Factor Models for Futures Contracts to Improve Estimation of the Correlation Matrix

Ellen Ek

Master's thesis
2022:E40

Lund University

Faculty of Engineering
Centre for Mathematical Sciences
Mathematical Statistics

Centrum Scientiarum Mathematicarum

Master's Theses in Mathematical Sciences 2022:E40
ISSN 1404-6342

LUTFMS-3444-2022

Mathematical Statistics
Centre for Mathematical Sciences
Lund University
Box 118, SE-221 00 Lund, Sweden

http://www.maths.lu.se/

Abstract

In this paper regularization of the correlation matrix between futures contracts is examined. With starting point in the recently established HPCA framework (Avellaneda, 2019), a couple of different extensions to the one-factor model is suggested. Extensions are made in terms of adjusting the model according to different cluster structures. The data consists of futures contracts on a wide variety of underlying assets. Naturally they can be partitioned by asset class, asset sub class and/or region. The considered asset classes are equities, bonds, FX and commodities. Equities, bonds and FX are further partitioned by region - Europe, North/Latin America and Asia/Oceania. Commodities are partitioned into metals, energies and agriculturals. Metals are also divided into precious metals and industry metals, and agriculturals are divided into grains, livestock and miscellaneous agriculturals. These clusters are modelled both hierarchically and non-hierarchically where region is considered a second dimension rather than a child cluster. A completely different approach to HPCA is also presented, which is based on the assumption that sparseness in the eigenvectors is favourable. The proposed methods to modify the correlation matrix are evaluated with respect to ability of predicting eigenportfolio risk, interpretability/sparseness of eigenvectors and portfolio performance. Three different allocation methods are applied - minimum variance, mean-variance and equal risk contribution. All proposed methods turns out to predict eigenportfolio risk very well. Sparseness of the eigenvectors vary significantly between the different methods. The methods based on the sparseness-assumption turns out to perform best regarding Sharpe Ratio in portfolio performance.

Keywords: Correlation matrix, Hierarchical Principal Component Analysis, Factor Model, Clusters, Portfolio Optimization, Futures Contracts

Acknowledgements

I am very grateful for the opportunity of writing this master thesis in cooperation with Lynx Asset Management. A special thanks to Tobias Rydén and Florian Koch for their great ideas and seemingly unlimited knowledge. Also, a big thanks to Magnus Wiktorsson - my academic supervisor, for all valuable feedback throughout this project.

Contents

1 Introduction

Within portfolio optimization, the covariance matrix is often used to derive optimal portfolio weights. A major problem is that a covariance matrix estimated from financial return series data, in general is very noisy.

A popular method to regularize the covariance matrix is to apply some kind of shrinkage. However, this method doesn't add any more information about the correlation structure, it merely makes the matrix more well-behaved in general. Another approach to solve this problem could therefore be to include some prior knowledge about the correlation structure in the estimator. A method called hierarchical principal component analysis (HPCA), involving a factor model based on PCA and some partitioning of the assets, has been successfully tested in terms of interpretability on US stocks (S&P500) partitioned by sector (Avellaneda, 2019) [1]. It is however not tested on larger trading universes, for example futures contracts, which behave a little differently. It is neither tested on more complex partitionings as for example hierarchical or two-dimensional clusters. In this paper it will therefore be examined how the HPCA framework can be extended to more complex cluster structures and if those can be successfully applied on futures contracts.

More specifically, it will be examined if HPCA applied on assets partitioned by asset class, asset sub class and region, both hierarchically and non-hierarchically, can improve performance of the correlation matrix in terms of interpretability and portfolio performance. Also a somewhat different approach to HPCA will be implemented, based on the assumption that interpretability and sparseness can improve performance.

1.1 What Correlation Structures to Utilize?

Since HPCA is based on some predefined clusters, we first need to form these. In this paper the market universe consists of 100 different futures contracts on various underlying assets, ranging from commodities such as metals, energies and agriculturals - to FX, stocks and bonds tied to different regions such as Europe, North America, Latin America, Asia and Oceania. A more detailed description of the data can be found in section 4.3 and a complete list over the assets can be found in Appendix A.

Clusters should then be formed according to the modeller's own beliefs regarding common driving factors behind the returns (Avellaneda, 2019) [1]. Two natural driving forces could for instance be regional factors or asset class-specific factors - which is exactly what we will use here.

However, one could of course form clusters based on some machine learning technique, but a main assumption in this paper is that we can improve performance by manually deciding what clusters to use. An argument for this could for example be that our data is too noisy to yield reliable clusters with a data-driven approach.

Nevertheless, even though we are suspecting that there are some common driving factors behind different asset classes and different regions it is not obvious how we should translate those into clusters. One quite natural way to form clusters is in a hierarchical way. In our data there are four main asset classes: bonds, equities, commodities and FX. Commodities can be further partitioned into agriculturals, energies and metals. Agriculturals can be even further partitioned into grains, livestock and miscellaneous agriculturals. Also metals can be partitioned once more, into precious metals and industrial metals. Regarding regions, our data provides futures contracts with underlying assets/currencies from Europe, North America, Latin America, Asia and Oceania. Since Latin America and Oceania are quite small economies, we may suspect that Latin America

and North America may be modeled as one common region, as well as Oceania and Asia as one common region. However, since most commodities are produced/harvested all over the world it doesn't make much sense to model any regional covariation within this asset class. An overview of these hierarchical clusters can be found in Fig. 1 below.

A natural question to consider now is whether the regional components from different asset classes could originate from the same source, and if so - maybe it is better to model the clusters in two dimensions, ie. asset class and region non-hierarchically? An illustration of this structure can be found in Fig. 2 below.

One last suggestion could be to keep it simple and skip the regional clusters, meaning that we only utilize the left (red) part of the cluster structure in Fig. 2.

In the following section some different methods to modify the correlation matrix based on these clusters will be presented.

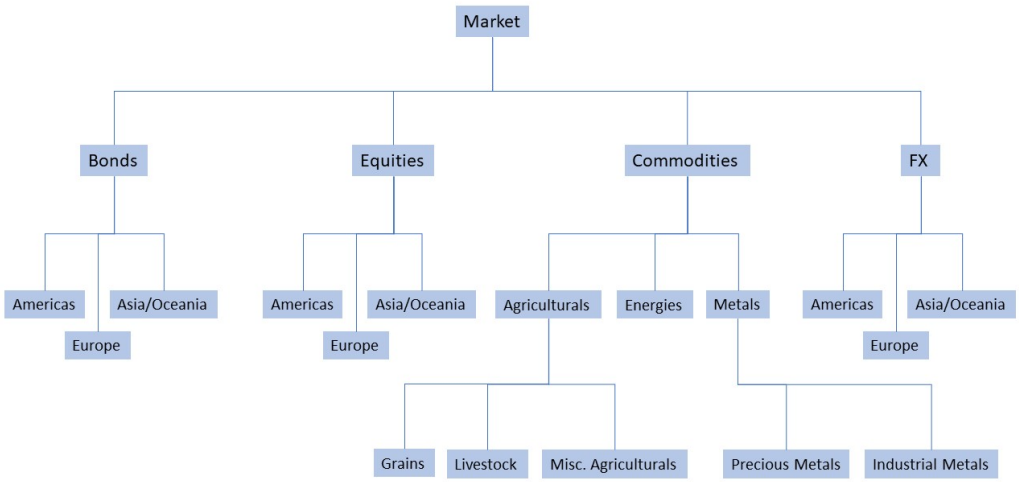


Figure 1: Suggestion of a hierarchical correlation structure. Each asset belongs to only one cluster.

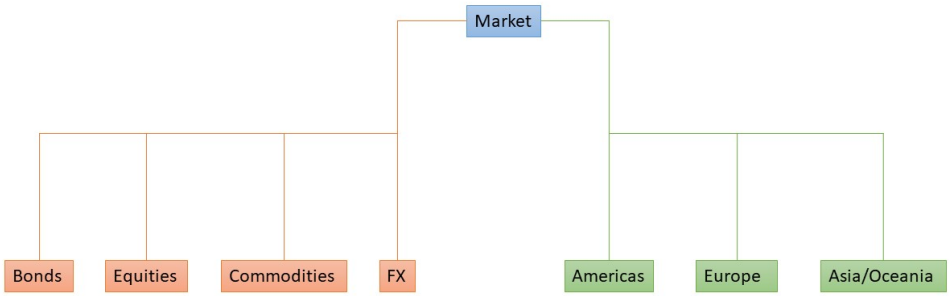


Figure 2: Suggestion of a non-hierarchical correlation structure. Assets can belong to both an asset class cluster (red) and a regional cluster (green).

2 Hierarchical PCA

The main idea behind hierarchical principal component analysis is to modify the correlation matrix by partitioning the assets into clusters and replace the correlation between assets from different clusters with a measure of the correlation between the clusters (Avellaneda, 2019) [1].

2.1 HPCA with Non-Hierarchical Partitioning

The most simple approach to HPCA goes under non-hierarchical partitioning since it involves only one level of clusters, for example asset classes. A slightly more complex approach is to have multidimensional clusters, meaning that all assets can belong to multiple clusters from different partitionings, for example one asset class and one region. Both approaches will be thoroughly explained in the following sections.

2.1.1 One-Dimensional Partitioning

The simplest way to model correlation structures is presumably a partitioning in one dimension with no sub-groups, for example by asset class only (left/red part of Fig. 2). Correlation between assets from different clusters can then be measured as the correlation between the first principal component of the clusters, as proposed by Avellaneda (2019) [1]. This can be achieved by describing the returns of each asset with a one-factor model consisting of the strongest principal component associated with its cluster. An important note here is that the following model only applies to assets from different clusters, for assets in the same cluster we do not want to modify the correlations.

Nevertheless, with N observations of M assets collected in an N x M return series matrix X, then for asset j we can describe its N x 1 time series of returns X_j, as

$$X_j = U^{(1)}_{I(j)}\beta_j + \epsilon_j, \tag{1}$$

where $U^{(1)}_{I(j)}$ is the N x 1, first principal component of cluster $I(j)$, β_j is the regression coefficient for asset j and ϵ_j is some noise. The noise ϵ_j is by definition uncorrelated with the principal component $U^{(1)}_{I(j)}$, but not necessarily uncorrelated with the noise-term of other assets. However, an important assumption here - the so-called 'HPCA-assumption', is that the residuals of two assets from different clusters ϵ_i and ϵ_j are uncorrelated (Avellaneda, 2019) [1], ie.

$$\text{if } I(i) \neq I(j), \text{ then } Corr(\epsilon_i, \epsilon_j) = 0.$$

The principal component $U^{(1)}_{I(j)}$ is obtained from a singular value decomposition of the returns from all assets in cluster $I(j)$. More about the singular value decomposition can be found in 'Matrix Theory' (Holst & Ufnarovski, 2014) [4]. For any cluster K with k constituents, we have

$$X_K = U_K S_K V_K^T, \tag{2}$$

where U_K is a matrix of size N x k with the principal components, S_K is a diagonal matrix of size k x k with the singular values and V_K is a matrix of size k x k with the component weights, all of them associated with cluster K. Since all principal components are orthogonal to each other, the β-coefficients for all assets in cluster K can be obtained without any regression from

$$\beta_K = S^{(1)}_K V^{(1)T}_K,$$

where β_K is a 1 x k vector, $S^{(1)}_K$ is the largest (first) singular value and $V^{(1)}_K$ is the first column of the component weights (size k x 1), all of them associated with cluster K.

However, the purpose of this paper is not to construct a factor model for the raw return series themselves, but rather to use a factor model to modify the correlation matrix. It is therefore more convenient to express everything in terms of the correlation matrix instead of raw returns.

That also allows for use of some fancier method than the basic sample covariance estimator when computing the raw correlations in the first place. A popular, slightly more decent method is for instance an exponentially weighted moving average.

A useful property of the correlation matrix is that it is real and symmetric, meaning that a singular value decomposition is equivalent to an eigenvalue decomposition (Holst & Ufnarovski, 2014) [4] - where the eigenvectors nicely coincide with the component weights V (from the SVD of the return series matrix X) and the eigenvalues D with the squared singular values S^2 (also from the SVD of the return series matrix X). A singular value decomposition of the correlation matrix consisting only of the assets within a specific cluster K can hence be written as

$$\Sigma_K = V_K D_K V_K^T,$$
$$D_K = S_K^2.$$

Meaning that the β-coefficients for all assets in cluster K can be obtained from the correlation matrix instead of the raw return series according to

$$\beta_K = \sqrt{D^{(1)}_K} V^{(1)T}_K.$$

We can also express the first principal component $U^{(1)}$ in terms of V and D, by rewriting Eq. 2 as

$$U_K^{(1)} = X_K V_K^{(1)} S_K^{(1)-1} = X_K V_K^{(1)} D_K^{(1)-1/2}. \tag{3}$$

To simplify notation we introduce F_K, which is a column vector of length M with the entries of $V_K^{(1)} D_K^{(1)-1/2}$ on the rows corresponding to the assets of cluster K and zeros elsewhere. This allows us to write

$$U_K = XF_K.$$

Finally we have everything we need to reconstruct the correlation matrix according to the factor model described by Eq. 1, in terms of the raw correlation matrix Σ. By placing all coefficients in one column vector β of size M x 1 and all factors F_K from each cluster K adjacent to each other (column wise) in a larger matrix say F of size M x [number of clusters], the modified correlation matrix $\tilde{\Sigma}$ can be written as

$$\begin{aligned}
\tilde{\Sigma} &= \frac{1}{N} X^T X \\
&= \frac{1}{N} (XF\beta)^T (XF\beta) \\
&= \frac{1}{N} \beta^T F^T X^T XF\beta \\
&= \beta^T F^T \Sigma F\beta.
\end{aligned}$$

Note that all residuals have vanished due to the HPCA-assumption. The final step is then to replace all internal correlations within each cluster, with the corresponding raw correlations from Σ. In summary, the following expression describes each element in the modified correlation matrix:

$$\tilde{\Sigma}_{i,j} = \begin{cases} \beta_i F_{I(i)}^T \Sigma F_{I(j)} \beta_j & \text{if } I(i) \neq I(j) \\ \Sigma_{i,j} & \text{otherwise} \end{cases}$$

Keep in mind here that F_K is filled out with zeros everywhere except for the rows corresponding to the constituents of cluster K.

