

Filtering the covariance matrix of nonstationary systems with time-independent eigenvalues

Christian Bongiorno^{1,*}, Damien Challet¹, and Grégoire Loeper²

¹Université Paris-Saclay, CentraleSupélec, Laboratoire de Mathématiques et Informatique pour la Complexité et les Systèmes, 91192 Gif-sur-Yvette, France

*christian.bongiorno@centralesupelec.fr

²BNP Paribas, 20 boulevard des Italiens, 75009 Paris, France

July 22, 2022

Abstract

We propose a data-driven way to reduce the noise of covariance matrices of nonstationary systems. In the case of stationary systems, asymptotic approaches were proved to converge to the optimal solutions. Such methods produce eigenvalues that are highly dependent on the inputs, as common sense would suggest. Our approach proposes instead to use a set of eigenvalues totally independent from the inputs and that encode the long-term averaging of the influence of the future on present eigenvalues. Such an influence can be the predominant factor in nonstationary systems. Using real and synthetic data, we show that our data-driven method outperforms optimal methods designed for stationary systems for the filtering of both covariance matrix and its inverse, as illustrated by financial portfolio variance minimization, which makes our method generically relevant to many problems of multivariate inference.

Introduction

In multivariate systems, many statistical inference problems require estimating the covariance matrix or its inverse. In the simplest case, a system of interest is stationary and produces Gaussian features. Even in these favorable circumstances, a direct estimation of covariance matrices is very noisy when the number of data points is comparable with or smaller than the number of features, which is known as the curse of dimensionality, or high-dimensional setting. Using only the most recent data points is a necessity for nonstationary systems as one faces the conundrum of using as few data points as possible so as to use the most recent information while still estimating the covariance matrix precisely enough. Thus,

generically, covariance filtering requires removing two sources of systematic bias and noise: sampling noise and nonstationarity (which results in covariate shift).

Filtering sampling noise out of covariance matrices has a long history. There are two main approaches: either to coerce the matrix to follow a specified structure (i.e., to use an ansatz) Tumminello et al. [2007a], Bongiorno and Challet [2021b], or to use its decomposition into eigenvalues and eigenvectors (see Bun et al. [2017] for a review) and filter them: for example, Random Matrix Theory provides tools to compute the noisy influence of sampling errors and reversely gives methods to denoise covariance matrices (e.g. Laloux et al. [1999], Plerou et al. [1999b], Bun et al. [2017]). Estimators that only modify the covariance matrix eigenvalues are known as Rotationally Invariant Estimators (RIE thereafter). According to recent literature, an optimal RIE should minimize the Frobenius distance between the filtered and true covariance matrix. If the true covariance matrix is known, the optimal RIE is named the Oracle estimator, and can be obtained analytically. Obviously, the Oracle estimator does not make sense for forecasting, as the true covariance is unknown. In this case, it yields the lowest Frobenius norm that an RIE can achieve. Remarkably, asymptotically optimal RIEs that converge to the Oracle estimator can be obtained without the knowledge of the true covariance matrix Ledoit et al. [2012], Ledoit and Wolf [2017], Bun et al. [2016, 2017], Engle et al. [2019]; however, such estimators require that: i) the ground truth does not change (stationary systems), ii) the data matrix is very large, and iii) the data has finite fourth moments. In the following, we shall call this family of estimators NLS, which stands for non-linear shrinkage.

Yet, the most interesting complex systems are rarely stationary and often produce heavy-tailed features. Dissipative quantum systems, ecosystems, and many socio-economic systems are nonstationary in essence Chen et al. [2018]. Here, we propose a purely data-driven covariance filtering method that outperforms the optimal stationary RIE as soon the system is nonstationary. Remarkably, our method rests on an averaging procedure of Oracle eigenvalues in a long calibration dataset, which we call the Average Oracle. In other words, our method consists in replacing the eigenvalues of nonstationary systems with time-invariant eigenvalues. Because the Average Oracle leads to appreciably better estimation of covariance matrices in nonstationary systems, we expect its application domain to be wide. We apply it below to dynamic optimal resource allocation which has interdisciplinary applications (financial portfolios, wind farm locations, marketing channels, and more generally optimization problems with a quadratic cost).

Covariance matrix filtering

At time t , given n time-series (features), one needs to predict their covariance matrix Σ_{test} in the test interval $\mathcal{I}_{\text{test}} = [t, t + \delta_{\text{test}}[$ from the information known in the train interval $\mathcal{I}_{\text{train}} = [t - \delta_{\text{train}}, t[$. Even in the assumption of a stationary world, the sample estimator Σ_{train} is biased and noisy as soon the ratio $q = \frac{n}{t}$

is not negligible. Fortunately, one can improve the sample estimator by using a suitable filtering scheme, which yields a new estimator, denoted by Ξ_{train} . The idea is to bring Ξ_{train} as close as possible to Σ_{test} , which is quantified by a distance, such as the Frobenius distance (squared element-wise difference) $\|\Sigma_{\text{test}} - \Xi_{\text{train}}\|_F$.

A special class of rotationally invariant estimators uses the decomposition of the covariance matrix into eigenvectors and eigenvalues. Indeed, the spectral decomposition theorem states that $\Sigma_{\text{train}} = V_{\text{train}} \Lambda_{\text{train}} V_{\text{train}}^\dagger$, where V_{train} is the $n \times n$ eigenvector matrix and Λ_{train} is the diagonal matrix of eigenvalue monotonically ordered.

Let us keep focus on eigenvalue-based filtering (i.e., build an RIE) and thus use the empirical eigenvectors V_{train} . Generically, if Ξ is an RIE, it can be written as

$$\Xi(\Lambda) := V_{\text{train}} \Lambda V_{\text{train}}^\dagger, \quad (1)$$

where Λ is a diagonal matrix with well-chosen eigenvalues. For example, if one knows the future covariance matrix Σ_{test} , the optimal eigenvalue matrix is the so-called Oracle and can be shown [Bun et al., 2016] to be

$$\Lambda_O = \text{diag}(V_{\text{train}}^\dagger \Sigma_{\text{test}} V_{\text{train}}), \quad (2)$$

where the diag operator only keeps the diagonal of a matrix and sets the elements to zero elsewhere. These eigenvalues express the future empirical covariance matrix in the basis of the current one. They are optimal in the following sense: the related RIE

$$\Xi(\Lambda_O) = V_{\text{train}} \Lambda_O V_{\text{train}}^\dagger, \quad (3)$$

minimizes the (element-wise) Frobenius distance $\|\Xi(\Lambda_O) - \Sigma_{\text{test}}\|_F$. Although for practical purpose the exact Oracle estimator cannot be used, since Σ_{test} is in the future, asymptotical estimators that converge to the Oracle estimator are known [Ledoit et al., 2012, Bartz, 2016, Ledoit and Wolf, 2017, Bun et al., 2017]. In other words, this optimal RIE exploits a way to express Λ_O as a function of the past information only, provided that the hypotheses listed above hold, stationarity being a crucial one.

The above asymptotically optimal RIEs rest on neglecting any evolution of the covariance matrix. Although this makes sense for stationary systems, it likely discards relevant information in nonstationary systems and indeed Oracle eigenvalues do contain valuable information to filter Σ_{train} as they encode the link between the past and the future in an RIE setting. With the aim of capturing the average transition from two consecutive time-windows, our method rests on the averaging Eq. (2), rank-wise, over many randomly selected consecutive intervals taken from a long calibration window. The latter must be much larger than the one used to compute V_{train} .

More precisely (see Fig. 1 for a graphical explanation), we need to define a long calibration window $\mathcal{I}_{\text{cal}} = [t - \delta_{\text{max}}, t - \delta]$, with $\delta_{\text{max}} \gg \delta_{\text{train}}$, and δ not necessarily linked to the actual test window size δ_{test} , as shown in appendix C. Then, we select B random times $t^{(b)} \in \mathcal{I}_{\text{cal}}$. For a given $t^{(b)}$, two consecutive

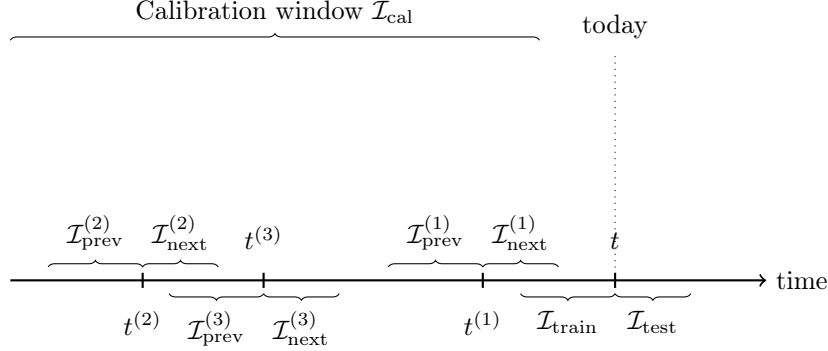


Figure 1: Average Oracle eigenvalues computation: Oracle eigenvalues are computed in many sub-intervals of a long calibration window and then averaged rank-wise. These eigenvalues can then be used outside of the calibration window.

intervals must be defined: the first interval $\mathcal{I}_{prev}^{(b)} = [t^{(b)} - \delta_{\text{train}}, t^{(b)}[$ must be of size δ_{train} , while the length of the next one $\mathcal{I}_{next}^{(b)} = [t^{(b)}, t^{(b)} + \delta[$ will be of size δ . It is worth mentioning that the next interval $\mathcal{I}_{next}^{(b)}$ is in the future with respect to $t^{(b)}$ but it is in the past with respect to t , i.e., with respect to \mathcal{I}_{test} ; therefore, an Oracle-like scheme can be applied.

