

Detection and Analysis of Correlation Clusters and Market Risk Concentration

Factor analysis is successfully used to analyze correlation structures in a variety of areas. In this paper, Thorsten Oest and Jörg Rollbühler focus on its applications in the context of market risk. Factor analysis provides a framework to analyze the correlation structure of financial time series and therefore is suited to quality management of correlation matrices. Furthermore, we discuss the evaluation of market risk concentration for which we combine factor analysis and the k-Means cluster search algorithm. From a technical point of view, we describe an efficient decomposition algorithm for large correlation matrices of Wishart type. In all of the methods applied in this paper, we provide algorithms ready for implementation. Proof of concept for these algorithms is demonstrated by application to real data.

Introduction

With improved methodologies, market risk management faces an increasing number of risk drivers, with numbers of 5,000 to 10,000 risk drivers not uncommon. With respect to complexity, the main challenge is to handle the correlations between these risk drivers. A typical correlation matrix can easily contain millions of correlation coefficients. The vast amount of coefficients not only raises a problem for data processing but also for quality assurance and risk analysis.

In data processing, the decomposition of the correlation matrix is particularly ambitious. For a sample correlation matrix determined from the time series of risk driver returns, we propose a decomposition approach which takes advantage of the low rank of the matrix. With this approach, only a sub-matrix of the correlation matrix is decomposed using standard methods. The full decomposition is deduced by applying additional simple operations. This reduces the computational complexity considerably.

Apart from this technical aspect, we have studied analysis methods for large correlation matrices. Here factor analysis is a useful tool. In finance, factor models have a long history. A main area of application is portfolio optimization, where factor models were first introduced via the capital asset pricing model by Sharpe (1964), Lintner (1965), and Mossin (1966). Today, factor models are widely used and integrated in commercial risk models like those by MSCI Barra or APT. Whilst in asset management

diversification gives importance to small correlations, in market risk management the main focus of attention is on strong correlations. For the identification of the most relevant correlations, factor analysis is well suited.

In this paper, we will concentrate on two practical aspects. Firstly, we discuss a method for quality checks of large correlation matrices. The idea is to focus on those few factors that describe most of the correlation structure and verify the expectation that these factors represent markets or segments where significant correlations are present.

In a second application, we combine factor analysis with cluster analysis to evaluate risk concentration. As a result of the recent financial crisis, risk concentration attracts increasing attention (CEBS, 2009). Market risk concentration depends mainly on the highly complex correlation and investment structure. We propose a method which reduces this complexity by building clusters with similar correlation and investment characteristics. Having identified these clusters, the effect of specific changes in the condensed correlation structure can be studied by means of stress testing or stressed value at risk (VaR) to quantify the contribution to concentration risk.

For the purpose of demonstration, all methods proposed in this paper were applied on publicly available historical data (July 2008 to July 2009) for about 5,000 risk factors. The main fraction of risk factors are US equities ($\approx 2,800$), European equities ($\approx 1,600$), South American equities (≈ 70), Australian equities (≈ 40), US, European, and Asian yield curves (≈ 400), British inflation rates (≈ 75), and FX rates (30). Covariance coefficients are calculated on log returns for equities and FX rates and on absolute returns for interest rates. To clean up data, only a simple outlier rejection is applied by removing returns in excess of 10 standard deviations.

Setting up the Multi-factor Model

Factor model for risk factors

In this section we describe the multi-factor framework. Given are returns for N risk factors which we assume to be described by a multivariate normal distribution.¹ The distribution parameters are estimated from time series of observed returns:

r_{it} Return of risk factor i for the time period $[(t-1), t]$ with $t \in \{1, 2, \dots, T\}$.

$\mathbf{R} = (r_{it})_{(N,T)}$ Historical return matrix.

$\Sigma = \frac{1}{T} \cdot \mathbf{R}\mathbf{R}'$ Sample covariance matrix which is by definition a matrix of Wishart type. For simplicity, we assume that the mean return is zero, which is a good approximation within the scope of market risk. \mathbf{R}' denotes the transpose of \mathbf{R} .

$\sigma_i = \sqrt{\Sigma_{ii}}$ The volatility of risk factor i .