An advantage with this model is the simpleness - only one dimension, one layer of clusters and only one factor to explain cluster correlations. This also leads to sparse and interpretable eigenvectors, which is proved in Avellaneda (2019) [1].

2.1.2 Two-Dimensional Partitioning

One extension of the aforementioned one-dimensional approach found in Avellaneda (2019) [1] is to include some other partitionings of the assets. An interesting application of such a model regarding futures contracts could be to partition assets both in terms of asset class but also in terms of region, non-hierarchically. See for example Fig. 2. The following theory holds for multiple dimensions/partitionings but is presented in two dimensions to simplify notation. The extended factor model describing the returns X of asset j now consists of two factors, ie.

$$X_j = U_{I(j)}^{(1)} \beta_{1,j} + U_{J(j)}^{(1)} \beta_{2,j} + \epsilon_j, \tag{4}$$

where $I(j)$ and $J(j)$ denotes the two different clusters asset j belongs to. The HPCA-assumption must also be extended, we suggest that for two different assets i and j

$$\text{if } I(i) \neq I(j) \text{ or } J(i) \neq J(j), \text{ then } Corr(\epsilon_i, \epsilon_j) = 0.$$

Furthermore, unfortunately the two principal components $U^{(1)}_{I(j)}$ and $U^{(1)}_{J(j)}$ in Eq. 4 are not necessarily orthogonal to each other meaning that we cannot find the β-coefficients directly from the singular value decomposition as earlier. But simple regression will do the work. To find the regression coefficients we therefore solve for each asset j, the following minimization problem

$$\beta_j = \arg\min_{\beta_j} \frac{1}{2}||X_j - U_{I(j),J(j)}\beta_j||_2^2, \tag{5}$$

where $\beta_j = [\beta_{1,j}\ \beta_{2,j}]^T$ is a vector containing the two regression coefficients and $U_{I(j),J(j)} = \left[U^{(1)}_{I(j)}\ U^{(1)}_{J(j)}\right]$ is a matrix whose columns consist of the first principal component of cluster $I(j)$ and $J(j)$ respectively.

For the same reason as earlier, we would rather work with the correlation matrix than the raw return series, meaning that we would rather do the regression in terms of Σ than X. Similar to the one-dimensional case we rewrite U as

$$\begin{aligned} U_{I(j),J(j)} &= \left[U^{(1)}_{I(j)}\ \ U^{(1)}_{J(j)}\right] \\ &= \left[X_{I(j)}V^{(1)}_{I(j)}{D^{(1)}_{I(j)}}^{-1/2}\ \ X_{J(j)}V^{(1)}_{J(j)}{D^{(1)}_{J(j)}}^{-1/2}\right], \end{aligned}$$

and introduce $F_{I(j),J(j)}$, which now is a M x 2 matrix consisting of the entries of the two column vectors $V^{(1)}_{I(j)}{D^{(1)}_{I(j)}}^{-1/2}$ and $V^{(1)}_{J(j)}{D^{(1)}_{J(j)}}^{-1/2}$ on the rows corresponding to the assets of cluster $I(j)$ and $J(j)$ respectively, zeros elsewhere. This allows us to express U as

$$U_{I(j),J(j)} = XF_{I(j),J(j)}.$$

We can now rewrite the minimization problem in Eq. 5 as

$$\begin{aligned} \beta_j = \arg\min_{\beta_j} \frac{1}{2}||X_j - U_{I(j),J(j)}\beta_j||_2^2 = \\ \arg\min_{\beta_j} \frac{1}{2}(X_j - U_{I(j),J(j)}\beta_j)^T(X_j - U_{I(j),J(j)}\beta_j) = \\ \arg\min_{\beta_j} \frac{1}{2}(X_j^TX_j - 2X_j^TU_{I(j),J(j)}\beta_j + (U_{I(j),J(j)}\beta_j)^TU_{I(j),J(j)}\beta_j) = \\ \arg\min_{\beta_j} \frac{1}{2}(X_j^TX_j - 2X_j^TXF_{I(j),J(j)}\beta_j + \beta_j^TF^T_{I(j),J(j)}X^TXF_{I(j),J(j)}\beta_j) = \\ \arg\min_{\beta_j} \frac{N}{2}(\Sigma_{j,j} - 2\Sigma_{j,:}F_{I(j),J(j)}\beta_j + \beta_j^TF^T_{I(j),J(j)}\Sigma F_{I(j),J(j)}\beta_j) = \\ \arg\min_{\beta_j} (\Sigma_{j,:}F_{I(j),J(j)}\beta_j + \frac{1}{2}\beta_j^TF^T_{I(j),J(j)}\Sigma F_{I(j),J(j)}\beta_j), \end{aligned}$$

which is a quadratic problem that can be solved analytically. After differentiating and setting the expression equal to zero, the solution is found to be

$$\beta_j = (F^T_{I(j),J(j)}\Sigma F_{I(j),J(j)})^{-1}(\Sigma_{j,:}F_{I(j),J(j)})^T.$$

As in the one-dimensional case, we place all coefficients β_j adjacent to each other (row wise) in a larger matrix β and all $F_{I(j),J(j)}$ adjacent to each other (column wise) in a larger matrix F, and calculate the modified correlation matrix $\tilde{\Sigma}$ according to

$$\tilde{\Sigma} = \beta^T F^T \Sigma F \beta.$$

The final step is then to replace all internal correlations within each cluster, with the corresponding raw correlations from Σ. In summary, the following expression describes each element in the modified correlation matrix:

$$\tilde{\Sigma}_{i,j} = \begin{cases} \beta_i F^T_{I(i),J(i)} \Sigma F_{I(j),J(j)} \beta_j & \text{if } I(i) \neq I(j) \text{ or } J(i) \neq J(j) \\ \Sigma_{i,j} & \text{otherwise} \end{cases}$$

Keep in mind here that each column in $F_{K,L}$ is filled out with zeros everywhere except for the rows corresponding to the constituents of cluster K and L respectively.

A possible disadvantage with this model when using for example an asset class partitioning and a regional partitioning is that some non-interpretable cross-correlation terms appear. Since the model consists of both an asset class term and a regional term, when we estimate correlations this will result in a sum of one asset class correlation term, one regional correlation term and two cross terms describing correlations between each asset class and region.

2.2 HPCA with Hierarchical Partitioning

There are many ways to incorporate a more hierarchical structure in the correlation matrix. Obviously all of them require some predefined tree structure of the clusters, see for example Fig. 1. Here, two different suggestions on how to make use of a hierarchical partitioning will be presented. Both based on the one-dimensional non-hierarchical model proposed by Avellaneda (2019) [1].

2.2.1 One-Factor Model

One way to incorporate this structure into the correlation matrix is to estimate correlations as the correlation between the first principal component of the closest relevant branch. To be more specific, the returns X of asset j to be used for estimating the correlation with asset i, can be written as a one-factor model according to

$$X_j = U_{I(j,i)} \beta_j + \epsilon_j,$$

where $I(j, i)$ returns the child cluster (that asset j belongs to) of the smallest cluster that both asset j and i belong to. Let us illustrate with an example.

When estimating the correlation between a European bond and an American bond, the first principal component of European bonds are used to describe the European bond and the first principal component of American bonds are used to describe the American bond since these are the child clusters of the smallest common cluster (bonds). If instead a European bond is compared to a precious metal, then the first principal component of bonds is used to describe the European bond and the first principal component of commodities is used to describe the precious metal as those are the child clusters of the smallest common cluster (the market), see again Fig. 1.

The intuition behind this model is that correlation between assets further away from each other in the tree should be estimated with some coarser method, in this case a principal component from a larger subgroup.

2.2.2 'All'-Factor Model

An extension of the aforementioned one-factor model could be to utilize all first principal components of each cluster as well as child clusters. The returns X of asset j would then be described by a linear combination of all k first principal components asset j is related to,

$$X_j = U_{I(1,j)}\beta_{1,j} + U_{I(2,j)}\beta_{2,j} + \ldots + U_{I(k,j)}\beta_{k,j},$$

regardless of which other asset the correlation is to be estimated with. Here, $I(i,j)$ returns the i:th cluster of asset j. For instance, a European bond would have 'Bonds' as first cluster and 'Europe' as second cluster. A precious metal would have 'Commodities', 'Metals' and 'Precious Metals' as first, second and third cluster, respectively - although the order doesn't matter here, it's still just a linear combination. See once again Fig. 1 for an example of correlation structure.

A consequence of this, is that deeper branches yield a more detailed model of the returns, which may (or may not) be favourable. A reason to prefer this model could for example be that the more clusters an asset can belong to, the more detailed its returns should be described. Another property of this model is that returns of an asset are described with the same factors regardless of what other asset the correlation is to be estimated with.

2.3 An 'Eigenvector Sparseness'-Approach to HPCA

Another quite different suggestion on how to extend the simple HPCA model proposed by Avellaneda (2019) [1], is to rely on the assumption that the eigenvectors of the denoised correlation matrix should be sparse. The other models are mainly based on the assumption that a coarser estimator of correlation between economically more or less unrelated assets is favourable - resulting in hopefully sparse eigenvectors. What if we instead *force* this property onto the correlation matrix? That's the purpose of this approach. For clarification, this framework can be used regardless of whether the clusters are formed hierarchically or non-hierarchically.

2.3.1 The Basic Framework

The main idea is to formulate the problem as a minimization of the difference between the unmodified and the modified correlation matrix, under the constraint that the eigenvectors of the modified correlation matrix have some predefined sparse structure. For interpretability reasons, it could be desirable that some eigenvectors have non-zero elements only on a subset of the assets. For example some eigenvectors describing intra-cluster relationships and some other describing inter-cluster relationships – with zeros elsewhere. A convenient way to express these constraints is according to the following.

Let G be the matrix consisting of all relevant eigenvectors. Depending on how we want to model correlations it could for example be the eigenvectors of each cluster (if modeled non-hierarchically) or the first eigenvector of each (possibly) hierarchical cluster. It should be mentioned here that even though we discuss cluster correlations in terms of eigenvectors, we implicitly mean principal components. A reminder of the relationship between those can be found in Eq. 3.

Another important aspect to consider here is whether we want to span the full space with these eigenvectors or if we are satisfied with a low-rank approximation. If we have non-hierarchical clusters and put all eigenvectors from each cluster in the matrix G, then those will span the full space. On the other hand, if we have hierarchical clusters and want one or more eigenvectors from each cluster, it is a little trickier to span the full space as some clusters overlap. Let us illustrate

with two examples, the first one with non-hierarchical clusters and the second one with hierarchical clusters.

In the case of non-hierarchical clusters, say A, B and C, we can just put all eigenvectors from each cluster V_A, V_B, V_C into G without any further considerations, see Fig. 3 below.

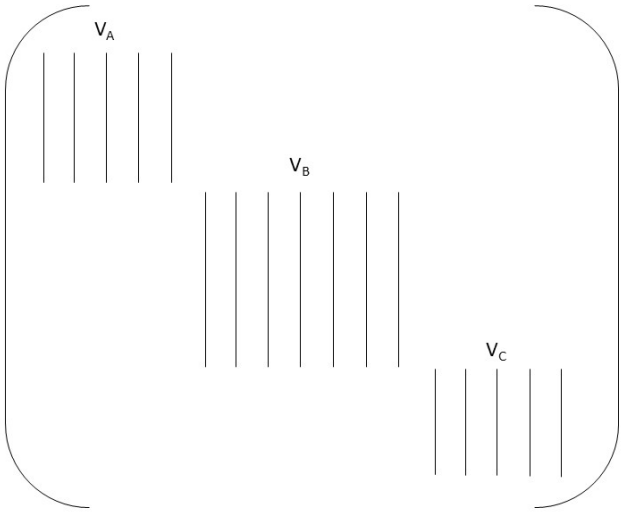


Figure 3: An example of the matrix G when composed by eigenvectors from non-hierarchical clusters A, B and C. Each line represents an eigenvector - zeros everywhere else.

Then, by imposing some specific constraints on $G^T\tilde{\Sigma}G$, where $\tilde{\Sigma}$ is the resulting modified correlation matrix, an arbitrary sparse eigenvector structure can be achieved. Before stating the constraints we let Q denote the resulting matrix from the matrix multiplication $G^T\tilde{\Sigma}G$ and take a quick look at the appearance of this matrix Q. If we reuse the earlier example of G with non-hierarchical clusters (Fig. 3) and the principle of using the first eigenvector to estimate cluster correlations (as proposed by Avellaneda, 2019) [1], we will get the structure in Fig. 4.

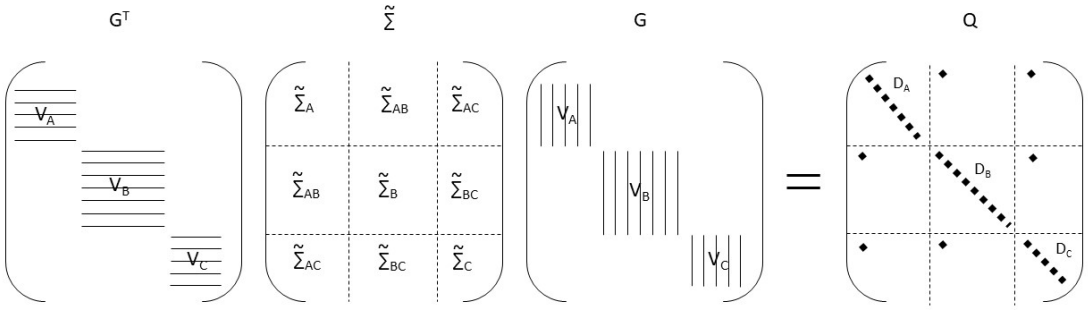


Figure 4: An example of the block structure in $G^T\tilde{\Sigma}G$.