Each sub-sampling has an associated set of Oracle eigenvalues

$$\Lambda_O^{(b)} = \text{diag}(V_{\text{prev}}^{(b)\dagger} \Sigma_{\text{next}}^{(b)} V_{\text{prev}}^{(b)}). \quad (4)$$

With $V_{\text{prev}}^{(b)}$ the eigenvectors of the sample covariance matrix computed in $\mathcal{I}_{prev}^{(b)}$, and $\Sigma_{\text{next}}^{(b)}$ the sample covariance of $\mathcal{I}_{next}^{(b)}$. The Average Oracle eigenvalues are then defined as the average element-wise:

$$\Lambda_{AO} := \frac{1}{B} \sum_{b=1}^B \Lambda_O^{(b)}. \quad (5)$$

It is important to stress that the columns of the eigenvectors $V_{\text{prev}}^{(b)}$ of (4) must follow always the same eigenvalue ordering convention chosen.

The AO-filtered covariance matrix is given by

$$\Xi(\Lambda_{AO}) = V_{\text{train}} \Lambda_{AO} V_{\text{train}}^\dagger. \quad (6)$$

The empirical eigenvalues from the train interval are completely discarded and replaced by the AO ones. The fact that the Average Oracle is a better estimator most of the time for nonstationary systems implies that the most recent information contained in the sample eigenvalues is less relevant (and more noisy) than the AO ones that focus on the average transition. On the other hand, the train eigenvectors contain some dynamical information and are kept.

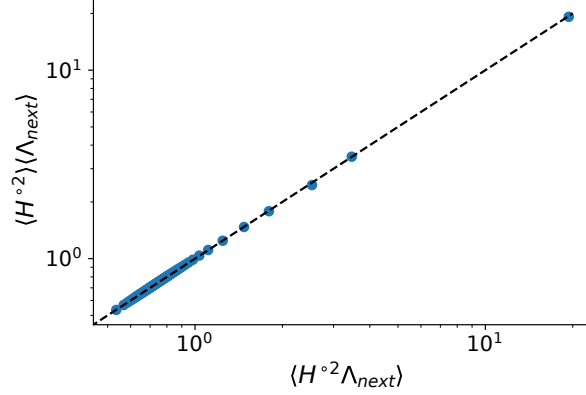


Figure 2: Empirical test of the linear independence of Λ_{next} and $H^{\circ 2}$ (Eq. (8)): $\langle H^{\circ 2} \rangle \langle \Lambda_{next} \rangle$ vs $\langle H^{\circ 2} \Lambda_{next} \rangle$.

Note also that our approach requires that the univariate variances are constant. If one deals with a system in which this is not the case, our method should be applied to the correlation matrix.

We thus propose to tackle nonstationarity with a time-invariant eigenvalue cleaning scheme. This is a zeroth order approximation, as the fluctuations of the optimal eigenvalue matrix around Λ_{AO} sometimes most probably contain valuable additional information (as may do those of the eigenvectors). Nevertheless, this approximation is a powerful filtering tool and is easily computed from data without any modelling assumption about the underlying system. In addition, our tests indicate that the filtering power does not decrease as a function of time in the nonstationary systems that we investigated.

The Role of Non-Stationarity

To understand how the Oracle eigenvalues are related to the non-stationarity of both the eigenvalues and eigenvectors of a covariance matrix, we decompose (2) as

$$\Lambda_O = (V_{prev}^\dagger V_{next})^{\circ 2} \Lambda_{next} = H^{\circ 2} \Lambda_{next}, \quad (7)$$

where $\bullet^{\circ 2}$ represents the Hadamard element-wise square of the matrix H . The matrix H is in fact a rotation matrix from the V_{prev} eigenvector basis to the V_{next} eigenvector basis and hence belongs in $SO(n)$.

This decomposition of the Oracle estimator makes it possible to analyse the respective contributions of H and Λ_{next} to the Oracle estimator, and, by linearity, to the Average Oracle. Indeed, according to Eq. 7, it is natural to test if the problem of overlap and test eigenvalues estimation can be separated, i.e., if

$$\langle H^{\circ 2} \Lambda_{next} \rangle \simeq \langle H^{\circ 2} \rangle \langle \Lambda_{next} \rangle. \quad (8)$$

It is worth noticing that although the average element-wise of an element of the $SO(n)$ group is not an element of the $SO(n)$ group, the element-wise average of $H^{\circ 2}$ still has the same relevant property as the original elements, i.e., the row-wise or column-wise sum equal one.

We test Eq. (8) with financial data. It is worth stressing, as detailed in the appendix B, that the variance of each time series is itself nonstationary and hence we focus on correlation rather than covariance matrix. In practice, we standardize the time series on every subinterval considered (see the appendix A for a full description of data handling).

We find that Eq. (8) holds remarkably well (Fig. 2): $H^{\circ 2}$ and Λ_{next} are linearly independent and thus both quantities can be assumed to contribute independently to the fluctuations of the Oracle eigenvalues. As a consequence, we are allowed to focus on the influence of nonstationary eigenvalues and eigenvectors separately.

Overlap matrix stability

The element-wise square h_{ij}^2 represents the projection of the i eigenvector from V_{prev} on the eigenvector j from V_{next} . In the case of the perfect overlap, only one $h_{ij} = 1$, all the other ones being zero. This never happens because of finite-sample size error and non-stationarity. The other extreme case is $h_{ij}^2 = \frac{1}{n}$ for all j , which corresponds to the lower overlap possible. Given these mathematical properties of the overlap matrix, Shannon entropy is an appropriate measure of the amount of overlap. For eigenvector i , we write

$$E_i = - \sum_{j=1}^n h_{ij}^2 \log_n h_{ij}^2 \quad (9)$$

with the standard Shannon entropy convention that $0 \log 0 = 0$ and normalization by $\log(n)$, in such a way that the highest overlap will be $E_i = 0$ while the lowest $E_i = 1$.

By using Eq. (9), we can test on real data if the average overlap in a stationary world, due only to sampling size error, is compatible with the measured overlap of the real world. We carried out such an experiment in the following way. We randomly select 10,000 random time-windows $\mathcal{I}_{\text{prev}}^{(b)}, \mathcal{I}_{\text{next}}^{(b)}$ with $\delta_{\text{prev}} = \delta = 252$ days and perform an independent random selection of n stocks for each time-window. We then measure the entropy of the overlap for each V_{prev} eigenvector ranked from the smallest eigenvalue to the largest one. The entropy is then averaged element-wise over these independent realizations, which yields a measure of real-world overlap, influenced both by sampling error and non-stationarity.

To show that the observed overlap is systematically lower than the one expected in a stationary world, we design a local data stationarization procedure and we measure the rank-wise average overlap E_i again. First we take the union of the intervals $\mathcal{I}_{\text{prev}}^{(b)} \cup \mathcal{I}_{\text{next}}^{(b)}$, then we shuffle the temporal ordering of the observations, then we split the shuffled data again into $\mathcal{I}_{\text{prev}}^{\prime(b)}, \mathcal{I}_{\text{next}}^{\prime(b)}$, but after the

shuffling both intervals will contain a mixture a past and future events; therefore, stationary.

Figure 3 displays the average entropy for both cases. One sees that the stationarization leads to smaller entropy, hence, larger apparent overlap. This means that in a fictional stationary world, the overlap is mechanically larger, which in turn implies that this assumption leads to a bias in the eigenvectors' stability in the real world. While for $n = 10$ the overestimation of the overlap for the stationary cases is clear, for $n = 100$ we must look at the difference between the two estimators. One notes that the overestimation of the stationary case is systematic even for large n .

Eigenvalues stability

In Eq. 2, the other part of Oracle the estimator is the expectation of $E[\Lambda_{\text{next}}]$. A reasonable assumption in a stationary world, if $\delta_{\text{train}} = \delta$, is that $E[\Lambda_{\text{next}}] = \Lambda_{\text{prev}}$.

Another possibility is that, because of very strong non-stationarity, the eigenvalues fluctuate so much that their average $\langle \Lambda_{\text{next}} \rangle$ computed over many time-windows within a much larger calibration time-window \mathcal{I}_{cal} is a better predictor of Λ_{next} than the closest past $\mathcal{I}_{\text{prev}}$.

