167	0.095	**0.105**	0.181	1.729	13.213
DCC-QIS	0.894	0.069	**0.105**	1.308	2.543	5.856
DCC-QuEST	0.895	0.069	**0.105**	1.307	2.542	5.852
QIS1200	1.123	0.098	0.112	0.15	2.536	7.931
QuEST1200	1.122	0.098	0.112	0.156	2.531	7.992
QIS240	1.022	0.083	0.108	0.427	2.538	9.675
QuEST240	1.027	0.084	0.108	0.402	2.512	10.011
1AFM-DCC-QIS	1.071	0.089	0.108	0.807	2.615	6.089
1AFM-DCC-QuEST	1.073	0.089	0.108	0.803	2.612	6.096

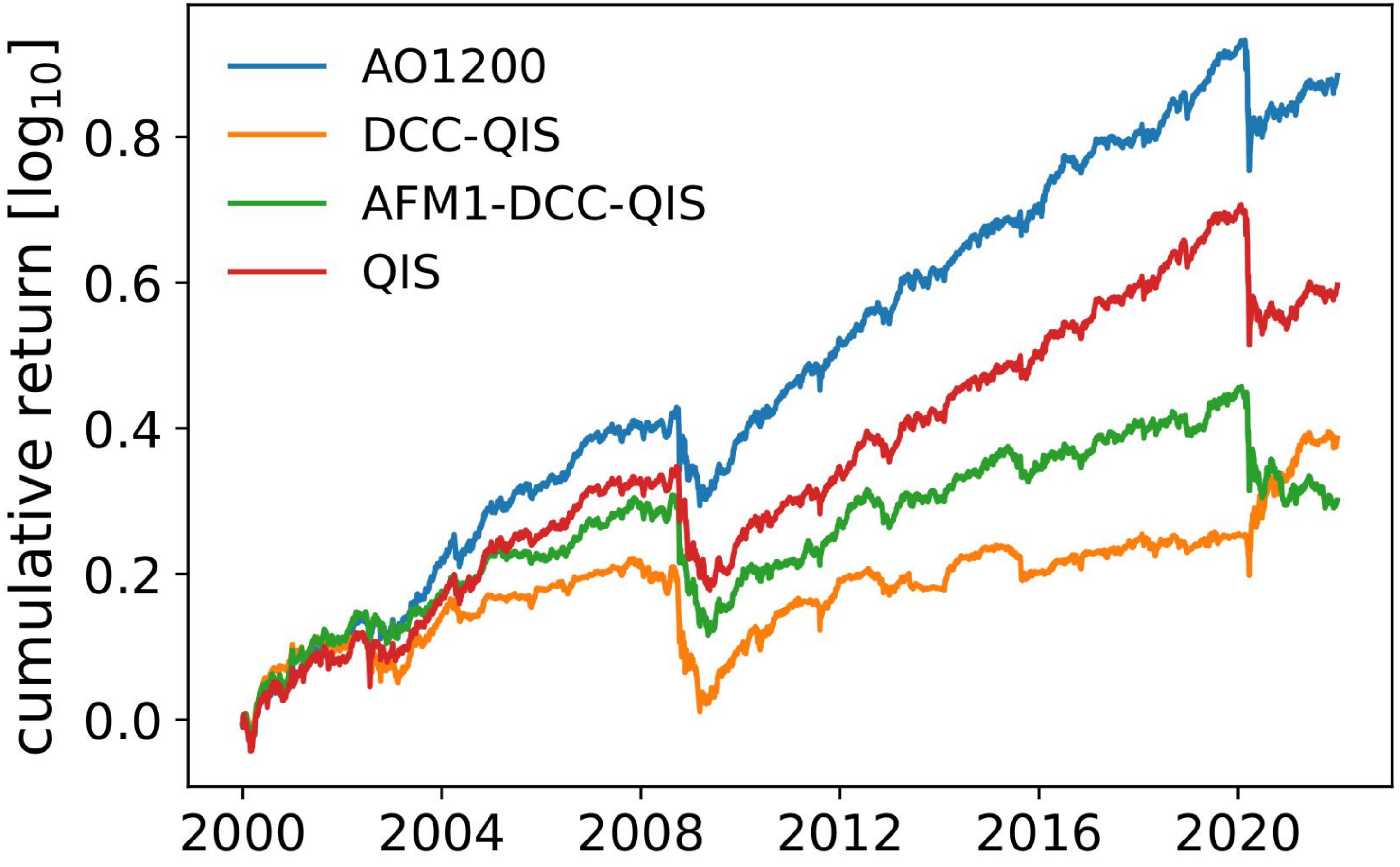
Table 1: $\Delta t_{out} = 5$ $n = 100$, average of $10,000$ simulations. GMV with long and short positions. All the values that are not statistically distinct from the best value according to a bootstrap confidence interval at 95% are marked in bold.

QIS and QuEST are alternative eigenvalue shrinkage approaches to CV (but very similar in performance)

1AFM is a single-factor model correction to the DCC model.

Bongiorno & Challet, In preparation, 2023

Methods Performances: Portfolio Management



Bongiorno & Challet, In preparation, 2023

Questions?

Assignment

- Implement the average oracle for n=100 stocks with a calibration window of 10 year and a train window of 1 year.
- Compare the performances with CV on a GMV portfolio for several realizations.
- Implement a portfolio management with the two methods with a weekly rebalancing on the remaining validation window.