Within the factor model, there are further restrictions imposed on the multivariate normal distribution to reduce the complexity of the correlation structure. Correlations are restricted on a limited number of factors s_{kt} ($k=1, \dots, m$ with $m \leq N$):

$$r_{it} \approx \mu_i + \sum_{k=1}^m \beta_{ik} \cdot s_{kt} + \varepsilon_{it}, \quad (1)$$

assuming that the systematic part $\varepsilon_{it} \sim N(0, \psi_i)$ and the factor returns $s_{kt} \sim N(0, 1)$ are all uncorrelated. For our purpose, the mean returns μ_i can be neglected. Nonetheless, in the treatment where the mean returns μ_i are included, r_{it} is replaced by $r_{it} - \mu_i$ in the following given formulas. Other classes of factor models, e.g. index factor models, do not require factors to be uncorrelated. For our applications uncorrelated factors are preferable.

The idea of the factor model is to reduce the correlation complexity but not to change the volatilities of the risk factors. From this point of view, it is more adequate to consider the correlation matrix instead of the covariance matrix, or in other words to set up a factor model for the normalized returns. This is obtained by replacing $r_{it} \rightarrow r_{it}^{\%}$ in the formulas above. In this case, Σ is the correlation matrix.

In order to reduce the $N \cdot (N-1)/2$ correlation coefficients, the correlation structure within the factor model is fixed by the β matrix of dimension $N \times m$:

$$\Sigma^{\text{FM}} = \beta \beta' + \text{diag}[\psi_1^2, \psi_2^2, \dots, \psi_N^2].$$

There are several ways to estimate which factors to use. In common practice, this is done by an eigendecomposition of the sample covariance matrix:

$$\Sigma = \sum_{k=1}^N \lambda_k \cdot \mathbf{v}_k \mathbf{v}_k',$$

where \mathbf{v}_i is the eigenvector corresponding to the eigenvalue λ_i , rank ordered $\lambda_1 \geq \lambda_2 \dots \lambda_N$. The eigenvalues can be seen as the variance of the risk factor combination given by the corresponding eigenvalue:

$$\text{Var}\left(\sum_{i=1}^N v_{ki} \cdot r_i\right) = \mathbf{v}_k' \Sigma \mathbf{v}_k = \lambda_k \cdot \mathbf{v}_k' \cdot \mathbf{v}_k = \lambda_k.$$

Obviously, the decomposition of Σ is of the same form as Σ^{FM} and we can set

$$\beta_{ik} = \sqrt{\lambda_k} \cdot v_{ki}.$$

For $m=N$ we get $\Sigma^{\text{FM}} = \Sigma$, with $\psi_i = 0$. In all other cases, the specific part is given by:

$$\psi_i^2 = (\Sigma - \beta' \beta)_{ii} = \sum_{k=m+1}^N \lambda_k \cdot (\mathbf{v}_k \mathbf{v}_k')_{ii} = \sum_{k=m+1}^N \lambda_k \cdot v_{ki}^2,$$

so that Σ and Σ^{FM} have the same diagonal elements but in general will be different off-diagonal. The contribution to the total variance from the specific part is $\sum_{i=1}^N \psi_i^2 = \sum_{k=m+1}^N \lambda_k$ and accordingly the contribution from the systematic part is given by the total communality $\text{trace}(\Sigma) - \sum_{i=1}^N \psi_i^2 = \sum_{k=1}^m \lambda_k$.

It should be mentioned that within this framework, we can also easily separate specific and systematic risk for a VaR calculated with historical simulation. The systematic and specific risk are obtained by running the historical simulation not with returns r_{it} but with $r_{it}^{\text{systematic}} = \sum_{k=1}^m \beta_{ik} \cdot s_{kt}$ or $r_{it}^{\text{specific}} = r_{it} - r_{it}^{\text{systematic}}$, respectively. For this we have to estimate the factor returns s_{kt} , which are derived by averaging equation (1) for $m=N$ over risk factors: $\sum_{i=1}^N \beta_{il} \cdot r_{it} = \lambda_l \cdot s_{lt}$. Thus, we have

$$s_{lt} = \frac{1}{\lambda_l} \cdot \sum_{i=1}^N \beta_{il} \cdot r_{it} = \frac{1}{\lambda_l} \cdot \sum_{i=1}^N v_{li} \cdot r_{it}.$$