In order to test these two hypotheses, we computed the average $\langle \Lambda_{\text{next}} \rangle$ over 10,000 randomly chosen time-windows of $\delta_{\text{prev}} = \delta = 252$ days in the calibration window [1995, 2005]. We then we tested the deviation $\Lambda_{\text{prev}} - \Lambda_{\text{next}}$ and $\langle \Lambda_{\text{next}} \rangle - \Lambda_{\text{next}}$ in the test window [2006, 2018] with 10,000 randomly chosen time-windows. We carried out such a comparison with L_1 and L_2 norms.

Specifically the L_1 norm is defined as

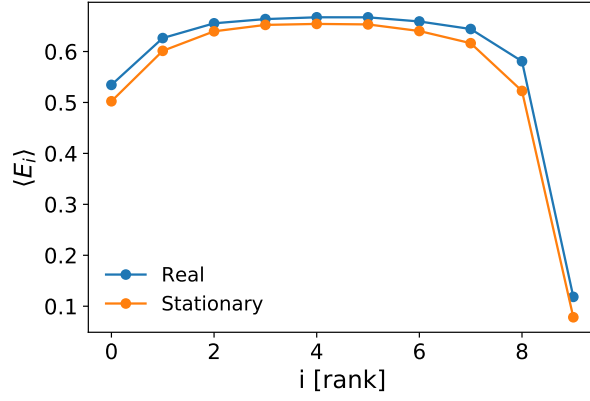
$$D_{\bullet}^{(1)} = \sum_{i=1}^n |\lambda_{i,\bullet} - \lambda_{i,\text{next}}| \quad (10)$$

and the L_2 norm is defined as

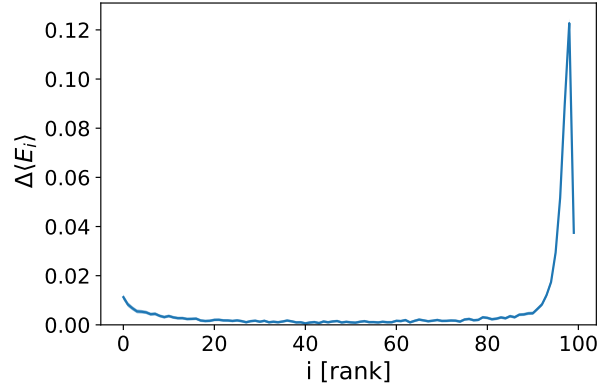
$$D_{\bullet}^{(2)} = \sqrt{\sum_{i=1}^n (\lambda_{i,\bullet} - \lambda_{i,\text{next}})^2} \quad (11)$$

where $\lambda_{i,\bullet}$ can be the $\langle \lambda_{i,\text{next}} \rangle$ or $\lambda_{i,\text{prev}}$.

In Fig. 4, we show the distribution of the eigenvalue deviation as $D_{\langle \text{next} \rangle}^{(2)} - D_{\text{prev}}^{(2)}$ and $D_{\langle \text{next} \rangle}^{(1)} - D_{\text{prev}}^{(1)}$. Each point of the distribution is a random selection of the time-window on the validation period. The average of both distributions is close to zero, supporting the idea that the most recent eigenvalues are approximately as good as the average historical ones. More precisely the distribution average is slightly negative: -0.1 for the L_2 , with a 95-percentile bootstrap p-value of 0.1 and -0.8 with a 95-percentile bootstrap p-value of zero (computed with 100,000 bootstrap re-sampling). However, such a difference is marginal, and the average approach does not bring any significant improvement.

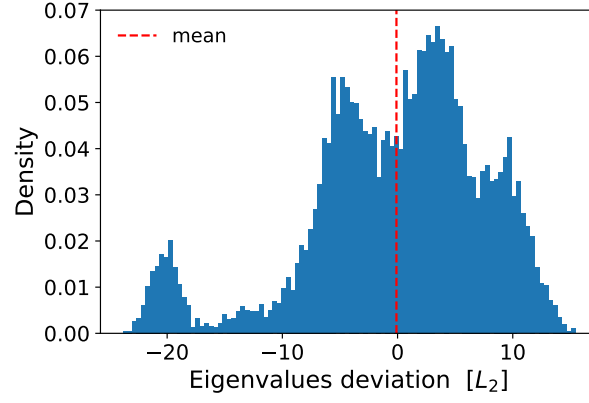


(a)

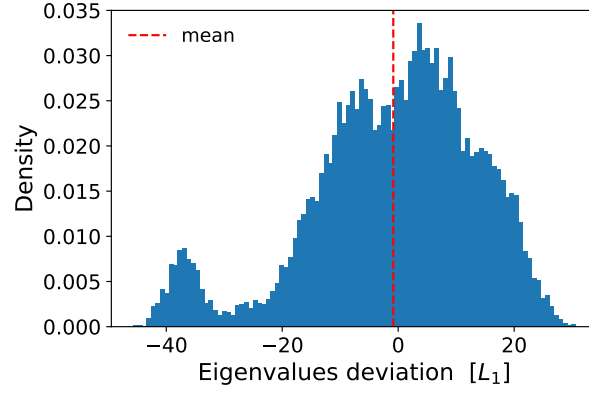


(b)

Figure 3: Entropy of the eigenvector overlap for US equities. Each point is an average of 10,000 independent realizations. The left plot refers to $n = 10$; the right plot is the difference between the real (nonstationary data) entropy and the stationarized data entropy for $n = 100$; the plot includes 95% confidence intervals which are not clearly visible due to the precision of the estimation.



(a)



(b)

Figure 4: Plot (a): distribution of eigenvalue deviation $D_{\langle \text{next} \rangle}^{(2)} - D_{\text{prev}}^{(2)}$; plot (b): distribution of eigenvalue deviation $D_{\langle \text{next} \rangle}^{(1)} - D_{\text{prev}}^{(1)}$; each element of the distribution refers to a randomly sampled time-window and a random set of $n = 100$ stocks; $\delta_{\text{prev}} = \delta = 252$.

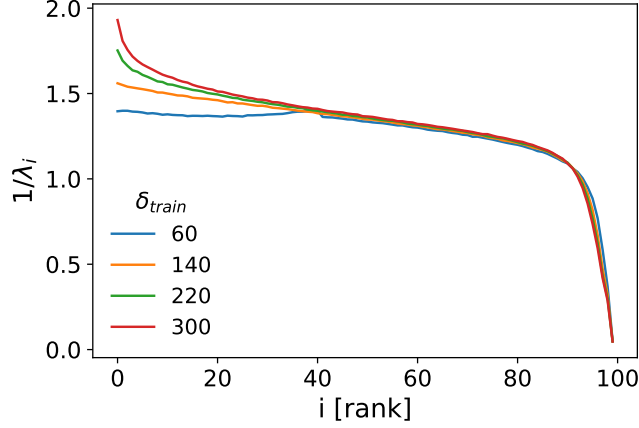


Figure 5: Inverse Average Oracle eigenvalues as a function of the eigenvalue rank for $n = 100$ and various train window sizes. US Equities correlation matrix; $B = 10000$ sub-intervals in the 1995-2006 period.

Matrix distance between filtered and realized covariance

We expect our method to be useful for any nonstationary system as a zeroth order correction. Here, we test AO vs. NLS with financial data as they are abundant, nonstationary, and heavy-tailed Gopikrishnan et al. [1999]. We also use synthetic data to assess the respective influence of nonstationarity and heavy tails on the performance of both AO and NLS in appendix D.

In some nonstationary systems such as financial markets, the order of magnitude of univariate variances Σ_{ii} strongly depends on time. In order to remove this source of non-stationarity, we focus on the eigenvalue correction of the correlation matrix and compute $\Xi(\Lambda_{AO}) = V_{\text{train}}^\dagger \Lambda_{AO} V_{\text{train}}$ where the eigenvalues and eigenvectors are those of correlation matrices. The elements of a filtered covariance matrix are obtained by multiplying the filtered correlation matrix by the respective univariate standard deviations

$$(\Sigma_{AO})_{ij} = \sqrt{(\Sigma_{\text{train}})_{ii}} \Xi(\Lambda_{AO})_{ij} \sqrt{(\Sigma_{\text{train}})_{jj}} \quad (12)$$

In the following, we use about 24 years of daily data for $N = 1000$ assets from the US stock market. The Average Oracle eigenvalues Λ_{AO} are calibrated in the 1995 to 2005 period from $B = 10000$ subintervals; for the sake of computation speed, we take $n < N$ random assets in each subinterval. Because we randomize asset selection, the resulting Average Oracle eigenvalues can be applied to any selection of assets (and to other markets). One could also choose to compute Λ_{AO} for a fixed set of assets. For a full description of the data handling, see the appendix A.

The resulting AO eigenvalues are reported in Fig. 5. Note that we plot the inverse of the average eigenvalues in order to emphasize the dependence on n of small eigenvalues, as many inference problems use the inverse covariance matrix (see below), and also to reduce the influence of the largest eigenvalues on the clarity of the figure.