This picture of course needs some further explanation. Since G is built from a nice collection of eigenvectors extracted from different parts of the correlation matrix, one could suspect that some eigenvalues D would appear in Q.[1] This is very accurate. Since we have a very clear block structure in G and $\tilde{\Sigma}$, we can find a block structure in the resulting matrix as well. All eigenvalues from each cluster (D_A, D_B, D_C) will appear as blocks on the main diagonal. In the off-diagonal elements of Q, associated with the inter-cluster correlations, we have another case. If we for instance take the top right block of Q which corresponds to the block multiplication $V_A^T\tilde{\Sigma}_{AC}V_C$, we will only get one non-zero element. This is because we decided earlier that cluster correlations should be described by the first eigenvectors only, in this case $V_A^{(1)}$ and $V_C^{(1)}$. This in turn means that $\tilde{\Sigma}_{AC}$ will be orthogonal to all other eigenvectors from cluster A and C, yielding zeros everywhere except in the top left corner where we have $V_A^{(1)T}\tilde{\Sigma}_{AC}V_C^{(1)}$. The same goes for all off-diagonal blocks, ie. one non-zero element in the top left corner and zeros elsewhere.

However, when we move on the case of hierarchical clusters there are some different options. Let us consider the same example as above but now B and C are child clusters to a new cluster, say D. One way to incorporate this new information is to utilize the first eigenvector of cluster D, denoted $V_D^{(1)}$. A question that arises is then if we really want to use the eigenvectors of B and C as they are, since they partly contain the same information as the first eigenvector of cluster D. A simple solution would be to remove the information contained in D by for example projection and then extract the eigenvectors from what is left, now called $\tilde{V}_B$ and $\tilde{V}_C$. If it is desirable to have a fully spanned space, then we must use one eigenvector less than earlier from each cluster and fill up with vectors that are orthogonal to both the first eigenvector of cluster D and the eigenvectors

[1] Remember that $V^T\Sigma V = D$.

of what is left in cluster B and C. In the case of two child clusters, there will be only one 'extra' vector. See illustration in Fig. 5.

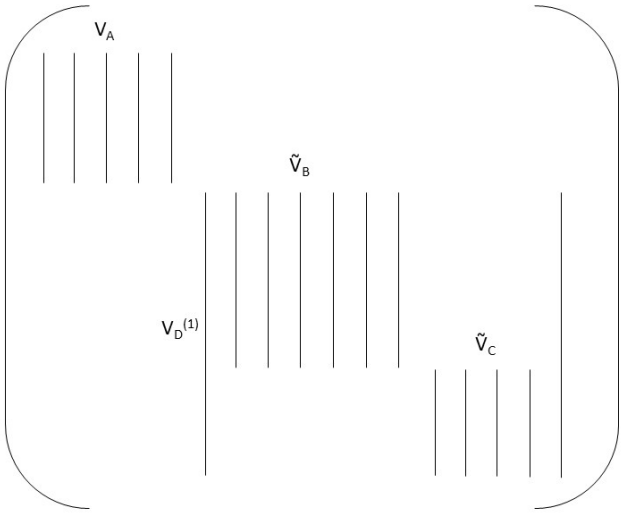


Figure 5: An example of the matrix G when composed by eigenvectors from hierarchical clusters, each line representing an eigenvector - zeros everywhere else.

For completeness an example of how to structure G in the case of triple layer hierarchical clusters can be found in Appendix B.

However, with G constructed according to Fig. 5 we get a slightly more complex structure of Q, see Fig. 6. Here we also assume that the principle of using the first eigenvector to estimate cluster correlations is applied. The earlier observed block structure from the non-hierarchical clusters in Fig. 4 is very pronounced here as well but in a more hierarchical manner. Since we have two main clusters (A and D) we have four main blocks, and in one of the main clusters (D) we have two child clusters (B and C) resulting in some additional blocks inside. As in the hierarchical example, the blocks corresponding to inter-cluster correlations (here A vs. D and B vs. C) will only have one non-zero element, corresponding to $V_A^{(1)T}\tilde{\Sigma}_{AD}V_D^{(1)}$ and $\tilde{V}_B^{(1)T}\tilde{\Sigma}_{BC}\tilde{V}_C^{(1)}$ respectively. Also, the eigenvalues of cluster A, the eigenvalues of what's left of cluster B and C after projection and the first eigenvalue of cluster D will appear on the main diagonal for natural reasons ($V^T\Sigma V = D$). The last element which we have not yet mentioned, is the bottom right element on the main diagonal. This will correspond to the result of multiplying the so called 'extra' vector - which is orthogonal to both $V_D^{(1)}$, $\tilde{V}_B$ and $\tilde{V}_C$ - with $\tilde{\Sigma}_D$ from both left and right. This element does not really have any interpretation, so if we are satisfied with a low-rank approximation of the correlation matrix then this is probably the first element to disregard.

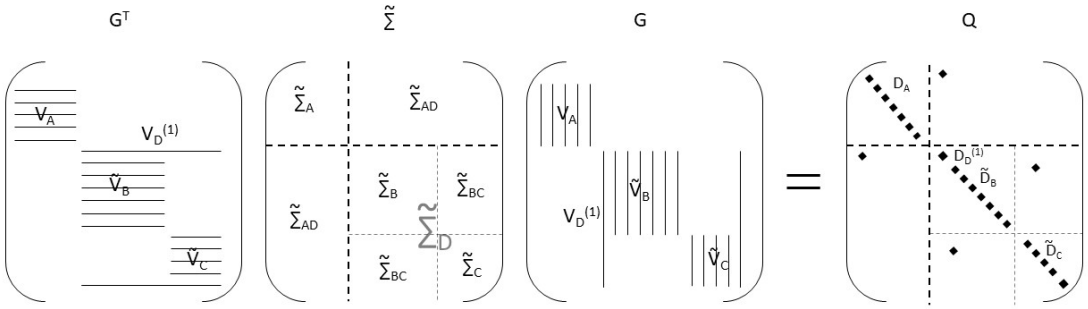


Figure 6: An example of the block structure in $G^T \tilde{\Sigma} G$.

How is this related to the eigenvectors of $\tilde{\Sigma}$? Well, from empirical studies we have found that the eigenvectors of $\tilde{\Sigma}$ are built from the cluster eigenvectors corresponding to the non-zero elements in Q. More precisely, each column (or row - it's symmetric) in Q will describe what cluster eigenvectors will form each eigenvector of $\tilde{\Sigma}$. For instance, if we examine the non-zero elements of Q in Fig. 4. There are three columns with three non-zero elements, all of them referring to the first eigenvector of each cluster. We will thus have three $\tilde{\Sigma}$-eigenvectors built from a combination of the first eigenvector of each cluster. We also see that all other columns only have one non-zero element implying that the rest of the $\tilde{\Sigma}$-eigenvectors will be built from only one of the cluster eigenvectors each. Similarly for Q in Fig. 6, we can see that two $\tilde{\Sigma}$-eigenvectors will be built from the first eigenvectors of cluster A and D explaining the higher level cluster correlations, and two other $\tilde{\Sigma}$-eigenvectors will be built from the first eigenvectors of what is left of clusters B and C after projection explaining the lower level cluster correlations. The rest of the $\tilde{\Sigma}$-eigenvectors will be built from only one of the cluster eigenvectors each.

Comparing this with the eigenvectors of the raw correlation matrix, which not necessarily have *any* zero-elements - we have achieved a very sparse structure. Although it should be noted here that this sparse structure is achieved also in the one-dimensional partitioning model described in section 2.1.1 (proven by Avellaneda, 2019) [1], but not necessarily the other models since the cluster eigenvectors are somewhat distorted.

So, by formulating this approach as an optimization problem with the constraint of some elements in Q to be zero we can both determine exactly which cluster eigenvectors to use when approximating different parts of the correlation matrix and achieve an arbitrarily sparse eigenvec-

tor structure of $\tilde{\Sigma}$. An exact formulation of the problem is the following:

$$\begin{aligned} &\tilde{\Sigma} = \arg\min_{\tilde{\Sigma}} ||\Sigma - \tilde{\Sigma}||_{Fr}^2 \\ &\text{s.t} \\ &Q_{i,j} = 0, \ (i,j) \in M \\ &Q = G^T \tilde{\Sigma} G \end{aligned} \tag{6}$$

where M is the set of all matrix coordinates in Q we prefer to be zero and G is a matrix consisting of all cluster eigenvectors we want to utilize. Some examples of G and what elements to constrain in Q can be found in Fig. 3 - 6.

This method will be applied according to the inter-cluster correlations of the one-factor model described in section 2.2.1. We will however discard the non-interpretable elements (see example related to Fig. 6), resulting in a low-rank approximation. This means that Q for this model will have similar appearance as Q in Fig. 6 but without all non-interpretable elements on the main diagonal. The advantage of this method is interpretability. We will get both fewer and sparser eigenvectors. A possible disadvantage is that the main diagonal in the correlation matrix (ones) not necessarily will be preserved, since we do not span the full space. Strictly speaking we then no longer have a correlation matrix.

2.3.2 Introduction of an L1-Penalty for Further Sparseness

In the previous section we discovered a neat way to force sparseness into the eigenvectors of the modified correlation matrix $\tilde{\Sigma}$ by putting constraints on what cluster eigenvectors to use. What if some of them still are superfluous? By introducing an L1-penalty on the coefficients we may be able to achieve an even sparser structure by letting an optimizer decide what cluster eigenvectors to keep or remove when approximating different parts of the correlation matrix. By adding the L1-penalty to Eq. 6, we end up with the new problem formulation

$$\begin{aligned} &\tilde{\Sigma} = \arg\min_{\tilde{\Sigma}} ||\Sigma - \tilde{\Sigma}||_{Fr}^2 + \lambda \sum_{(i,j) \in N} |Q_{i,j}| \\ &\text{s.t} \\ &Q_{i,j} = 0, \ (i,j) \in M \\ &Q = G^T \tilde{\Sigma} G \end{aligned}$$

where M is the set of all matrix coordinates in Q we prefer to be zero and N is the set of all other coordinates in Q. A problem with numerical solvers and the L1-penalty is that the reduced elements in Q will only be *approximately* zero and not necessarily equal to zero. Consequently, some elements in the eigenvectors of $\tilde{\Sigma}$ will also be *approximately* zero, as well as some eigenvalues. This can have hazardous effects on portfolio performance[2] so we also have to introduce a threshold of what we consider to be zero in Q. We put this threshold as the smallest eigenvalue of the raw correlation matrix (as the entries of Q are closely related to eigenvalues) and let our optimizer do the thing once more. This time with the constraint of zeros also at the coordinates of Q that did not exceed the threshold and without the L1-penalty as we already know what coefficients were superfluous.

An optimal λ we will be searched for with respect to a qualitative analysis of the eigenvectors.

[2]See for example Hauser & Schmeltzer (2013) [7]

3 Analysis of the Principal Components and Their Ability to Model Cluster Correlations

Theoretically the first principal component of a cluster should describe the most pronounced behaviour and therefore be suitable for modelling inter-cluster correlations in the HPCA framework. Since the theory relies completely on this assumption, it is of great importance that we analyze the behaviour of those.

The principal components themselves are not very interesting to visualize, so we will instead look at the eigenvectors. This can be justified by Eq. 3 from earlier, which shows that the first eigenvector essentially describes the contribution of each asset to the principal component. So, by observing the behaviour of the eigenvectors we will understand how the assets are related to the principal components.

To verify that our choice of clusters is reasonable, we will also examine some eigenvectors from the full correlation matrix.

3.1 Analysis From an Asset Class Perspective

We earlier decided to form four main clusters based on asset classes, namely 'Bonds', 'Equities', 'Commodities' and 'FX'. To verify that the behaviour of the assets in each cluster is similar, we look at the first eigenvectors of the raw correlation matrix, see Fig. 7 below.

In the first two eigenvectors we can see some typical market behaviour, namely one vector with opposite directions of bonds and essentially all other assets ('Eigenvector 1') as well as one vector with the same direction for all assets ('Eigenvector 2'). This originates from investors acting differently depending on the state of the global economy. From one perspective, during good times investors tend to put their money into equities and other fairly risky assets, while during tougher time periods rather invest in safer assets like bonds. This creates a negative correlation between bonds and essentially all other assets, which usually is apparent in the first eigenvector, ie. explains most of the variance. From another perspective, during good times investors tend to invest in the market in general, while during tougher periods rather sell - regardless of whether the relevant assets are bonds or other types of assets. This creates another dimension of the variability in the correlation matrix, namely a slightly positive correlation between all assets - which appears as an eigenvector with the same sign for all elements. This behaviour is often a little less pronounced, thus ending up as the second eigenvector, ie. explains second most of the variance. However, for various reasons this 'buy all' or 'sell all' behaviour sometimes dominates the market, making the two first eigenvectors switch place. Keep this market behaviour in mind when examining the regional eigenvectors.

Further, in the third and the fourth eigenvector of Fig. 7 we can see that most assets within FX seem to covariate, as well as most assets within commodities. The common factor that drives all assets within FX in the same direction for this data set, is the US dollar. All currencies are written in US dollar which creates a strong dependency to the exchange rate of the USD. More about the data can be found in section 4.3. From these observations it seems reasonable to believe in the four main clusters stated before.

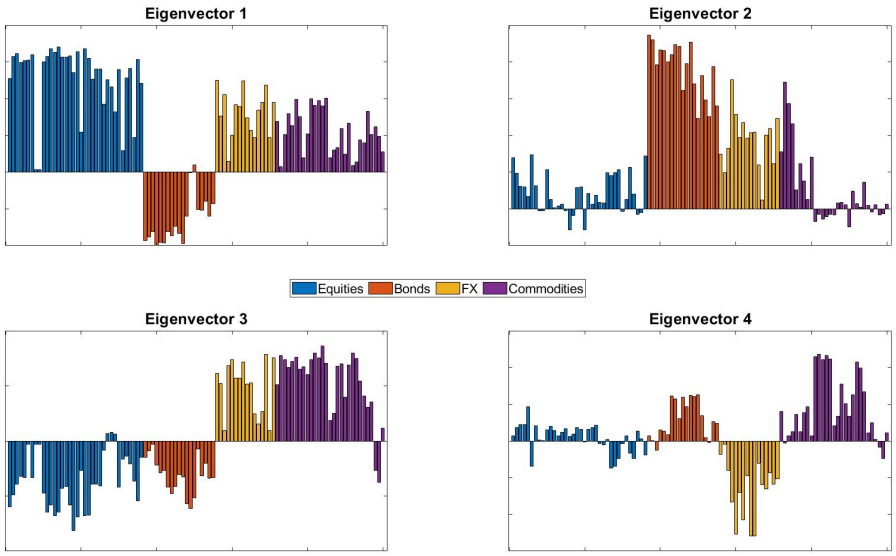


Figure 7: First four eigenvectors of the sample correlations estimated with an exponentially weighted moving average, last observation taken from 2022-03-31.