Factor model for risk positions

The previous section focused only on features of the risk factors. In this section, we discuss the application of factor analysis in the context of value at risk, where we restrict ourselves to the delta-normal approach. The delta-normal VaR for a portfolio is given by

$$\text{VaR}^2 = c \cdot \Delta' \Sigma \Delta = c \cdot \sum_{i,j} \Delta_i \cdot \Sigma_{ij} \cdot \Delta_j \equiv c \cdot \sum_{i,j} \hat{\Sigma}_{ij},$$

where $\Delta_i = \frac{\partial \text{PV}}{\partial r_i}$ is the sensitivity of the portfolio value with respect to the risk factors, c is a factor which accounts for the confidence level and time horizon, and Σ is the covariance matrix. If we replace $\hat{\Sigma}$ by its eigendecomposition

$$\hat{\Sigma} = \sum_{k=1}^N \hat{\lambda}_k \cdot \hat{\mathbf{v}}_k \hat{\mathbf{v}}_k',$$

where $\hat{\lambda}_i$ are the eigenvalues and $\hat{\mathbf{v}}_i$ the corresponding eigenvectors of $\hat{\Sigma}$, we get

$$\text{VaR}^2 = c \cdot \sum_{i,j} \hat{\Sigma}_{ij} = c \cdot \sum_{k=1}^N \hat{\lambda}_k \cdot \left(\sum_{i=1}^N \hat{v}_{ki} \right)^2 = c \cdot \sum_{k=1}^N \left(\sum_{i=1}^N \hat{\beta}_{ik} \right)^2 \quad (2)$$

with $\hat{\beta}_{ik} = \sqrt{\hat{\lambda}_k} \cdot \hat{v}_{ki}$. Risk concentration is mainly associated with large eigenvalues, although the total risk contribution might be small, due to hedging or due to anti-correlation of risk factors. Nevertheless, the main potential risk sources are associated with large eigenvalues. We will come back to the analysis of risk concentration in the case study described below.

The corresponding factor model for the risk positions is

$$\hat{r}_{it} := \Delta_i \cdot r_{it} = \sum_{k=1}^m \hat{\beta}_{ik} \cdot \hat{s}_{kt} + \hat{\varepsilon}_{it}.$$

If the focus does not lie on risk factors but on other aggregation categories, e.g. currencies or portfolios, the returns of the corresponding aggregation categories have to be considered. Suppose we are interested in the risk concentration with respect to trading portfolios. The investment in each portfolio p can be described by sensitivities Δ_{pi} with respect to the risk factors i . The VaR can be rewritten as

$$VaR^2 = c \cdot \sum_{p,q} \sum_{i,j} \Delta_{pi} \cdot \Sigma_{ij} \cdot \Delta_{qj} \equiv c \cdot \sum_{p,q} \tilde{\Sigma}_{pq}$$

and a factor analysis for portfolio returns can be achieved by an eigendecomposition of $\tilde{\Sigma}_{pq}$ followed by a risk cluster analysis.

Estimation of the numbers of factors

So far we have not addressed the question of how many factors should be used in a factor model. The simplest way is to deduce it from the total communality spectrum using common sense, i.e. taking only those factors which account for a well-above-average contribution to the total communality. For an overview of more objective methods, see Bai and Ng (2008). Most published studies were done with the focus on portfolio optimization using relatively long time series. In market risk management, the time horizon of historical data is typically one year. In the following, we compare two estimates for the number of factors for a one-year horizon.

A method proposed by Laloux *et al.* (1999, 2000) is based on random matrix theory. For a Wishart-type random matrix, the eigenvalue spectrum is known to be (Marčenko & Pastur, 1967):