The first way to compare the performance of both NLS and AO is to compute the average Frobenius norm in the out-of-sample period (2006–2018). We carried out extensive simulations with various $\delta_{\text{train}}, \delta_{\text{test}} \in \{40, \dots, 500\}$, selecting 100 random selections of $n = 100$ assets for each $t^{(b)}$. While Λ^{AO} depends on n and δ_{train} , it does not seem to depend on δ (see appendix C) which we fix henceforth to 252 (one year of daily data). We compared the Average Oracle approach with an efficient and provably good numerical implementation of NLS Bartz [2016], Bun et al. [2018] based on cross-validation within the train window.

The average Frobenius norm in the out-of-sample period for NLS and AO is reported in the left plot of Fig. 6. AO clearly does better than NLS, even if the latter is designed to minimize this norm in the stationary case. For the sake of completeness, we also added the unrealistic case where the Oracle eigenvalues are computed from the future as in Eq. (2), which shows how much the AO could still be improved with a predictive model of eigenvalues. Note that the advantage of AO over NLS increases with n (appendix G). We also check in appendix F that the same results hold for the Kullback-Leibler distance.

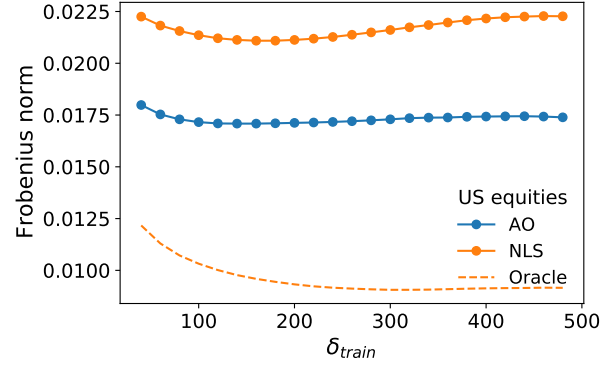
The superior performance of the Average Oracle is mainly due to nonstationarity. Indeed, when we stationarize the data in the train and test interval as described in Sec. the advantage of using AO disappears. In particular, in Fig. 6 (right plot) we show that NLS clearly outperforms the Average Oracle on the Frobenius distance on stationarized data. Thus, the advantage of the Average Oracle is precisely that it captures some part of the average dynamics that is discarded by the assumption of stationarity (see also appendix D).

Application to portfolio optimization

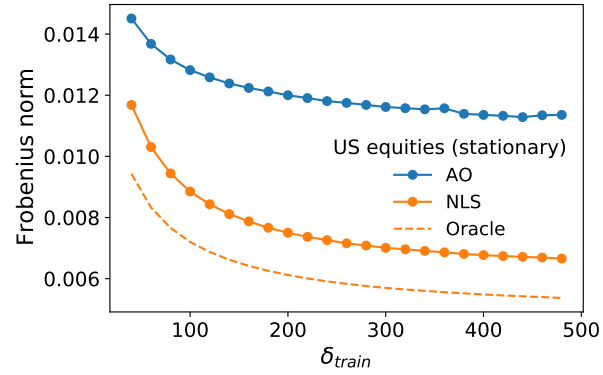
A canonical application of covariance matrices in a resource allocation context is portfolio optimisation. The simplest case only uses the covariance matrix and aims at minimizing the realized variance of the value of a portfolio of assets from the knowledge of data in the train interval. Mathematically, a portfolio is defined by the fraction w_i of wealth invested into each available asset $i = 1, \dots, n$. In other words, the performance of a portfolio with weights $w \in \mathbb{R}^n$ is the weighted sum of the performance of all the assets, i.e., $r_P = \sum_{i=1}^n w_i r_i$, where r_i is the price return of asset i , and its variance is $w^\dagger \Sigma w$. Practically, the weights are computed from the train window data and the covariance is that of the test window, thus the realized portfolio volatility σ_P is given by

$$(\sigma_P)^2 = w^\dagger \Sigma_{\text{test}} w. \quad (13)$$

The optimization problem usually adds the normalization constraint $\sum_{i=1}^N w_i = 1$. This defines the Global Minimum Variance Portfolio problem (GMV). Intu-

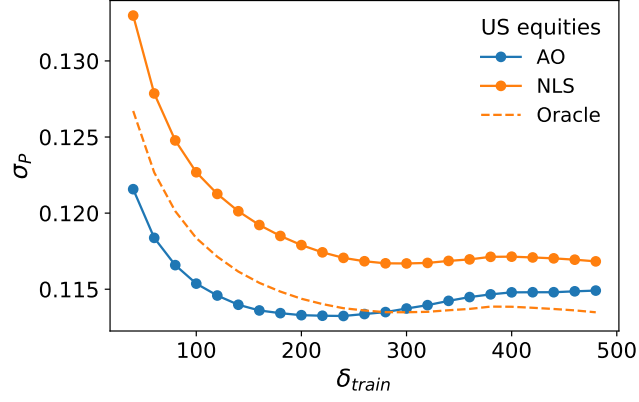


(a)

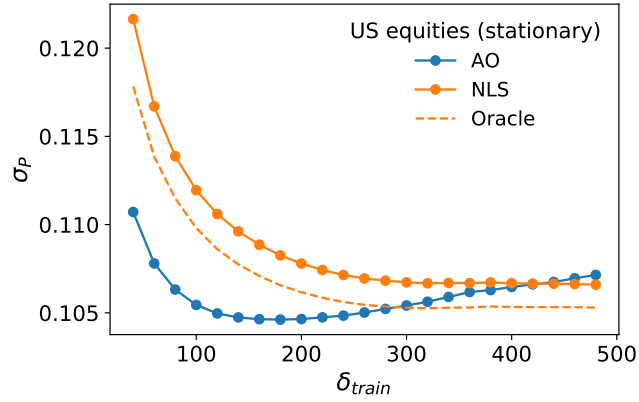


(b)

Figure 6: Average Frobenius distance between the filtered and test covariance matrices as a function of the train window length in the out-of-sample period. Plot (a) refers to the original data set, plot (b) to the stationarized data. 100 portfolios with random $n = 100$ assets are computed for each day of the out-of-sample.



(a)



(b)

Figure 7: Average realized volatility of Global Minimum Variance portfolios as a function of the calibration window length. The plot (a) refers to the original dataset, the plot (b) to the stationary data. 100 portfolios with random $n = 100$ assets are computed for each day of the out-of-sample.

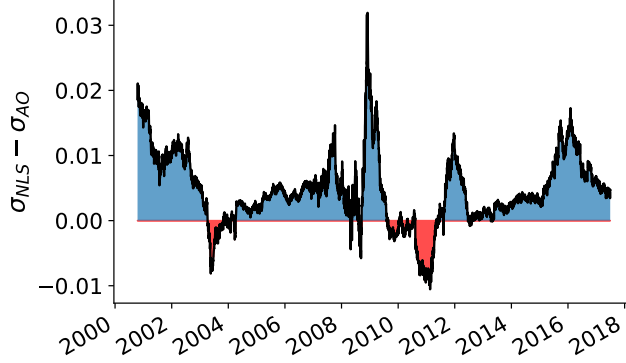


Figure 8: Difference of realized volatility as a function of time between NLS and the Average Oracle. $n = 100$, averages over 100 portfolios per date.

itively, minimizing σ_P requires a small distance between Σ_{train} and Σ_{test} , hence the importance of the Frobenius norm (see Fig. 6), which is the usual criterion in the covariance matrix filtering literature Bun et al. [2017]. It is a reasonable heuristics, but does not lead to the optimal portfolio weights [Bongiorno and Challet, 2021c]. Simple computations indeed lead to the optimal weights

$$w^* = \frac{\Sigma^{-1}\mathbb{1}}{\mathbb{1}^\dagger \Sigma^{-1}\mathbb{1}}. \quad (14)$$

One immediately notices that portfolio optimization also requires that the inverse of the covariance matrix (the precision matrix) be also well filtered as the optimal weights are much influenced by the smallest eigenvalues of Σ .

Figure 7 shows that the realized volatility of GMV portfolios is smaller when using the Average Oracle than when using NLS, as expected from the Frobenius norm. However, quite remarkably AO is also better than the Oracle eigenvalues when $\delta_{train} < 270$. We use the same way to check the importance of nonstationarity. The Average Oracle still outperforms the Oracle and NLS for small enough δ_{train} . i.e. in the high-dimensional case. This is due to the fact that minimising the Frobenius norm is not optimal for portfolio optimisation Bongiorno and Challet [2021c].

The Average Oracle outperforms NLS most of the time: plotting the average realized volatility as a function of time (Fig. 8), one sees that there are only a few periods during which AO loses to NLS, that there seems to be no difference between the AO calibration period (until 2005) and the testing period (from 2006), and finally that the advantage of AO does not decrease as time goes on.