To investigate this further we examine the eigenvectors of each main cluster separately. First we want to verify the choice of our hierarchical clusters under 'Commodities' in Fig. 1. By observing the first four eigenvectors of commodities (Fig. 8) we can indeed distinguish the three child clusters - metals, energies and agriculturals. We can also see that all elements in the first eigenvector have the same sign, meaning that the assets in general tend to be positively correlated with each other. If we then look at eigenvector five (Fig. 9) we can within metals also see some internal variation, more precisely precious metals vs. industry metals. In eigenvector six (Fig. 9) we can see some internal variation within agriculturals as well, this time grains and livestock vs. miscellaneous agriculturals. The hierarchical clusters under 'Commodities' in Fig. 1 thus seem to be very reasonable.

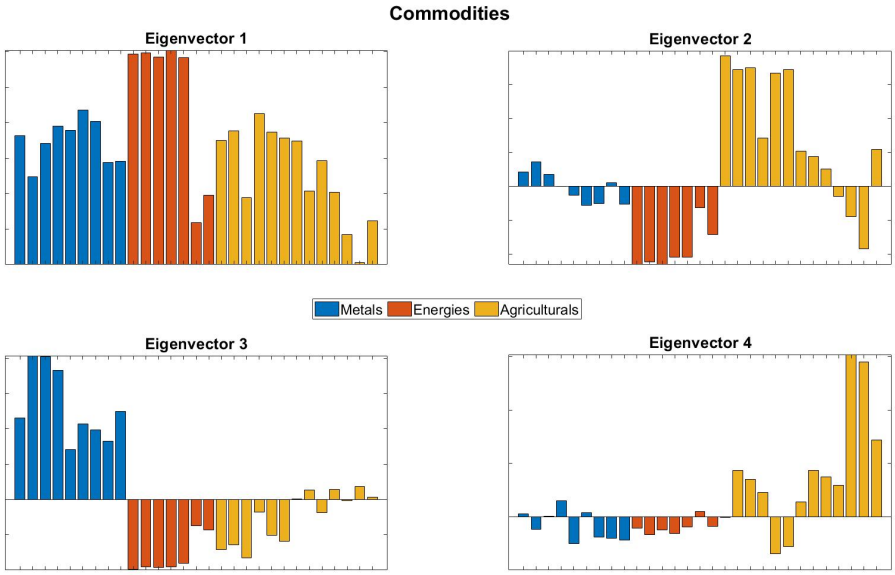


Figure 8: First four eigenvectors of the sample correlations between assets classified as 'Commodities'. Estimated with an exponentially weighted moving average, last observation taken from 2022-03-31. The height of each bar corresponds to the value of each entry in the eigenvector.

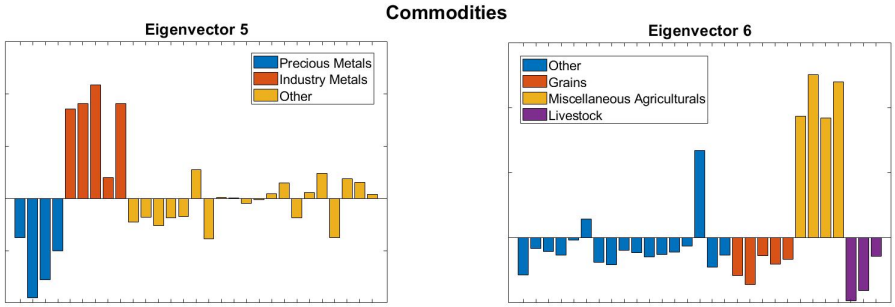


Figure 9: Eigenvector five and six of the sample correlations between assets classified as 'Commodities'. Estimated with an exponentially weighted moving average, last observation taken from 2022-03-31.

3.2 Analysis From a Regional Perspective

By examining the first three eigenvectors of bonds, equities and FX separately we can verify that there indeed seem to be some regional covariation, see Fig. 10.

For instance, in the second eigenvectors we can see that for equities, North + Latin America and Europe tend to vary with each other but against Asia, and for bonds and FX we instead have North + Latin America and Asia varying together but against Europe. In the third eigenvector of equities and bonds we can see Europe and Asia varying together against North + Latin America. The third eigenvector of FX is harder to interpret, maybe because the regional component is not as strong as for bonds and equities. This may be explained by the different risk levels of currencies within the same region. Currencies associated with less stable economies are usually considered more risky, for example Mexican peso or Brazilian real. Another important note is that all elements in the first eigenvector of all three clusters have the same sign, implying that the assets tend to be positively correlated with each other.

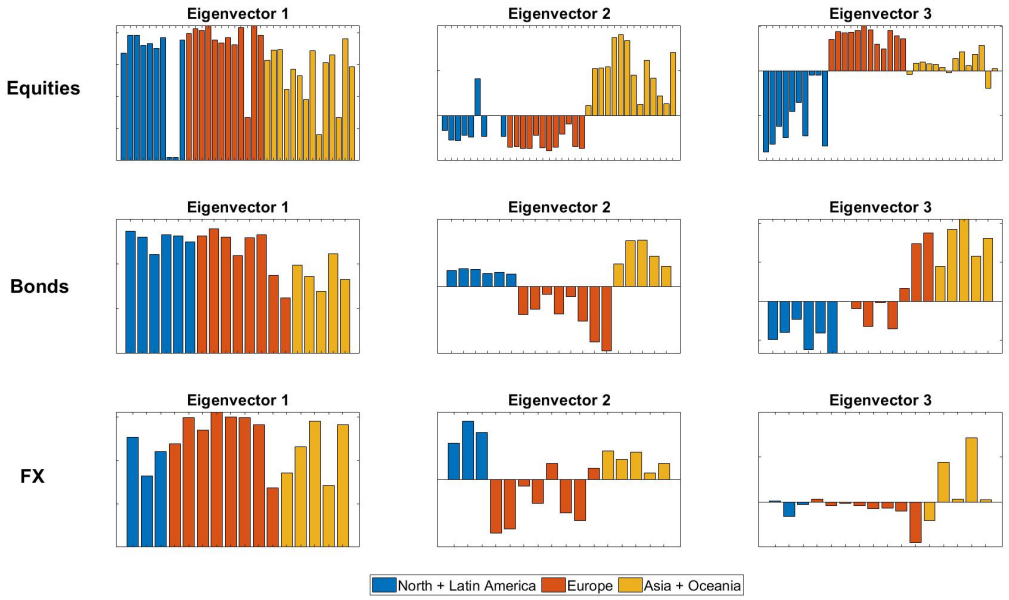


Figure 10: First three eigenvectors of the sample correlations between assets classified as 'Equities', 'Bonds' and 'FX', respectively. Estimated with an exponentially weighted moving average, last observation taken from 2022-03-31.

However, for the non-hierarchical model in Fig. 2 where the regional factor is assumed to be common for both equities, bonds and FX, we need to examine the eigenvectors of all those assets together, see Fig. 11. We directly see the characteristic market behaviour we encountered earlier in the eigenvectors of the full correlation matrix. Namely, one eigenvector showing that bonds are negatively correlated to most other assets and one eigenvector with essentially all assets in the same direction.

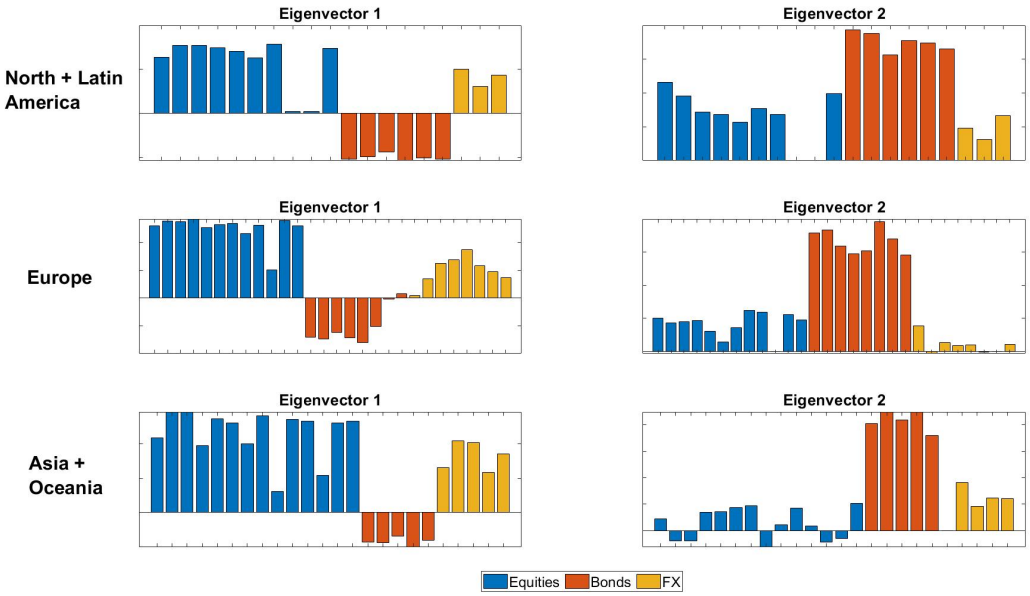


Figure 11: First two eigenvectors of the sample correlations within each region. Estimated with an exponentially weighted moving average, last observation taken from 2022-03-31.

The question to consider now is thus whether the first principal component of the non-hierarchical regional clusters is suitable to model inter-cluster correlations. To examine this we make a heatmap over the inter-cluster correlations, estimated from the first principal component of each cluster, see Fig. 12. As has been noted earlier, there will be some unintuitive correlations between asset classes and regions. If we examine these correlations carefully we detect some very interesting values. Correlations between equities and all regions are strongly negative, correlations between bonds and all regions are considerably positive and correlations between FX + commodities and all regions are considerably negative. This is probably not what one would expect the cross-correlations to look like.

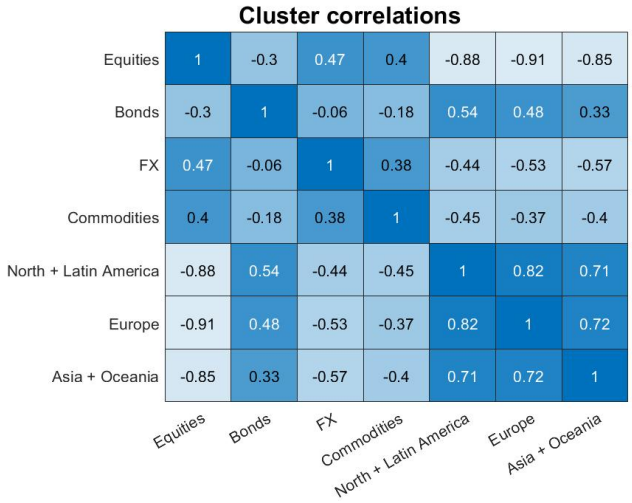


Figure 12: Cluster correlations. Estimated with an exponentially weighted moving average, last observation taken from 2022-03-31.

To examine whether this has any negative effect or not, we compare the raw correlation matrix to a modified version according to the non-hierarchical model in section 2.1.2, see Fig. 13. The reconstruction seems to work fine for all correlations except for equities vs. bonds where it is negatively inflated from a general correlation of about -0.3 to almost -0.6. This is not how the reconstruction should work and it is most likely caused by the somewhat odd behaviour of the regional principal components, so we make another attempt. This time with both the first regional principal component and the second one (making it a 3-factor model), see Fig. 14. This time the reconstruction seem to be much better and we therefore choose to proceed with this 3-factor model instead of the earlier proposed 2-factor model described in section 2.1.2.

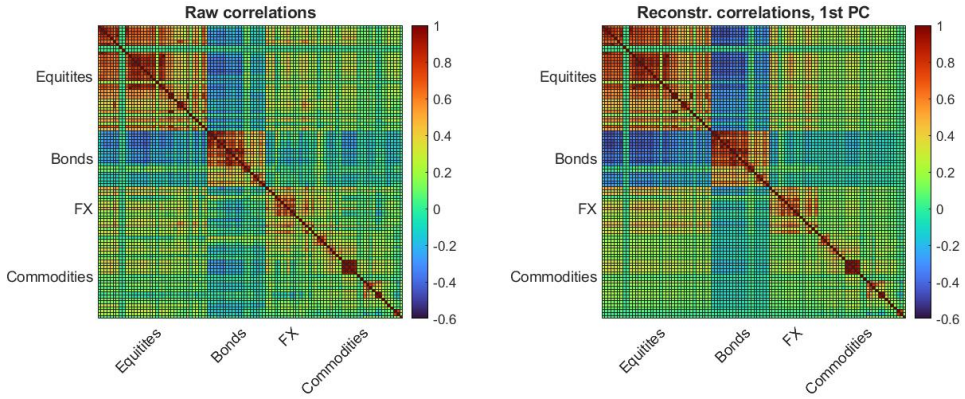


Figure 13: Raw correlations and reconstructed correlations according to section 2.1.2. Estimated with an exponentially weighted moving average, last observation taken from 2022-03-31.