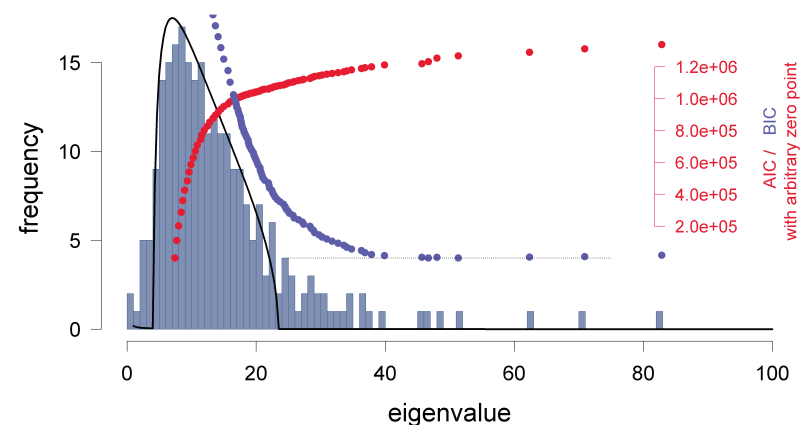
$$\frac{1}{N} \frac{dn(\lambda)}{d\lambda} = \begin{cases} (1-Q) \cdot \delta(\lambda) & \text{for } Q < 1, \lambda < \lambda_{min} \\ \frac{Q}{2\pi\sigma^2} \frac{\sqrt{(\lambda_{max}-\lambda)(\lambda-\lambda_{min})}}{\lambda} & \text{for } \lambda_{min} \leq \lambda \leq \lambda_{max} \\ 0 & \text{otherwise} \end{cases}$$

$$\lambda_{min,max} = \sigma^2 \cdot \left(1 \pm \frac{1}{\sqrt{Q}} \right)^2$$

in the limit $N \rightarrow \infty$ and $T \rightarrow \infty$ with $Q = T/N$. Fitting this distribution with free parameters Q and σ^2 to our data set gives a reasonable description of the lower part of the eigenvalue spectrum (see Figure 1). In our example, there are 37 eigenvalues above λ_{max} .

Other popular methods to estimate the number of model parameters are based on the Akaike's Information Criterion (AIC) (Akaike, 1974) and the Bayesian Information Criterion (BIC) (Schwarz, 1978). The optimal number of parameters is obtained when the value of the information criterion is minimal. The AIC is defined by (Akaike, 1974): $AIC = -2 \ln \mathcal{L} + 2n_p$, where $\ln \mathcal{L}$ is the mean expected logarithm of the likelihood of finding the sample given by the time series and n_p the number of model parameters. The BIC has a larger penalty factor for additional model parameters: $BIC = -2 \ln \mathcal{L} + n_p \cdot \ln T$. For a factor model, we have (Akaike, 1987):

Figure 1: Lower part of the eigenvalue spectrum of the correlation matrix (histogram) fitted to the theoretical spectrum of a random matrix (line). There are three eigenvalues at higher values of 913, 295, and 115. The Akaike's and the Bayesian Information Criterion are displayed as red and blue dots as a function of the eigenvalue which corresponds to the smallest principal component included in the factor model.



$$\ln \mathcal{L} = -\frac{T}{2} \cdot \left[\ln \det(\Sigma^{FM}) + \text{tr}(\Sigma \cdot \Sigma^{FM^{-1}}) \right]$$

and

$$n_p = N \cdot (m+1) - \frac{1}{2} \cdot m \cdot (m-1).$$

Figure 1 shows the values of AIC and BIC for different numbers of principal components used. There are nine factors above the local minimum of the BIC value. The global minimum of both, AIC and BIC, is on the left side of Figure 1 (values not shown), which would suggest usage of all factors. None of the information criteria is in agreement with the boundary of the fitted eigenvalue spectrum of a random matrix. See also McDonald and Marsh (1990) for a critical appraisal of the AIC approach for multivariate models. It should be mentioned that for the calculation of AIC, Σ^{FM} is usually determined from a maximum likelihood estimation to reduce the bias on $\ln \mathcal{L}$. But even in this case there remains a bias for small sample sizes. For a multiple regression model the final sample correction is known to be substantial (Sugiura, 1978).

Based on the random matrix analysis, not more than about 37 factors should be used in our example. Given these factors, the full correlation matrix of the factor model is determined by adjusting the specific factors accordingly. This construction of a correlation matrix can be seen as an attempt to remove noise from the correlation matrix. Since we are mainly interested in the identification of those factors which are not dominated by noise, we do not go into more detail concerning noise reduction in correlation matrices. There is a wide area of publications on this topic, see for example the recent comparison of methods with respect to risk forecasting by Briner and Connor (2008).