As a final test, we applied the Average Oracle calibrated with US data to Hong-Kong equity data and found qualitatively similar data (see appendix I). This strongly suggests that the AO captures a systematic nonstationary effect found in two different nonstationary systems.

Interestingly, the Average Oracle and the method we use to compute the

NLS share a common ingredient: this method computes the Average Oracle eigenvalues of bootstrapped calibration and out-of-sample periods restricted to δ_{train} , whereas the Average Oracle method uses 20 years of data and respects causality (time order) between the calibration and the out-of-sample time windows. The reason why AO outperforms NLS is because causality, hence, time ordering, needs to be conserved in nonstationary systems.

We note that NLS can also be applied to z-scores of asset price returns instead of on the return themselves. In this case, we find that NLS leads to smaller GMV portfolio variance on average, still larger than AO, but with a significantly larger Frobenius distance. This means that the Average Oracle provides a hands-off approach to covariance cleaning and does not require complex computations. Once calibrated, the AO is much faster than NLS.

Discussion

By outperforming the optimal estimators for stationary systems, our method shows the need (and the possibility) of further improvement. Indeed, real systems are rarely stationary and thus one needs to account for the evolution of the real covariance matrix. In such circumstances, we have shown that a simple zeroth-order eigenvalues filtering method, that only retains an average dependence between past and future, already outperforms optimal known solutions for stationary systems.

The Average Oracle is a first step towards accounting for nonstationarities when filtering covariance matrices, as it is a zeroth-order correction. Any additional knowledge about the underlying system may help design higher-order corrections to our method. This knowledge may not come from the covariance matrix itself, but from possibly higher-order dependence measures, such as triads, which are much better at predicting the instability of the sign of correlation coefficients (see for example Ref. Bongiorno and Challet [2021a]).

The fact that the average influence of the future on the correlation matrix eigenvalues is more informative than the empirical eigenvalues themselves in nonstationary systems suggests ways to extend analytical results. Similarly, it is natural to quantify how much exploitable information lies in the difference between the empirical and Oracle eigenvalues, i.e., how to mix the Average Oracle eigenvalues with the empirical ones from the training period.

References

- Daniel Bartz. Cross-validation based nonlinear shrinkage. *arXiv preprint arXiv:1611.00798*, 2016.
- Christian Bongiorno and Damien Challet. Non-parametric sign prediction of high-dimensional correlation matrix coefficients. *EPL (Europhysics Letters)*, 133(4):48001, 2021a.

- Christian Bongiorno and Damien Challet. Reactive global minimum variance portfolios with k-bahc covariance cleaning. *The European Journal of Finance*, pages 1–17, 2021b.
- Christian Bongiorno and Damien Challet. The Oracle estimator is suboptimal for global minimum variance portfolio optimisation. *arXiv preprint arXiv:2112.07521*, 2021c.
- Joël Bun, Romain Allez, Jean-Philippe Bouchaud, and Marc Potters. Rotational invariant estimator for general noisy matrices. *IEEE Transactions on Information Theory*, 62(12):7475–7490, 2016.
- Joël Bun, Jean-Philippe Bouchaud, and Marc Potters. Cleaning large correlation matrices: tools from random matrix theory. *Physics Reports*, 666:1–109, 2017.
- Joël Bun, Jean-Philippe Bouchaud, and Marc Potters. Overlaps between eigenvectors of correlated random matrices. *Physical Review E*, 98(5):052145, 2018.
- TT Chen, B Zheng, Y Li, and XF Jiang. Temporal correlation functions of dynamic systems in non-stationary states. *New Journal of Physics*, 20(7):073005, 2018.
- Robert F. Engle, Olivier Ledoit, and Michael Wolf. Large dynamic covariance matrices. *Journal of Business & Economic Statistics*, 37(2):363–375, 2019.
- Parameswaran Gopikrishnan, Vasiliki Plerou, Luis A Nunes Amaral, Martin Meyer, and H Eugene Stanley. Scaling of the distribution of fluctuations of financial market indices. *Physical Review E*, 60(5):5305, 1999.
- Laurent Laloux, Pierre Cizeau, Jean-Philippe Bouchaud, and Marc Potters. Noise dressing of financial correlation matrices. *Physical review letters*, 83(7):1467, 1999.
- Olivier Ledoit and Michael Wolf. Nonlinear shrinkage of the covariance matrix for portfolio selection: Markowitz meets goldilocks. *The Review of Financial Studies*, 30(12):4349–4388, 2017.
- Olivier Ledoit, Michael Wolf, et al. Nonlinear shrinkage estimation of large-dimensional covariance matrices. *The Annals of Statistics*, 40(2):1024–1060, 2012.
- Vasiliki Plerou, Parameswaran Gopikrishnan, Luis A Nunes Amaral, Martin Meyer, and H Eugene Stanley. Scaling of the distribution of price fluctuations of individual companies. *Physical review E*, 60(6):6519, 1999a.
- Vasiliki Plerou, Parameswaran Gopikrishnan, Bernd Rosenow, Luís A Nunes Amaral, and H Eugene Stanley. Universal and nonuniversal properties of cross correlations in financial time series. *Physical Review Letters*, 83(7):1471, 1999b.

Michele Tumminello, Fabrizio Lillo, and Rosario N Mantegna. Hierarchically nested factor model from multivariate data. *EPL (Europhysics Letters)*, 78(3):30006, 2007a.

Michele Tumminello, Fabrizio Lillo, and Rosario N Mantegna. Kullback-leibler distance as a measure of the information filtered from multivariate data. *Physical Review E*, 76(3):031123, 2007b.

Acknowledgements

This publication stems from a partnership between CentraleSupélec and BNP Paribas and used HPC resources from the “Mésocentre” computing center of CentraleSupélec and École Normale Supérieure Paris-Saclay supported by CNRS and Région Île-de-France.

Code: a notebook is available at <https://gitlab-research.centralesupelec.fr/2019bongiornc/average-oracle-cleaning>.

A Data

We use about 25 years of daily data for about $N = 1000$ US equities from which we compute returns $r_{i,t} = p_i(t)/p_i(t-1) - 1$ adjusted for splits, reverse splits and dividends. When computing Oracle eigenvalues, we applied two asset selection filters.

We applied two asset selection filters. First, for a given $\mathcal{I}_{prev}^{(b)}$ time subinterval and its corresponding $\mathcal{I}_{next}^{(b)}$ subinterval, we only keep the assets that have less than 20% of zero or missing values in the train window in order to avoid undefined standard deviation during the stationarization resampling procedure. Some assets do not have data for the whole period. Other causes for missing data or zero values are technical issues or trading stops. We did not apply the same filter to the test window so as to avoid using future information in our analysis.

In addition, we require that no pair of assets in our subset have in the train window a correlation coefficient larger than 0.95 in order to avoid duplicated assets (for example different asset classes of the same company).

B Nonstationary univariate variance

When the variance of individual time series is not constant, one should apply the AO method on the correlation matrix, i.e., compute the AO eigenvalues on data standardized in $\mathcal{I}_{prev}^{(b)}$ and $\mathcal{I}_{next}^{(b)}$ separately, and use them to replace the eigenvalues of the correlation matrix computed in \mathcal{I}_{train} . As this is clearly the case of financial data, all the results reported here that use financial data use this kind of standardization. The filtered covariance matrix is then defined as

the filtered correlation matrix suitably multiplied by the individual standard deviation (see Eq. (12)).

C Dependence of the Average Oracle eigenvalues on the next subinterval length

We checked that the Oracle eigenvalues can be considered independent from the $\mathcal{I}_{\text{next}}$ window length δ . In Fig. 9, we show that the AO eigenvalues for different δ fixing $\delta_{\text{train}} = 252$ days: the estimator is only weakly sensitive to the test window length. This observation supports the idea that the AO procedure can extract the underlying stationary part of the eigenvalue dynamic. Note that by reducing the test window, the estimation becomes noisier and thus requires more train and test windows (denoted by B) to yield average eigenvalues with the same level of precision.

D A model of nonstationary correlation matrices

The empirical findings about eigenvector overlaps suggests a simple model with a fixed set of true eigenvalues Λ_{true} and a dynamical set of eigenvectors V_t . To mimic the real heterogeneity of the eigenvalues, we choose a geometric progression of the eigenvalues, whereas the initial $t = 0$ eigenvector basis V_0 is chosen randomly from the $SO(n)$ group. At each time-step, the eigenvectors are rotated with rotation matrix $H_t \in SO(n)$, which yields

$$V_{t+1} = V_t H_t. \quad (15)$$

In order to control the amount of rotation, we decompose the rotation matrix as $n(n-1)/2$ elementary plane rotation according to the canonical basis, which corresponds to the decomposition of the rotation matrix in Euler angles

$$H_t = H(\alpha_{1,t})H(\alpha_{2,t}) \cdots H(\alpha_{n(n-1)/2,t}) \quad (16)$$

The limit $\alpha_{i,t} \rightarrow 0$ corresponds to constant eigenvectors and the system is stationary. In order to test different levels of non-stationarity, we sampled independently $\alpha_{i,t} \sim \mathcal{N}(0, s^2)$ from a normal distribution with expected value 0 and standard deviation s .