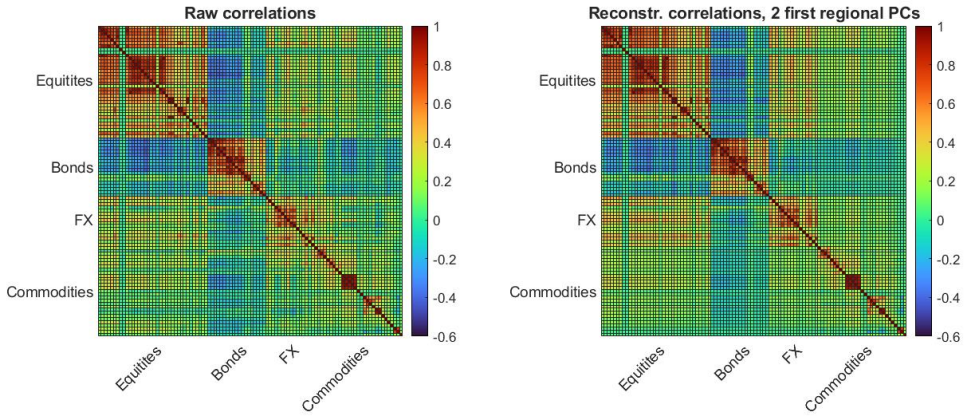


Figure 14: Raw correlations and reconstructed correlations according to section 2.1.2 but with the two first regional principal components. Estimated with an exponentially weighted moving average, last observation taken from 2022-03-31.

4 Models, Metrics and Performance Measures

The aforementioned methods to modify the correlation matrix will be evaluated with respect to sparseness/interpretability of eigenvectors, ability of forecasting risk and portfolio performance. As benchmarks, the unmodified (raw) correlation matrix estimated with an exponentially weighted moving average (EWMA), as well as with a simple shrinkage will be used.

4.1 Models

An explanation of how raw correlations are estimated, a brief summary of the different models that will be assessed and how shrinkage is applied can be found below.

Raw Correlations
So called raw or unmodified covariance is estimated with an exponentially weighted moving average, according to

$$C_t = \gamma \cdot X_t^T X_t + (1 - \gamma) \cdot C_{t-1},$$

with a decay of $\gamma = 1/500$ implying a center of mass at 500 days. X_t is a vector containing the (standardized) returns of each asset at time t. The correlation matrix Σ_t is then obtained by normalizing each element in the covariance matrix C_t with the standard deviation of the two relevant assets. This is the basic correlation estimator all modification methods are applied upon.

One-Dimensional Partitioning by Asset Class
This model is described in section 2.1.1 with a non-hierarchical partitioning by asset class, according to left/red part of Fig. 2.

Two-Dimensional Partitioning by Asset Class and Region

This model is described in section 2.1.2 with a non-hierarchical partitioning by asset class and region, according to Fig. 2. Note however that two regional factors are used, which is justified by the earlier analysis of the regional eigenvectors in section 3.2.

Hierarchical One-Factor Model
This model is described in section 2.2.1 with a hierarchical partitioning according to Fig. 1.

Hierarchical 'All'-Factor Model
This model is described in section 2.2.2 with a hierarchical partitioning according to Fig. 1.

Sparse Eigenvector Approach
This model is described in section 2.3 with a hierarchical partitioning according to Fig. 1. Note that this is a low-rank approximation of the correlation matrix.

Sparse Eigenvector Approach with L1-Penalty
This model is described in section 2.3.2 with a hierarchical partitioning according to Fig. 1. From a qualitative analysis of the eigenvectors for different values on the L1-penalty coefficient λ, it was found that $\lambda = 0.2$ yielded sufficiently sparse eigenvectors with only significant cluster-eigenvectors. A comparison of the eigenvectors for $\lambda = 0$ and $\lambda = 0.2$ can be found in Appendix C. For clarification, with $\lambda = 0$ this is the same model as the 'Sparse eigenvector approach' above and for $\lambda = 0.2$ also a low-rank approximation.

Shrinkage - a Simple but Effective Benchmark Method
The main idea behind shrinkage is to incorporate structure by a combination of the raw correlation matrix and some very structured estimate. More about shrinkage can be found in 'Honey, I Shrunk the Sample Covariance Matrix' (Ledoit & Wolf, 2004) [5]. However, this structured estimate is often very simple, for example no correlation at all or the mean correlation between all assets, as in Ledoit & Wolf (2004) [5]. In other words, shrinkage can be achieved by pulling all off-diagonal elements towards some value, for example 0 or the mean correlation. Here, shrinkage towards 0 with a shrinkage factor of $\alpha = 0.2$ will be used, ie.

$$\Sigma_{shr} = \alpha \cdot I + (1 - \alpha) \cdot \Sigma,$$

where I is the identity matrix and $0 < \alpha < 1$. The shrinkage factor $\alpha = 0.2$ is chosen with respect to a trade-off between portfolio performance and eigenvalue risk-prediction performance, which will be further explained in section 4.2. Nevertheless, this method is very simple but yet so effective. Even a small modification like this, has huge impact on the well-behaviour of the matrix. Take the eigenvalues for example. From the characteristic polynomial it can be found that each eigenvalue will be transformed according to

$$\tilde{D}^{(k)}_{shr} = \alpha + (1 - \alpha) \cdot D^{(k)}_{shr}, \tag{7}$$

where $\tilde{D}^{(k)}_{shr}$ and $D^{(k)}_{shr}$ denotes eigenvalue k after and before transformation, respectively. Eq. 7 can also be rewritten as

$$\tilde{D}^{(k)}_{shr} = \alpha + (1 - \alpha) \cdot D^{(k)}_{shr} = \alpha(1 - D^{(k)}_{shr}) + D^{(k)}_{shr}.$$

This implies that

$$\tilde{D}_{shr}^{(k)} > D_{shr}^{(k)} \text{ if } D_{shr}^{(k)} < 1 \text{ and}$$
$$\tilde{D}_{shr}^{(k)} < D_{shr}^{(k)} \text{ if } D_{shr}^{(k)} > 1.$$

Shrinkage thus increases eigenvalues smaller than 1 and decreases eigenvalues larger than one. This in turn leads to a dramatic improvement of the condition number (smaller numerator and larger denominator). In practice it reduces the extreme trading positions that often occur due to noise in the smaller eigenvalues of the raw correlation matrix.

4.2 Performance Measures

As stated earlier, the quality of the modified correlation matrices will be assessed in terms of portfolio performance, interpretability/sparseness of the eigenvectors and ability to forecast risk.

4.2.1 Portfolio Allocation Methods

To evaluate performance of the modified correlation matrices, some different portfolio optimization methods that rely with various extent on the correlation matrix will be implemented. More specific, minimum variance, mean-variance and equal risk contribution (ERC). Inverse volatility/Equal weights will be used as benchmark portfolios.

Mean-Variance
The idea behind Mean-Variance is to maximize expected returns while minimizing the variance (Markowitz, 1952) [6]. This can be formulated as

$$w_t = \arg\max_{w_t} \left(\mu_t^T w_t - \frac{1}{2} w_t^T \Sigma_{t-1} w_t \right),$$

where μ is a vector of expected returns of the assets. One solution to this problem is

$$w_t = \Sigma_{t-1}^{-1} \mu_t,$$

which can be scaled by a constant c to obtain all solutions, ie. $w_t = c \cdot \Sigma_{t-1}^{-1} \mu_t$. We will chose c such that the portfolio volatility is equal to 1, see Eq. 8 in the end of this section.

Expected returns can be tricky to estimate accurately and various fancy methods exist, but to keep it simple we will use an exponentially weighted moving average. Expected returns will thus be estimated according to

$$\mu_t = \gamma \cdot X_t + (1 - \gamma) \cdot \mu_{t-1},$$

where X_t is the returns of each asset at time t and γ is the decay. Here a decay of $\gamma = 1/300$, ie. a center of mass at 300 days, will be used.

Minimum Variance
Minimum variance is - except for inverse volatility/equal weights - the simplest allocation method we will use here as it only requires an estimation of the covariance/correlation matrix. One could therefore expect that a better estimated correlation matrix would result in a better performing

portfolio. The main idea behind minimum variance is exactly what it sounds like - to minimize the variance. This can be formulated as

$$\begin{aligned} w_t = \ & \arg\ \min_{w_t} \frac{1}{2} \cdot w_t^T \Sigma_{t-1} w_t \\ \text{s.t} \ \ & w_t^T \mu = 1, \ \text{where} \ \ \mu = [1\ 1\ \ldots\ 1]^T. \end{aligned}$$

The solution to this is unique, according to

$$w_t = \Sigma_{t-1}^{-1} \mu.$$

We will however scale the solution such that the portfolio volatility is equal to 1, see Eq. 8 in the end of this section. The only difference compared to the weights obtained from 'Mean-Variance' is thus that μ is a constant vector of ones.

The intuition behind the model is to minimize risk by utilizing how the assets are correlated with each other. No effort is put into adding an estimation of expected returns, correlation is enough according to this approach.

Equal Risk Contribution

The idea behind ERC is to derive optimal portfolio weights based on equal risk contribution from each asset. In the more general case, where the risk contributions are not necessarily equal but described by a vector $b = [b_1\ b_2\ \ldots\ b_n]$, the problem can be formulated as the following system of equations

$$\begin{cases} RC_i(w) = b_i R(w) \\ b_i > 0 \\ w_i > 0 \\ \sum_{i=1}^n b_i = 1 \\ \sum_{i=1}^n w_i = 1 \end{cases}$$

where $RC_i(w)$ is the risk contribution of asset i and $R(w)$ is the risk measure of the portfolio with portfolio weights $w = [w_1\ w_2\ \ldots\ w_n]$ (Griveau-Billion et al., 2013) [3]. The risk measure $R(w)$ will be the portfolio volatility. This system can efficiently be solved by the cyclical coordinate (CCD) descent algorithm, as described in 'A Fast Algorithm for Computing High-dimensional Risk Parity Portfolios' (Griveau-Billion et al., 2013) [3]. As the title reveals, we will use equal risk contribution from each asset, ie. $b_1 = b_2 = \ldots = b_n$ where n is the number of assets. It should be noted here that the weights are constrained to be positive, unlike the weights estimated with 'Minimum Variance' and 'Mean-Variance'.

Inverse Volatility/Equal Weights

Inverse volatility is a relatively simple allocation method, assigning weights proportional the inverse of the volatility of each asset. In practice this means

$$w_t = \frac{1}{\sqrt{diag(\Sigma_{t-1})}}.$$

However, in this setting where we use the correlation matrix instead of the covariance matrix all assets will have volatility equal to 1. This implies that all assets will be assigned the same weight,

which is yet another simple allocation method often referred to as 'Equal weights'. More about the advantages with this naive equal weights allocation method can be found in 'Optimal versus naive diversification: How inefficient is the 1/N portfolio strategy?' (DeMiguel et al., 2007) [2].

To make the different portfolio allocation methods comparable we normalize all weights by expected portfolio volatility. Expected portfolio volatility is simply estimated as the realised volatility at $t-1$ but with the newly estimated weights, resulting in the normalized weights $\tilde{w}_t$ according to

$$\tilde{w}_t = \frac{w_t}{\sqrt{w_t^T \Sigma_{t-1} w_t}}. \tag{8}$$

In this way we can set our preferred risk level, here a portfolio volatility equal to 1. It should also be mentioned that weights are updated daily. Also, when these allocation methods are applied on low-rank approximations of the correlation matrix (the sparse approaches), a pseudo inverse of the matrices is used when necessary.

To evaluate portfolio performance - average portfolio return (daily), realised portfolio volatility (daily), Sharpe ratio (yearly) and average turnover/holding period will be used. The performance measures will be estimated according to the following:

Average Portfolio Return:	$E[dR]$
Realised Portfolio Volatility:	$\sqrt{V[dR]}$
Sharpe Ratio:	$\frac{\text{Average Portfolio Return}}{\text{Realised Portfolio Volatility}} \cdot \sqrt{252}$
Average Turnover:	$\frac{1}{E[\lvert w \rvert]} \cdot \sum_{k=1}^{M} \sum_{t=1}^{N} \frac{\lvert w_{t+1}^k - w_t^k \rvert}{M \cdot N}$
Average Holding Period	$\frac{1}{\text{Average Turnover}}$

Here dR denotes daily portfolio return, w_t^k denotes the weight of asset k at time t and $E[|w|]$ is the average absolute weight over time, amongst all assets. M and N are the number of assets and daily portfolio returns respectively.

4.2.2 Eigenvalues and Their Ability to Predict Risk

Portfolio weights derived according to Mean-Variance can be decomposed into eigenvalues and eigenvectors, which in turn can be interpreted as a sum of scaled eigenportfolios. The column

vector of portfolio weights w can be expressed as a sum of M scaled eigenportfolios according to

$$\begin{aligned} w &= \Sigma^{-1}\mu \\ &= VD^{-1}V^T\mu \\ &= \sum_{k=1}^{M} V^{(k)} \frac{1}{D^{(k)}} V^{(k)T}\mu \\ &= \sum_{k=1}^{M} \frac{V^{(k)}}{\sqrt{D^{(k)}}} \frac{V^{(k)T}\mu}{\sqrt{D^{(k)}}} \\ &= \sum_{k=1}^{M} \frac{V^{(k)}}{\sqrt{D^{(k)}}} \frac{\mu^{(k)}}{\sqrt{D^{(k)}}}. \end{aligned}$$

Each eigenvector $V^{(k)}$ can thus be interpreted as a set of eigenportfolio weights (the eigenportfolio), scaled by its volatility $\frac{1}{\sqrt{D^{(k)}}}$ to obtain unit variance, and Sharpe ratio (expected return/volatility). The relationships between eigenportfolio risk (variance), expected eigenportfolio return $\mu^{(k)}$ and eigenvalues $D^{(k)}$ can be shown by

$$\begin{aligned} Var[V^{(k)T}X] &= V^{(k)T}Var[X]V^{(k)} \\ &= V^{(k)T}\Sigma V^{(k)} \\ &= D^{(k)}, \\ E[V^{(k)T}X] &= V^{(K)}E[X] \\ &= V^{(K)}\mu \\ &= \mu^{(k)}, \end{aligned}$$

where X now is a *random vector* describing the returns of each asset. We can therefore evaluate the performance of a correlation matrix by comparing its eigenvalues to the realised risk of each eigenportfolio. Realised eigenportfolio risk will be measured as the variance of the return of the portfolios constructed from each eigenvector during some different time periods, depending on the horizon of interest. These variances will then be compared to the eigenvalues corresponding to the eigenvectors used to construct the portfolios, to evaluate the predictive performance of the eigenvalues.