To simulate the data, we used a factor model

$$X_t = P_t A_t, \quad (17)$$

where

$$P_t = V_t \Lambda_{\text{true}}^{\frac{1}{2}} \quad (18)$$

and A_t are sampled from n independent standardized normal or Student t -distributions.

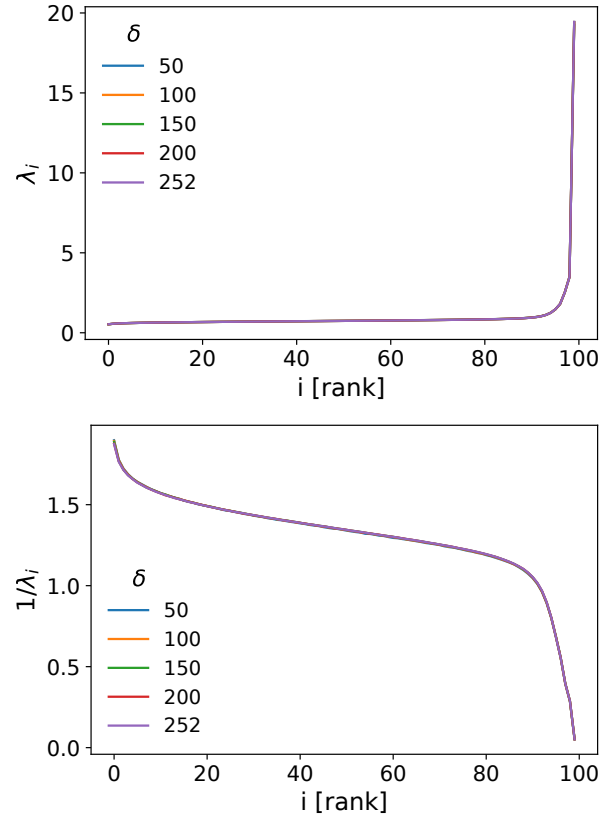


Figure 9: Average Oracle Eigenvalues estimated with a train window $\delta_{train} = 252$ days and various test windows δ . The upper plot displays the averaged eigenvalues, whereas the lower plot refers to their inverse. The calibration is performed with an average over $B = 100,000$ elements.

According to this definition,

$$\mathcal{E}[X_t X_t^\dagger] = \mathcal{E}[P_t A_t A_t^\dagger P_t^\dagger] = P_t P_t^\dagger = V_t \Lambda_{\text{true}} V_t^\dagger \quad (19)$$

since $\mathcal{E}[A_t A_t^\dagger]$ is the identity and P_t are deterministic.

Because the eigenvalues are kept fixed in this model, Eq. (8) in the main text stipulates that it is enough to compute the average $\langle H^{\circ 2} \rangle$ in order to obtain the Average Oracle. To this end, we generate 1,000 simulations for each parameter s in the following way. For each simulation, we produce a data matrix X of $n = 10$ elements and $T = 10,000$ records which represents the full historical dataset. The last $\delta_{\text{test}} = 50$ records will be kept as the test window, while the first $T - \delta_{\text{test}}$ will be the calibration window. The small n choice was necessary due to the non-negligible computational effort to apply the Euler angles rotations.

We first compute the average overlap matrix $\langle H^{\circ 2} \rangle$ from 10,000 random consecutive time intervals of length $\delta t_{\text{train}} = 50$ drawn from the calibration window; this is done in two ways: first by keeping the original time order of the data, and then by building a stationarized data set using time shuffling, which yields the nonstationary and stationary average overlap matrices, respectively. Finally, we compute the train eigenvectors V_{train} from the last δt_{train} records of the calibration window.

Then, the two AO estimators corresponding to are obtained with

$$\Lambda_{AO} = H^{\circ 2} \Lambda_{\text{true}} \quad (20)$$

and the RIE estimator as

$$\Sigma_{AO} = V_{\text{train}} \Lambda_{AO} V_{\text{train}}^\dagger \quad (21)$$

We included the true eigenvalues Λ_{true} in the estimator to reduce the amount of noise in the benchmark and to focus on the nonstationary effect of the eigenvector rotation. Finally, we compute the Frobenius distance between the estimators and Σ_{test} from the test window.

In Fig. 10, we show that for $s \rightarrow 0$ no difference between the two estimators is detected, as such a case corresponds to an almost stationary world. Similarly, for s very large, we do not observe any significant difference between the two estimators, as the nonstationary effect is so strong as to destroy any relationship between past and future, and both estimators converge to the identity. Finally, for intermediate values of s , the nonstationary estimator outperforms the stationary one. It is worth reporting that for the case of Student t-distributed variables, the discrepancy between the two estimators increases when the degree of freedom ν of the t-distribution decreases. This is particularly relevant for financial applications, where the returns are characterized by heavy tails with $\nu \simeq 4$ on average Plerou et al. [1999a].

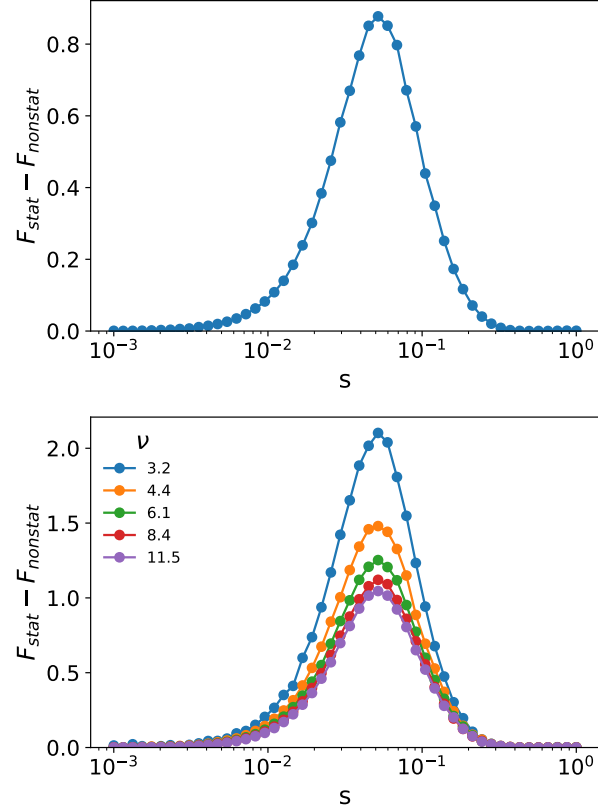


Figure 10: Non-stationary benchmark. The upper-plot refers to normal variable, the lower-plot to t-student with ν degrees of freedom. The plots shows the difference of Frobenius distance with the test covariance of the stationary minus the non-stationary estimator. The x-axis is the average amount of eigenvector rotation among two consecutive time-step.

E Influence of overlaps in synthetic nonstationary data

The empirical findings about overlaps in the main text suggests a simple model with a fixed set of true eigenvalues Λ_{true} and a dynamical set of eigenvectors V_t . To mimic the real heterogeneity of the eigenvalues, we choose a geometric progression of the eigenvalues, whereas the initial $t = 0$ eigenvector basis V_0 is chosen randomly from the $SO(n)$ group. At each time-step, the eigenvectors are rotated with rotation matrix $H_t \in SO(n)$, which yields

$$V_{t+1} = V_t H_t. \quad (22)$$

In order to control the amount of rotation, we decompose the rotation matrix as $n(n-1)/2$ elementary plane rotation according to the canonical basis, which corresponds to the decomposition of the rotation matrix in Euler angles

$$H_t = H(\alpha_{1,t})H(\alpha_{2,t}) \cdots H(\alpha_{n(n-1)/2,t}) \quad (23)$$

The limit $\alpha_{i,t} \rightarrow 0$ corresponds to constant eigenvectors and the system is stationary. In order to test different levels of non-stationarity, we sampled independently $\alpha_{i,t} \sim \mathcal{N}(0, s^2)$ from a normal distribution with expected value 0 and standard deviation s .

To simulate the data, we used a factor model

$$X_t = P_t A_t, \quad (24)$$

where

$$P_t = V_t \Lambda_{\text{true}}^{\frac{1}{2}} \quad (25)$$

and A_t are sampled from n independent standardized normal or Student t -distributions.

According to this definition,

$$\mathcal{E}[X_t X_t^\dagger] = \mathcal{E}[P_t A_t A_t^\dagger P_t^\dagger] = P_t P_t^\dagger = V_t \Lambda_{\text{true}} V_t^\dagger \quad (26)$$

since $\mathcal{E}[A_t A_t^\dagger]$ is the identity and P_t are deterministic.

E.1 Simulations

Because the eigenvalues are kept fixed in this model, Eq. (8) stipulates that it is enough to compute the average $\langle H^{\circ 2} \rangle$ in order to obtain the Average Oracle. To this end, we generate 1,000 simulations for each parameter s in the following way. For each simulation, we produce a data matrix X of $n = 10$ elements and $T = 10,000$ records which represents the full historical dataset. The last $\delta_{\text{test}} = 50$ records will be kept as the test window, while the first $T - \delta_{\text{test}}$ will be the calibration window. The small n choice was necessary due to the non-negligible computational effort to apply the Euler angles rotations.

We first compute the average overlap matrix $\langle H^{\circ 2} \rangle$ from 10,000 random consecutive time intervals of length $\delta t_{train} = 50$ drawn from the calibration window; this is done in two ways, as described in the main text: first by keeping the original time order of the data, and then by building a stationarized data set using time shuffling, which yields the nonstationary and stationary average overlap matrices. Finally, we compute the train eigenvectors V_{train} from the last δt_{train} records of the calibration window.

Then, the two AO estimators corresponding to are obtained with

$$\Lambda_{AO} = H^{\circ 2} \Lambda_{true} \quad (27)$$

and the RIE estimator as

$$\Sigma_{AO} = V_{train} \Lambda_{AO} V_{train}^{\dagger} \quad (28)$$

We included the true eigenvalues Λ_{true} in the estimator to reduce the amount of noise in the benchmark and to focus on nonstationary effect of the eigenvector rotation. Finally, we compute the Frobenius distance between the estimators and Σ_{test} from the test window.

In Fig. 11, we show that for $s \rightarrow 0$ no difference between the two estimators is detected, as such a case corresponds to an almost stationary world. Similarly, for s very large, we do not observe any significant difference between the two estimators, as the nonstationary effect is so strong as to destroy any relationship between past and future, and both estimators converge to the identity. Finally, for intermediate values of s , the nonstationary estimator outperforms the stationary one. It is worth reporting that for the case of Student t-distributed variables, the discrepancy between the two estimators increases when the degree of freedom ν of the t-distribution decreases. This is particularly relevant for financial applications, where the returns are characterized by heavy tails with $\nu \simeq 3$ on average Plerou et al. [1999a].

F Kullback–Leibler Divergence

For the sake of completeness, we include a comparison between the covariance estimators and realized covariance matrix by using the Kullback–Leibler (KL) divergence and compare our results with those for the Frobenius distance.

The KL divergence $KL(A||B)$ of the two distributions A and B measures the amount of information lost if B is used to approximate A . It is defined as

$$KL(A||B) = \int_{\mathcal{X}} \log \left(\frac{A}{B} \right) dA \quad (29)$$

which is the expectation according to distribution A of the log difference $\log(A) - \log(B)$. Some authors proposed the KL divergence as an alternative metric to the Frobenius distance when comparing correlation or covariance matrices Tumminello et al. [2007b].

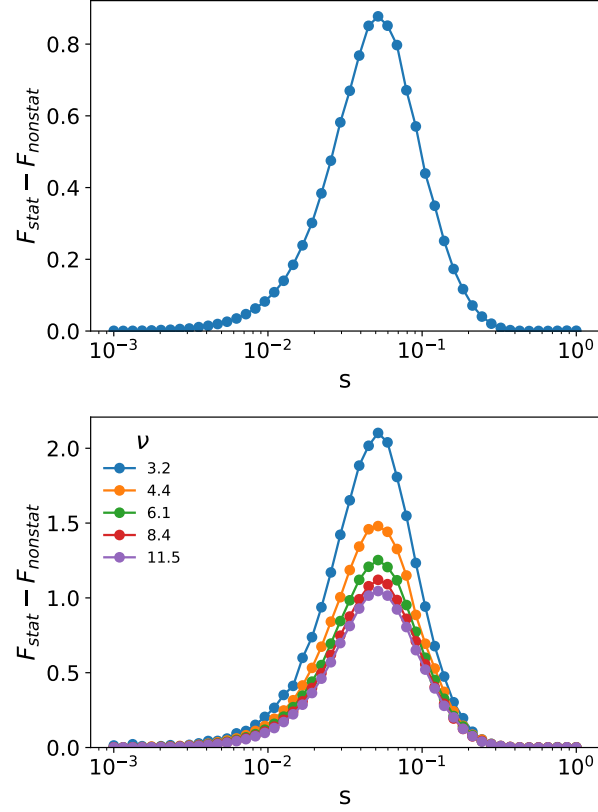


Figure 11: Non-stationary benchmark. The left plot refers to normal variable, the right plot to t-Student with ν degrees of freedom. The plots show the difference of Frobenius distance with the test covariance of the stationary minus the non-stationary estimator. The x-axis is the average amount of eigenvector rotation among two consecutive time-step.

In case of two multivariate central normal distribution data with respective covariance matrices Σ_{test} and Σ_{\bullet} , eq. (29) reduces to

$$KL(test||\bullet) = \frac{1}{2} \left(\text{tr} [\Sigma_{\bullet}^{-1} \Sigma_{test}] - n + \log \frac{|\Sigma_{test}|}{|\Sigma_{\bullet}|} \right) \quad (30)$$

Unfortunately, there is no closed analytical expression for the KL divergence for multivariate t-distributions

The average KL divergence in base $\log(n)$ over the validation window [2006, 2018] is reported in Fig. 12. It is important to remark that Eq. (30) is not defined if one of the two covariance matrices has a least one null eigenvalue. While the RIE is always positive-defined, it is possible that the test covariance matrix is not. Those cases were not considered in our analysis.

The results for the KL divergence reported for the time-ordered data are compatible with those for the Frobenius norm: the AO estimator outperforms NLS, with an optimal calibration window around $\delta_{train} = 180$ days. For the stationarized case, we observe an important difference from the Frobenius norm results since AO outperforms NLS for small calibration windows $\delta_{train} < 300$ days. The behavior of both curves, in this case, is monotonically decreasing, which is expected since, in a stationary world, more data always yields a better estimation. We suspect that such difference might be due to a higher sensitivity of the KL divergence to overfitting in the train window, thus a small δ_{train} should affect more NLS than AO. The fact that NLS outperforms AO for large δ_{train} in the stationarized case is to be expected, as NLS is optimal in the asymptotic regime of n and δ_{train} large for stationary data.

G Optimal Test Window Length of the Estimators

In this section, we explored how the performance changes as the length of the test window δ_{test} varies.

In the upper panel of Fig. 13 we show the Frobenius norm between the estimators and the test covariance. For very small time-horizons, all estimators reach a high distance with the test covariance. This is probably due to the high sample size error on the test covariance if it is computed over a short time-window. In addition, we observe that the Frobenius norm of the NLS estimator has a minimum around $\delta_{test} = 100$ days, a further increase of δ_{test} decreasing its performance. On the other hand, the AO estimator is much more stable for large δ_{test} . This supports the intuition that AO can really extract the stationary part of the system evolution.

Looking at the volatility of the GMV portfolio, we observe that the global minimum is reached for all estimators at the smallest δ_{test} then it increases approximately linearly with δ_{test} . The latter is not in contradiction with the results of the Frobenius norm since the global optimal minimum of the volatility changes as δ_{test} increases. In simple words, with a single portfolio held for two

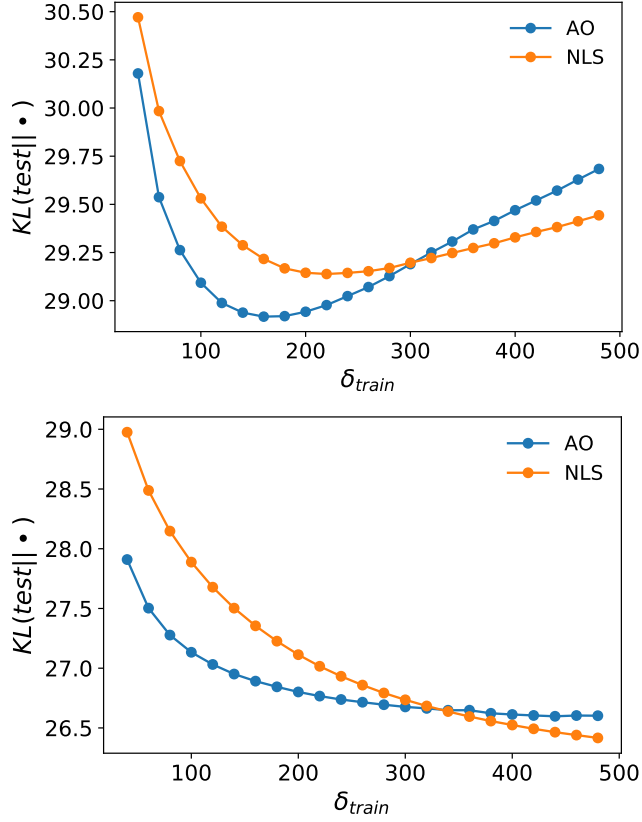


Figure 12: KL divergence between the estimators and test covariance versus the calibration window size δ_{train} . The upper plots refer to the original data set; the lower plot to the stationarized data. The metric is an average over 100 random samples of $n = 100$ assets for every day of the out-of-sample period; US equities. The KL divergence is computed in base n .