To investigate whether the risk is more often under- or overestimated during the time period the mean difference of the ratio [eigenvalue / eigenportfolio variance] from 1 will be examined for each eigenvalue. Benchmark eigenvalues and eigenvectors will be extracted from the raw correlation matrix and from the shrunken raw correlation matrix.

A typical problem with the raw correlation matrix is that its eigenvalues usually gives an overestimation of the risk associated with the first eigenportfolios as well as an underestimation of the risk related to the eigenportfolios associated with smaller eigenvalues. With shrinkage on the other hand, the opposite is often the case. As the value of the smallest eigenvalues is increased, the risk related to the eigenportfolios of those is typically overestimated. Ideally we want a non-biased small deviation from the realised risk, and rather overestimated than underestimated.

Regarding eigenvalues, the condition number is also an interesting measure of performance. The condition number of a matrix is calculated as the largest eigenvalue divided by the smallest

eigenvalue and roughly measures how prone to error the inverse of a matrix is. This is a nice indicator, but keep in mind that it does not necessarily mean that the actual portfolio performance will be improved.

4.2.3 Sparseness and Interpretability of Eigenvectors

Furthermore, sparseness of the eigenvectors can be used as a measure of how denoised a correlation matrix is. Sparseness is not only a nice indication of low levels of noise but also facilitates interpretation. Both the number of non-zero elements and some qualitative analysis of the eigenvectors are interesting here.

4.3 Data and Pre-Processing

The data used in this paper consists of price series from 100 different futures contracts taken from 1980-01-01 to 2022-03-31. Not all assets have data for the full time period and are therefore gradually added to the trading universe. A complete list of all utilized assets can be found in Appendix A. All prices are written in - or converted to - US dollars. Since each futures contract only exist for a limited amount of time (overlapping with its successor and predecessor), all price series on the same contract are merged to get one single coherent price series. This means that the actual price is not preserved, but the arithmetic difference is.

To somehow standardize things, closing prices are used to estimate arithmetic daily returns which in turn are volatility adjusted to obtain standardized daily returns. Volatility adjustment is achieved by scaling each return with an estimate of its current volatility, in this case estimated from both open, max, min and closing prices. To compensate for different opening hours and time zones, a 3-day smoothing filter is applied to the standardized returns. The mean values of the returns are not subtracted as it is assumed to be approximately zero, any deviation just noise.

5 Results

To understand how much impact each method have had on the correlation matrix itself, we provide a plot with the mean absolute difference per element compared to the raw correlation matrix (Fig. 15). The two sparse methods obviously have the largest impact, while 'Hierarchical 'all'-factor' is most similar to the raw correlation matrix.

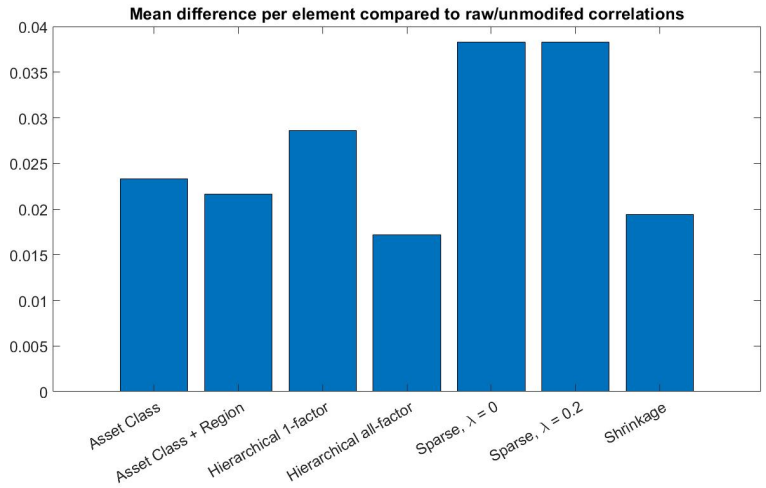


Figure 15: Mean absolute difference per element in the correlation matrix over the full time period 1980-01-01 to 2022-03-31.

5.1 Eigenvalues and Realised Eigenportfolio Risk

To get a feeling for what effect the different methods have on the eigenvalues, we make a plot over the mean eigenvalues see Fig. 16. It is very clear that shrinkage has a large impact on especially the smallest eigenvalues and that all other methods except for the two sparse ones yield eigenvalues that are very similar to the raw correlation matrix´s. The eigenvalues of the sparse method with $\lambda = 0$ seem to be a somewhat compressed version (along the x-axis) of the raw ones. The eigenvalues of the sparser method with $\lambda = 0.2$, discards many of the smallest eigenvalues keeping only ≈ 70 of the largest ones.

Closely related to the eigenvalues is the condition number, see Fig. 17. It is very clear that the size of the smallest eigenvalue drastically improves the condition number.

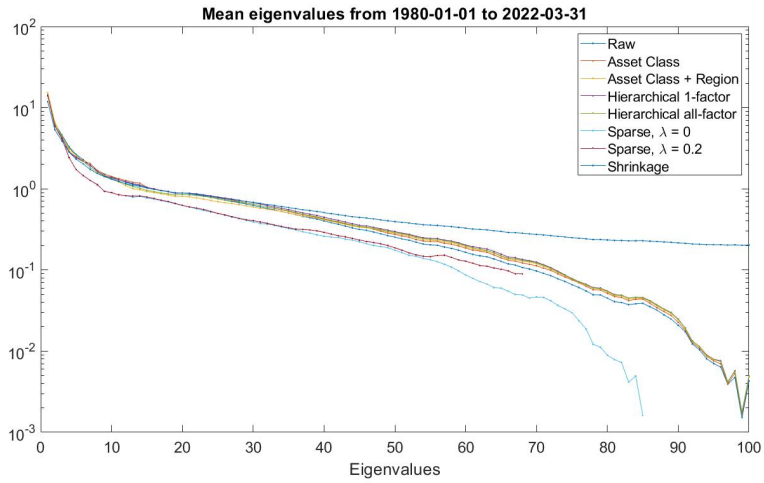


Figure 16: Mean eigenvalues.

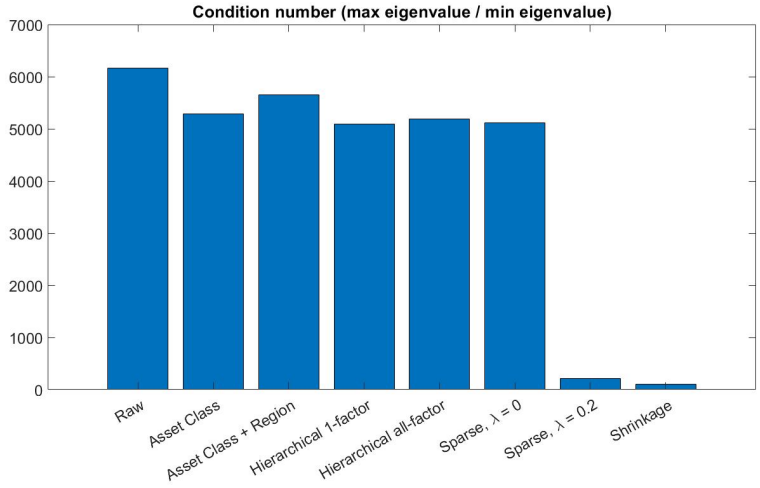


Figure 17: Condition number of the last correlation matrix, ie. 2022-03-31.

The 20-days predictive ability of the eigenvalues can be seen in Fig. 18 below. The earlier described 'bias' in the raw and shrunken eigenvalues are very apparent here. The raw eigenvalues gives a larger and larger underestimation of the risk associated with the smaller eigenportfolios while the shrunken gives a larger and larger overestimation of the risk associated with the smaller eigenportfolios. However, all other modification methods yield very accurate risk predictions, neither clearly underestimated nor clearly overestimated. In Fig. 19 we can see that the mean error behaves very similar regardless of the time horizon.

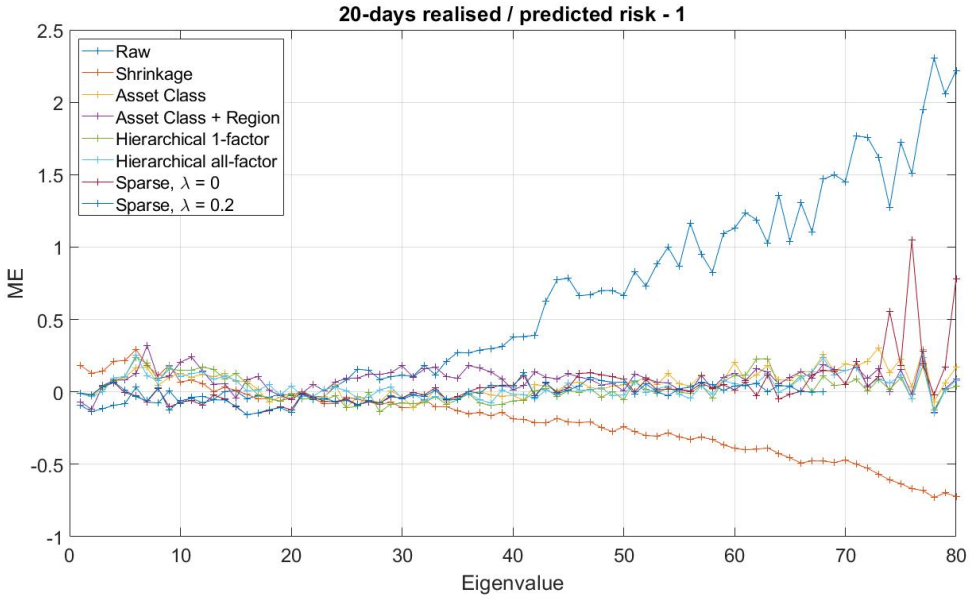


Figure 18: Mean error, realised risk vs predicted risk per eigenvalue/eigenportfolio.

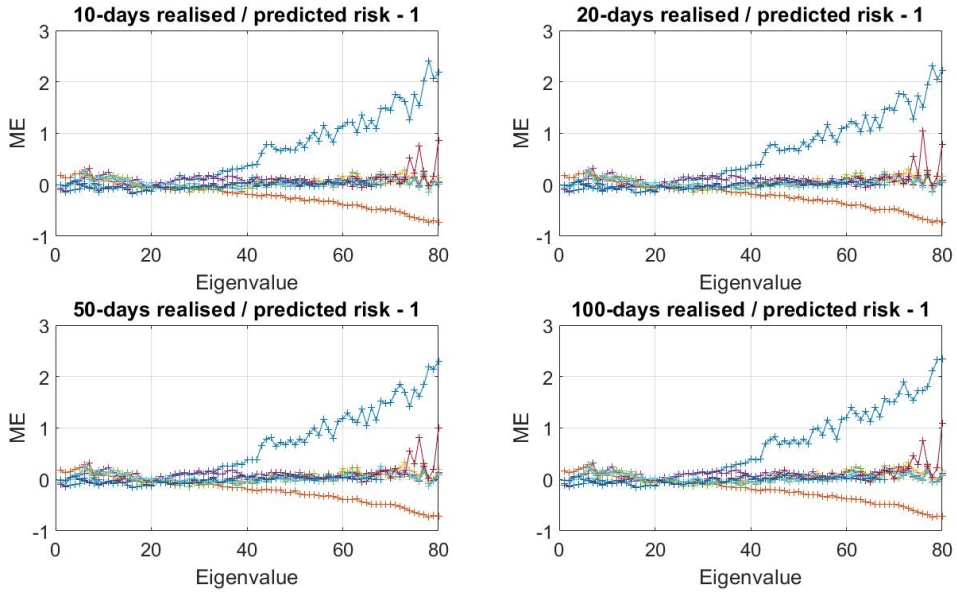


Figure 19: Mean error, realised risk vs predicted risk per eigenvalue/eigenportfolio, different time horizons.

5.2 Sparseness and Interpretability of Eigenvectors

A quick overview of the sparseness can be obtained in Fig. 20 below. We can see that the raw correlation matrix has essentially no zero elements in the eigenvectors, as well as the hierarchical 1-factor and 'all'-factor model. Slightly sparser is the eigenvectors from the non-hierarchical Asset Class + Region model, but sparsest are clearly the Asset Class and the two sparse approach models. This is not very surprising though as the purpose of the sparse approach is to produce sparse eigenvectors.

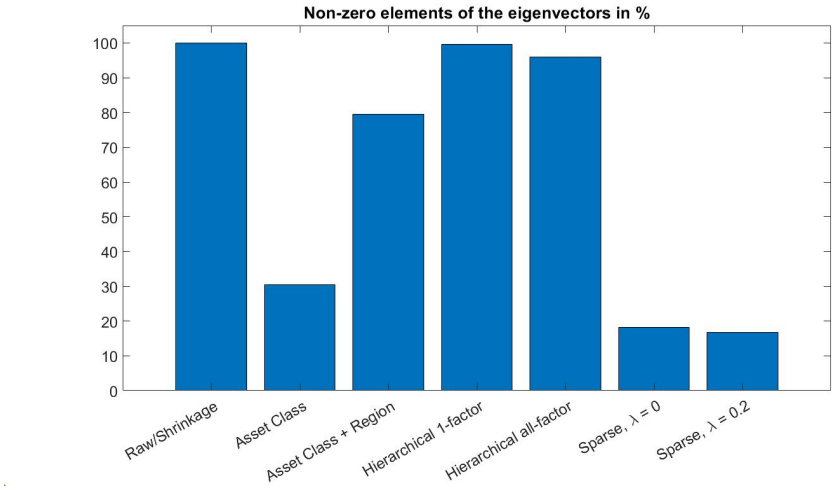


Figure 20: Percentage of non-zero elements in the last eigenvectors, ie. 2022-03-31.

5.3 Portfolio Performance

Portfolio performance for the three different allocation methods 'Minimum Variance', 'Mean-Variance' and 'Equal Risk Contribution' can be found in table 1, 2 and 3, respectively. Note that the performance of 'Inverse Volatility/Equal Weights' is the same in all tables and added to serve as a benchmark.