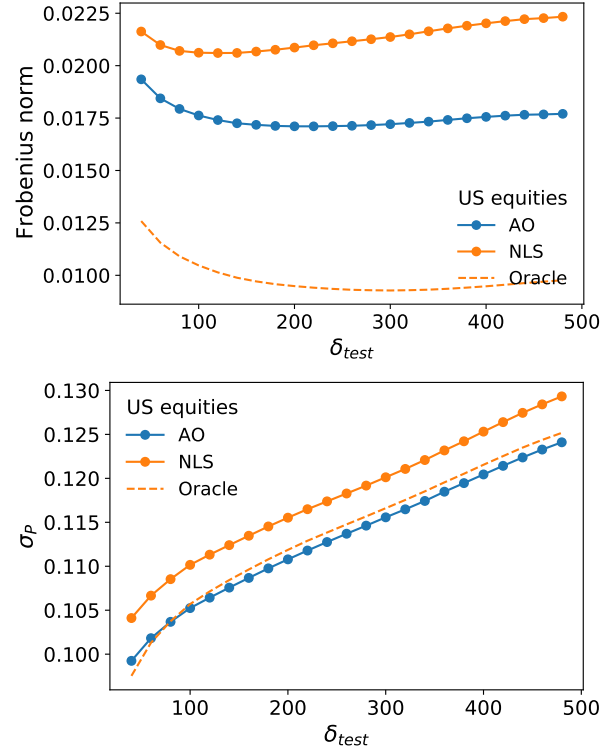


Figure 13: Out-of-sample performance measures of the filtered covariance matrices as a function of the test window length. The upper plot refers to Average Frobenius distance; the lower plot to the realized volatility of GMV portfolio. We fixed $\delta_{train} = 200$ days, and $n = 100$ random assets are selected 100 times for each day of the out-of-sample period.

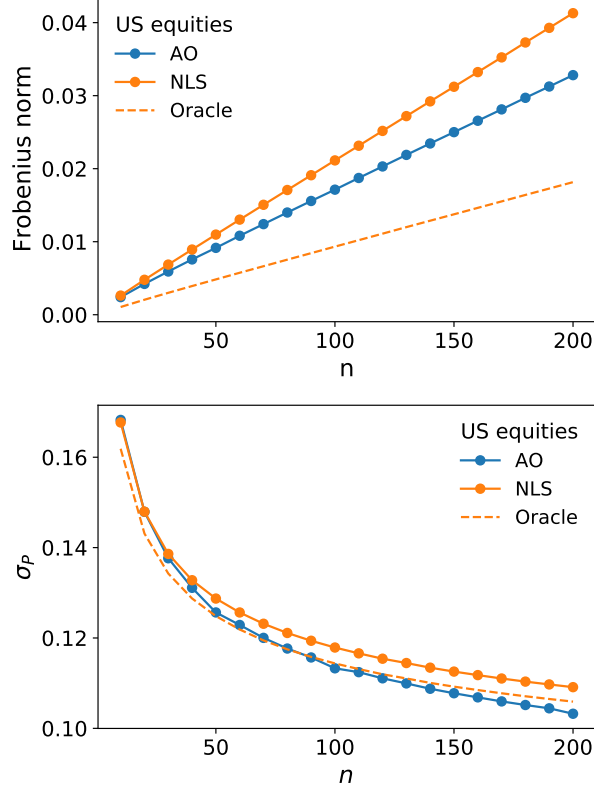


Figure 14: Out-of-sample performance measure of the filtered covariance matrices as a function of the system size. The upper plot refers to average Frobenius distance; the lower plot to the realized volatility of GMV portfolios. We fixed $\delta_{train} = 200$ days and $\delta_{test} = 252$ days, n random assets are selected 100 times for each day of the out-of-sample period.

years, it is not possible to reach the same low volatility of a weekly updated portfolio.

H System size dependence

In this section, we explored how the performance changes as the system size n varies.

In the upper panel of Fig. 14 we show the Frobenius norm distance for the three estimators. As n increases, the distance between the three estimators and the test covariance matrix increases. This is expected since the eigenvalue correction is only applied to $O(n)$ degrees of freedom, whereas the Frobenius norm increases as $n(n-1)/2$. It is worth remarking that the AO estimator

outperforms the NLS for all the values of n

In the lower panel of Fig. 14 we show the GMV portfolio volatility as a function of n . As in the previous case, changing the number of stocks changes the optimal minimum reachable on the test window: the larger n , the more possibilities one has to obtain a low-volatility portfolio. In addition, we observe that the relative discrepancy between the AO and NLS estimator increases with n .

I Hong Kong Stock Exchange

We took the AO eigenvalues calibrated with US equities data and used them to filter covariance matrices from the Hong Kong stock exchange data from the [2004-01-01,2017-06-23] period. In Fig. 15 we show the Frobenius distance between the covariance estimator and the out-of-sample covariance matrix. As for US equities, the AO provides a better estimator of the out-of-sample covariance matrix for the regular time-series, while being worse for stationarized data.

In Fig. 16, we show the realized variance of global minimum portfolios. As for the US equities, the AO yields lower variance than the Oracle for short calibration windows; AO also beats NLS for all the calibration window lengths that we tested, both for the regular and stationary case.

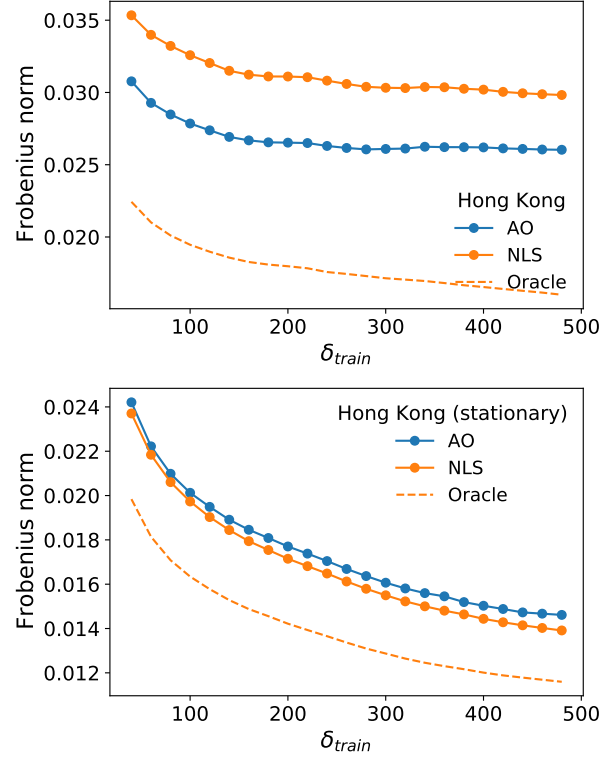


Figure 15: Average Frobenius distance between the filtered and test covariance matrices as a function of the calibration window length in the out-of-sample period. The upper plot refers to the original data set; the lower plot to the stationaryized data. 100 portfolios with $n = 100$ random assets are computed for each day of the out-of-sample period.

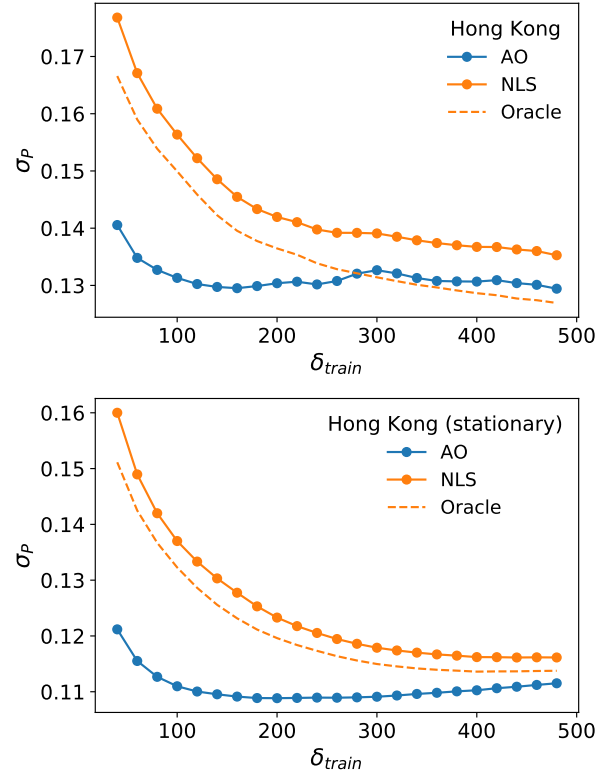


Figure 16: Average realized volatility of Global Minimum Variance portfolios as a function of the calibration window length. The upper plot refers to the original dataset, the lower plot to the stationarized data. 100 portfolios with $n = 100$ random assets are computed for each day of the out-of-sample period.