	Sharpe Ratio (yearly)	Portfolio Volatility (daily)	Avg. Portfolio Return (daily)	Avg. Holding Period (days)
Raw	0.769	1.11	0.054	38.6
Asset Class	0.802	1.11	0.056	103
Asset Class & Region	0.857	1.19	0.064	89.4
Hierarchical 1-factor	0.818	1.14	0.059	166
Hierarchical 'all'-factor	0.840	1.09	0.058	116
Sparse, $\lambda = 0$	0.939	1.09	0.065	88.1
Sparse, $\lambda = 0.2$	0.934	1.09	0.064	55.3
Shrinkage	0.818	1.14	0.059	83.8
Inverse vol./ Equal weights	0.683	1.01	0.043	1070

Table 1: Performance measures for the allocation method 'Minimum Variance' for all different modified correlation matrices. Inverse volatility/Equal weights included as benchmark.

	Sharpe Ratio (yearly)	Portfolio Volatility (daily)	Avg. Portfolio Return (daily)	Avg. Holding Period (days)
Raw	0.497	1.12	0.035	8.77
Asset Class	0.763	1.25	0.060	10.7
Asset Class & Region	0.838	1.29	0.068	11.2
Hierarchical 1-factor	0.815	1.30	0.067	11.6
Hierarchical 'all'-factor	0.779	1.25	0.061	11.3
Sparse, $\lambda = 0$	0.853	1.22	0.066	11.3
Sparse, $\lambda = 0.2$	0.888	1.20	0.067	12.2
Shrinkage	0.908	1.11	0.064	11.4
Inverse vol./ Equal weights	0.683	1.01	0.043	1070

Table 2: Performance measures for the allocation method 'Mean Variance' for all different modified correlation matrices. Inverse volatility/Equal weights included as benchmark.

	Sharpe Ratio (yearly)	Portfolio Volatility (daily)	Avg. Portfolio Return (daily)	Avg. Holding Period (days)
Raw	0.783	1.03	0.051	318
Asset Class	0.768	1.06	0.051	462
Asset Class & Region	0.772	1.03	0.050	329
Hierarchical 1-factor	0.759	1.07	0.051	500
Hierarchical 'all'-factor	0.791	1.03	0.052	370
Sparse, $\lambda = 0$	-	-	-	-
Sparse, $\lambda = 0.2$	-	-	-	-
Shrinkage	0.784	1.12	0.055	353
Inverse vol./ Equal weights	0.683	1.01	0.043	1070

Table 3: Performance measures for the allocation method 'Equal Risk Contribution' for all different modified correlation matrices. No values for the sparse methods since the optimization algorithm did not converge. Inverse volatility/Equal weights included as benchmark.

Within the 'Minimum Variance' portfolio performance in table 1, we can see that the Sharpe ratio is slightly improved for all modification methods compared to when using the raw correlation matrix. Most significant are the sparse approaches. Compared to shrinkage, the other modified matrices seem to perform about at least as well with respect to Sharpe ratio. Regarding holding period, all methods except for the sparsest ($\lambda = 0.2$) are longer than for the raw and shrunken matrix.

When examining the performance of 'Mean-Variance' in table 2, we can see a huge improvement of the Sharpe ratio for all modified matrices compared to the raw correlation matrix. Shrinkage only performs slightly better than the modified matrices. The holding period is very similar for all models.

Regarding 'Equal Risk Contribution', we note that the values for the sparse methods are missing. This is because the optimization did not converge, causing the portfolio weights to explode. Some debugging showed that this is most likely due to the fact that the sparse matrices not strictly are correlation matrices (not necessarily ones on the main diagonal) and also of very low rank, which caused problem for the convergence. However, seen to the other methods it is obvious that this allocation method is less sensitive to the correlation matrix. Essentially all methods yield the same performance regardless of if or how the correlation matrix is modified. It is also notable that the holding period is very long for all models.

An interesting note here is that we normalize all weights to obtain a daily portfolio volatility equal to 1, but still get a slightly higher volatility for all models and allocation methods.

6 Conclusions & Discussion

Regarding the ability of the eigenvalues to predict risk, we can conclude that all proposed HPCA methods yields a much more accurate estimation of the risk associated with each eigenportfolio compared to the raw and shrunken correlation matrices. This property is very favourable and most

likely contributing to the improvement of the Sharpe ratio when using the 'Minimum Variance' allocation method. This is not really the case for 'Mean-Variance' where the shrunken matrix still performs very well. This is probably caused by the estimation of expected returns, which tend to favour shrunken matrices. Although the sparse method with $\lambda = 0.2$ also performs well in 'Mean-Variance'. This can maybe be explained by the similarity of its eigenvalues and the shrunken eigenvalues.

Comparing the two non-hierarchical methods 'Asset Class' and 'Asset Class & Region' we can conclude that regarding portfolio performance, we need both an asset class specific factor and a regional factor to get a higher Sharpe ratio. If we instead look at sparseness of the eigenvectors, only an asset class-factor yield much sparser result. It could though be questionable if that is the right kind of sparseness. In the case of only asset class, the eigenvectors are built from the asset class cluster eigenvectors. This means that if there are any other structures in the correlation (for example regional clusters), this will not be visible in the eigenvectors of the full matrix since they are constrained to be a combination of the asset class specific eigenvectors. So, sparseness itself is only a good thing if we strongly believe in our clusters.

Regarding the two hierarchical methods, their performance are very similar from all aspects. The 'all'-factor model yields a higher Sharpe ratio for 'Minimum Variance' while the one-factor yields a higher Sharpe ratio for 'Mean-Variance'. It is therefore hard to conclude whether one of them is better than the other.

Comparing all HPCA methods altogether, the sparse ones seem to be best performing as they yield both sparse and interpretable eigenvectors, accurately estimated eigenportfolio risk and decent portfolio performance. There is however a convergence problem when utilizing these matrices in ERC, but since essentially all matrices performs the same it is meaningless to utilize something else than the raw correlation matrix in ERC.

When comparing the two sparse models their overall performance is quite similar, except for the fact that the sparsest model ($\lambda = 0.2$) has disregarded many of the smallest eigenvalues. This results in a much better condition number. It also has slightly higher Sharpe ratio and a little sparser eigenvector structure. We can thus conclude that the sparsest model slightly outperforms the less sparse model ($\lambda = 0$) and performs as good as shrinkage but with the advantage of being regularized in a much more meaningful way.

Another interesting topic is the relevancy of the clusters. In this paper it has been proposed that clusters should be modeled as constant over time, but this may not necessarily be the case. However, when partitioning the assets in a natural way - for instance asset class and region - then there is no reason to model this dynamically since regions and asset classes are very unlikely to change.[3] Time-varying clusters would thus be more relevant if one would utilize some data-driven clustering method instead. Also, one could argue that the regional clusters are becoming less and less relevant as the economy is getting more and more globalized. A couple of hundred years ago, the regional factor would probably have been very evident as people were mostly trading within their own closest area but nowadays everyone can trade with essentially anyone.

A possible problem, which we luckily have not encountered here, arises from the fact that we have not proven that the modified correlation matrices from 'Asset Class & Region' and the two hierarchical methods are positive semi-definite, implying that there is a possibility of negative eigenvalues. This could have terrible consequences for portfolio performance since we then have an eigenvector associated with *negative* risk, see for example 'Seven sins in portfolio optimization' (Hauser & Schmeltzer, 2013) [7] for further explanation.

[3]It would be weird if the location of USA suddenly was in Europe or if gold turned into a grain.

References

[1] Marco Avellaneda. "Hierarchical PCA and applications to portfolio management". In: *Revista mexicana de economıa y finanzas* 15.1 (2020), pp. 1–16.

[2] Victor DeMiguel, Lorenzo Garlappi, and Raman Uppal. "Optimal versus naive diversification: How inefficient is the 1/N portfolio strategy?" In: *The review of Financial studies* 22.5 (2009), pp. 1915–1953.

[3] Théophile Griveau-Billion, Jean-Charles Richard, and Thierry Roncalli. "A fast algorithm for computing high-dimensional risk parity portfolios". In: *Available at SSRN 2325255* (2013).

[4] Anders Holst and Victor Ufnarovski. *Matrix Theory*. Studentlitteratur, 2014.

[5] Olivier Ledoit and Michael Wolf. "Honey, I shrunk the sample covariance matrix". In: *The Journal of Portfolio Management* 30.4 (2004), pp. 110–119.

[6] Harry M Markowitz. "Portfolio Selection, 1959". In: *Journal of Finance* 7 (1952), pp. 77–91.

[7] Thomas Schmelzer and Raphael Hauser. "Seven sins in portfolio optimization". In: *arXiv preprint arXiv:1310.3396* (2013).

A Appendix

Name	Asset Class	Asset Sub Class	Region
Nasdaq	Equities	-	North America
S&P 500	Equities	-	North America
Dow Jones	Equities	-	North America
Russell 2000	Equities	-	North America
Canada 60	Equities	-	North America
MEM	Equities	-	North America
S&P 400	Equities	-	North America
Russell 2000 CME	Equities	-	North America
Russell 2000 ICE	Equities	-	North America
S&P 500 ESG	Equities	-	North America
OMX	Equities	-	Europe
CAC 40	Equities	-	Europe
DAX	Equities	-	Europe
Euro STOXX 50	Equities	-	Europe
FTSE 100	Equities	-	Europe
IBEX	Equities	-	Europe
MIB	Equities	-	Europe
SMI	Equities	-	Europe
AEX	Equities	-	Europe
Euro STOXX Banks	Equities	-	Europe
STOXX 600 ESG	Equities	-	Europe
OMX ESG	Equities	-	Europe
SPI	Equities	-	Oceania
Nikkei 225 mini	Equities	-	Asia
Nikkei SIM	Equities	-	Asia
Taiwan	Equities	-	Asia
Hang Seng	Equities	-	Asia
Hang Seng China Enterprises	Equities	-	Asia
China 50	Equities	-	Asia
TOPIX	Equities	-	Asia
Nifty50	Equities	-	Asia
KOSPI	Equities	-	Asia
Singapore	Equities	-	Asia
Nikkei Mini	Equities	-	Asia
Nikkei CME	Equities	-	Asia
FTSE Taiwan	Equities	-	Asia
10-Year T-Note	Bonds	-	North America
5-Year T-Note	Bonds	-	North America
2-Year T-Note	Bonds	-	North America
U.S. Treasury Bond	Bonds	-	North America

Name	Asset Class	Asset Sub Class	Region
Ten-Year Government of Canada Bond	Bonds	-	North America
Ultra U.S. Treasury Bond	Bonds	-	North America
BOBL	Bonds	-	Europe
Euro Bund	Bonds	-	Europe
UK Gilt	Bonds	-	Europe
Euro Schatz	Bonds	-	Europe
Euro Buxl	Bonds	-	Europe
Euro OAT	Bonds	-	Europe
Euro BTP	Bonds	-	Europe
BTP Short	Bonds	-	Europe
mini-10year JGB	Bonds	-	Asia
KTB 10y	Bonds	-	Asia
KTB 3y	Bonds	-	Asia
10-Year Australian Government Bond	Bonds	-	Oceania
3-Year Australian Government Bond	Bonds	-	Oceania
Canadian Dollar	Foreign Exchange	-	North America
Brazilian Real vs. US Dollar	Foreign Exchange	-	Latin America
Mexican Peso vs. US Dollar	Foreign Exchange	-	Latin America
Swiss Franc	Foreign Exchange	-	Europe
Euro	Foreign Exchange	-	Europe
British Pound	Foreign Exchange	-	Europe
Swedish Krona vs. US Dollar	Foreign Exchange	-	Europe
Norwegian Krone vs. US Dollar	Foreign Exchange	-	Europe
Polish Zloty vs. US Dollar	Foreign Exchange	-	Europe
Hungarian Forint vs. US Dollar	Foreign Exchange	-	Europe
Turkish Lira vs. US Dollar	Foreign Exchange	-	Europe
Russian Ruble vs. US Dollar	Foreign Exchange	-	Europe
South Korean Won vs. US Dollar	Foreign Exchange	-	Asia
Australian Dollar	Foreign Exchange	-	Oceania
South Korean Won	Foreign Exchange	-	Asia
New Zealand Dollar	Foreign Exchange	-	Oceania
Palladium	Commodities	Metals - Precious Metals	-
Gold	Commodities	Metals - Precious Metals	-
Silver	Commodities	Metals - Precious Metals	-
Platinum	Commodities	Metals - Precious Metals	-
Aluminium	Commodities	Metals - Industry Metals	-
Copper	Commodities	Metals - Industry Metals	-
Zinc	Commodities	Metals - Industry Metals	-
Nickel	Commodities	Metals - Industry Metals	-
Lead	Commodities	Metals - Industry Metals	-
Brent Crude	Commodities	Energies	-
Low Sulphur Gasoil	Commodities	Energies	-
Crude Oil	Commodities	Energies	-

Name	Asset Class	Asset Sub Class	Region
NY Harbor ULSD	Commodities	Energies	-
RBOB Gasoline	Commodities	Energies	-
Henry Hub Natural Gas	Commodities	Energies	-
Crude ICE	Commodities	Energies	-
Corn	Commodities	Agriculturals - Grains	-
Soybeans	Commodities	Agriculturals - Grains	-
Soybean Meal	Commodities	Agriculturals - Grains	-
Soybean Oil	Commodities	Agriculturals - Grains	-
Wheat	Commodities	Agriculturals - Grains	-
Wheat Kansas	Commodities	Agriculturals - Grains	-
Sugar	Commodities	Agriculturals - Miscellaneous	-
Coffee	Commodities	Agriculturals - Miscellaneous	-
Cotton	Commodities	Agriculturals - Miscellaneous	-
Cocoa	Commodities	Agriculturals - Miscellaneous	-
Live Cattle	Commodities	Agriculturals - Livestock	-
Feeder Cattle	Commodities	Agriculturals - Livestock	-
Lean Hogs	Commodities	Agriculturals - Livestock	-

B Appendix

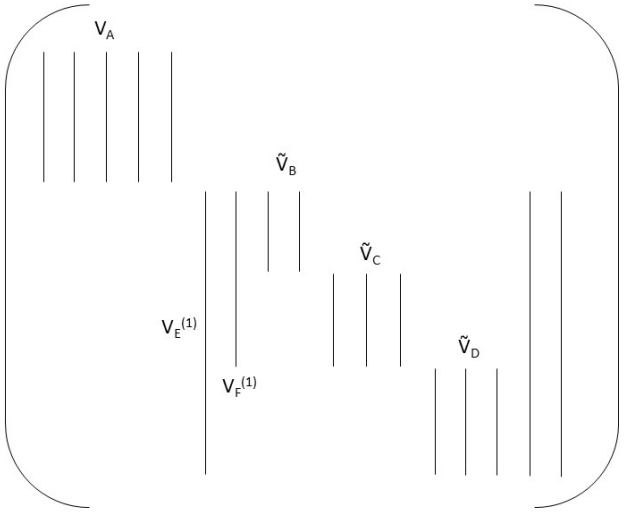
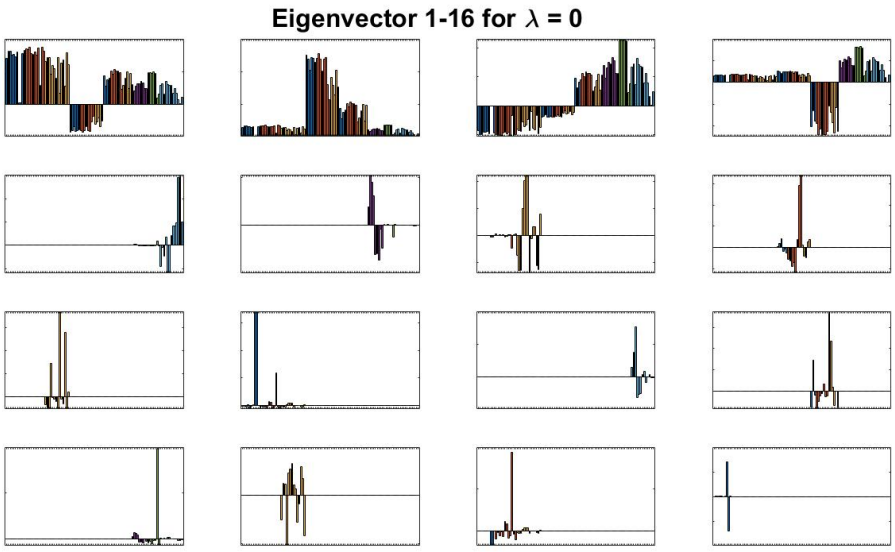
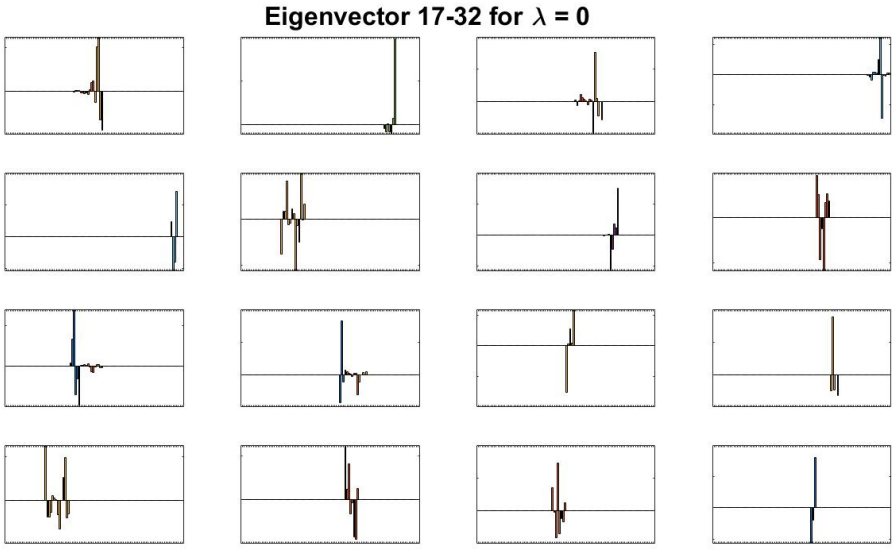


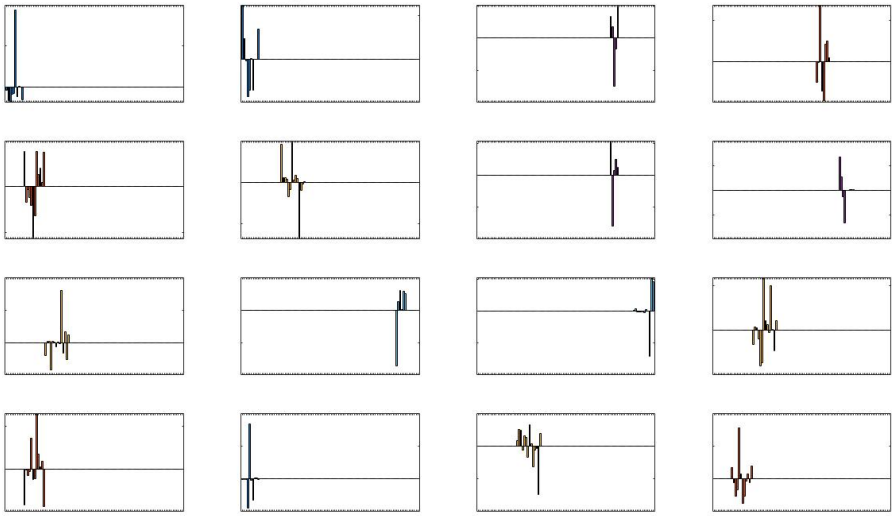
Figure 21: An example of the G matrix. In this case two main clusters A and E. Two child clusters of E - F and D. Also two child clusters of F - B and C.

C Appendix

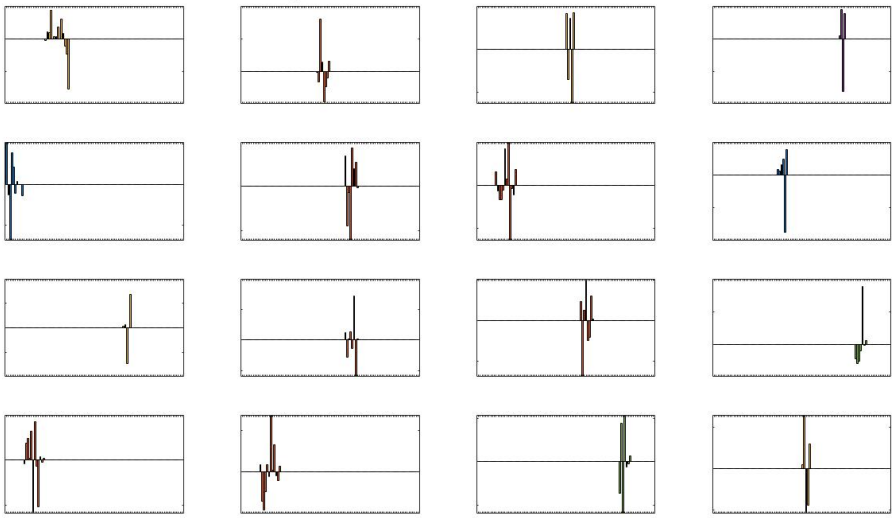




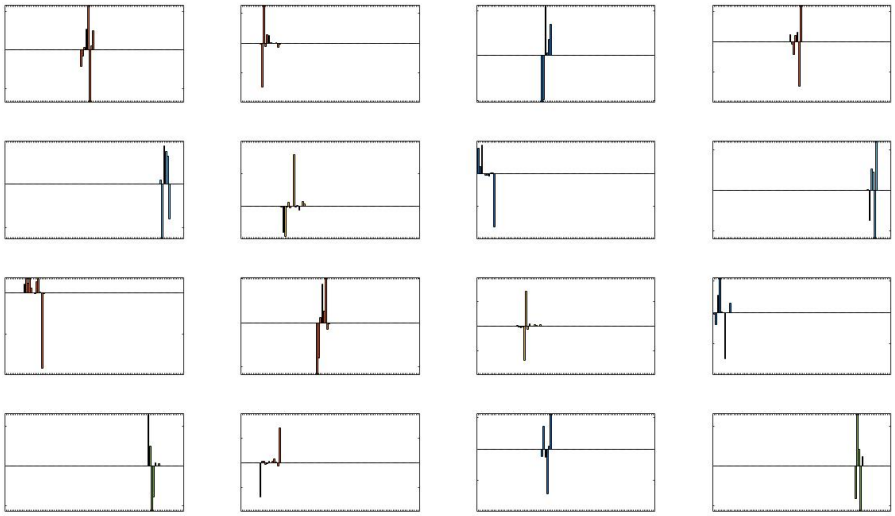
Eigenvector 33-48 for $\lambda = 0$



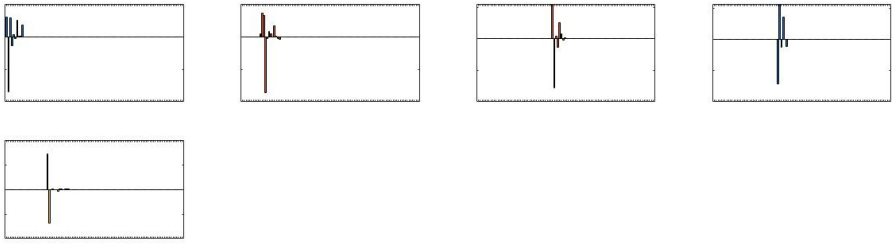
Eigenvector 49-64 for $\lambda = 0$



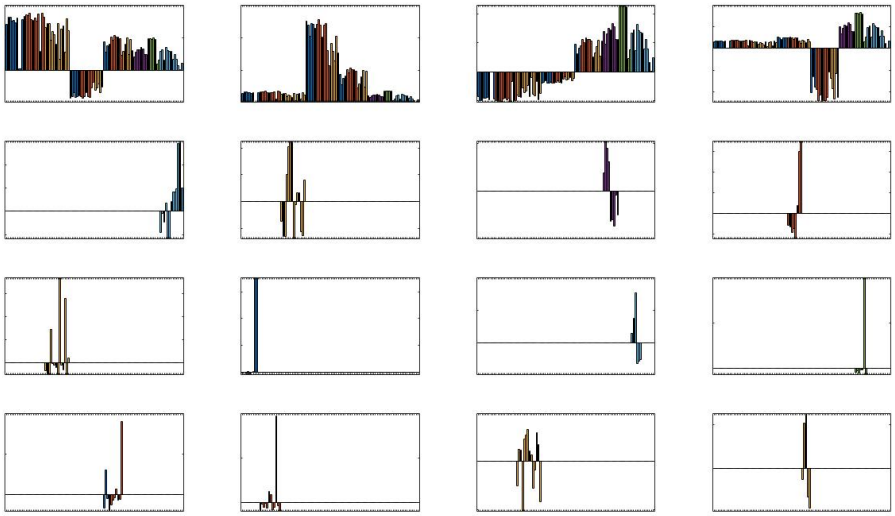
Eigenvector 65-80 for $\lambda = 0$



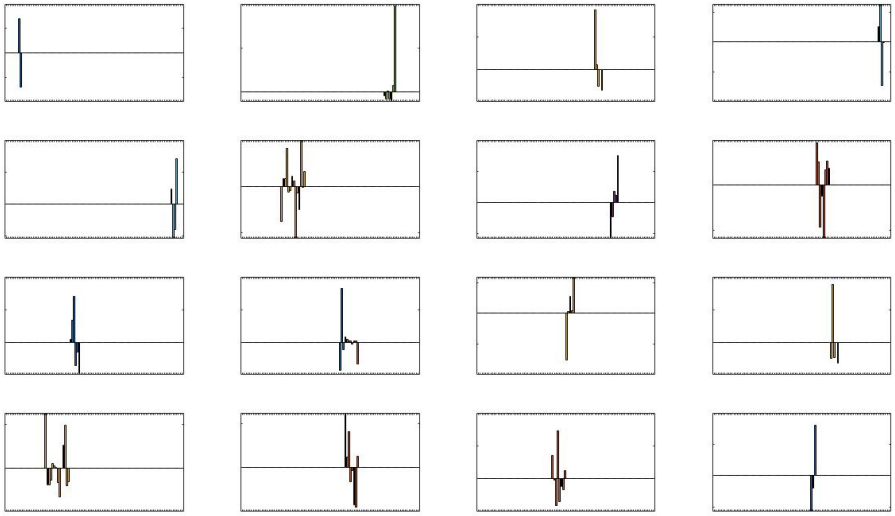
Eigenvector 81-85 for $\lambda = 0$



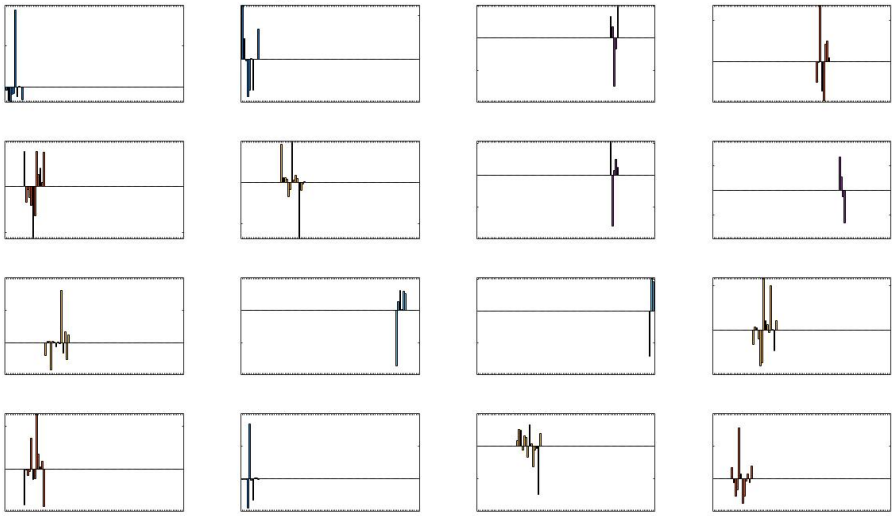
Eigenvector 1-16 for λ = 0.2



Eigenvector 17-32 for λ = 0.2



Eigenvector 33-48 for λ = 0.2



Eigenvector 49-64 for λ = 0.2