Decomposition of large matrices

Most decomposition algorithms scale with the cube of the matrix dimension, which can considerably slow down the factor analysis when using a large number of risk factors. Since empirical covariance matrices are calculated from finite return time series of length T , a typical scenario in the market risk context is that the rank of these matrices is much lower than the number of risk factors N . In those cases, the covariance matrix cannot contain more information than the return time series (consisting of $N \times T$ elements). It is therefore conceivable that this property of lower than full rank can be used to speed up eigenvalue decomposition. Depending on the input data available, there are different techniques to handle large amounts of data. Here we discuss the two possibilities, either that the matrix of return time series is given, or, alternatively, that the covariance is available. In the latter case, the decomposition should be done for the correlation matrix because in general it is better conditioned.

The method described is also useful for VaR calculations using the Monte Carlo simulation method.

Firstly we consider the case where the return matrix R is known. In order to find the eigenvalues of $\Sigma = \frac{1}{T} \cdot RR'$, we can decompose this rectangular matrix R by singular value decomposition (SVD) into $R = V S U'$. Here V is an orthogonal $N \times N$ matrix, S is a $T \times N$ matrix, while U is an orthogonal $T \times T$ matrix. The diagonal elements of the matrix S are non-negative and are called singular values. All other matrix elements vanish. From this, we obtain an eigenvalue decomposition of the form

$$\Sigma = V \Lambda V' \text{ with } \Lambda = \frac{1}{T} \cdot S S'$$

Let us now consider the task of quick eigenvalue decomposition for a given correlation matrix Σ with unknown return matrix. We assume that it is of Wishart type, i.e. it is estimated from N time series consisting of T returns each. Then, the rank of Σ cannot be higher than T . We first neglect numerical errors and additionally assume that the first T return time series are linearly independent. Then Σ has rank T . The matrix Σ can be divided in blocks as

$$\Sigma = \begin{pmatrix} \mathbf{F} & \mathbf{G}' \\ \mathbf{G} & \mathbf{H} \end{pmatrix}$$

with a regular $T \times T$ matrix \mathbf{F} , a $(N - T) \times T$ matrix \mathbf{G} , and a $(N - T) \times (N - T)$ matrix \mathbf{H} . First, we describe how a decomposition of the form $\Sigma = \mathbf{A} \mathbf{A}'$ with a $N \times T$ matrix \mathbf{A} can be achieved effectively. Since \mathbf{F} has full rank, a decomposition of the form $\mathbf{F} = \mathbf{a} \mathbf{a}'$ exists, which in this idealized case can be found by standard methods. Then

$$\mathbf{A} = \begin{pmatrix} \mathbf{a} \\ \mathbf{G}(\mathbf{a}^{-1})' \end{pmatrix}$$

In order to show this, we simply build the product $\mathbf{A} \mathbf{A}'$, which equals Σ exactly if $\mathbf{H} = \mathbf{G} \mathbf{F}^{-1} \mathbf{G}'$. This is indeed true, since the reduced rank of the matrix leads to a linear dependency $(\mathbf{G} \ \mathbf{H}) = \mathbf{L}(\mathbf{F} \ \mathbf{G}')$ with \mathbf{L} being a $N \times T$ matrix. Since \mathbf{F} is invertible, we find $\mathbf{L} = \mathbf{G} \mathbf{F}^{-1}$. Inserting this into $\mathbf{H} = \mathbf{L} \mathbf{G}'$ leads to the desired result. The eigenvalue decomposition is then easily found by applying the singular value decomposition as described in the last subsection.

In real applications, correlation matrices might be not well conditioned and one has to deal with numerical inaccuracies. Furthermore, returns may be linearly dependent, or gaps and manual corrections might be involved. We first address the issue of conditioning.

Numerical errors are of special importance if risk factors are ordered in risk factor groups and the matrix \mathbf{F} contains highly correlated risk factors. Then, the basis vectors do not span a wide enough angle and the inverse matrix \mathbf{a}^{-1} may not be well defined.

This problem can be solved by a careful choice of the risk factors included in the sub-matrix \mathbf{F} , but in most cases it is sufficient to look at a somewhat larger number of risk factors. Although this will add zero eigenvalues, it will lead to a better condition of the non-zero eigenvalues. For random matrices one can show that eigenvalues close to zero are extremely unlikely, if the sub-matrix is chosen larger than T . The eigenvalues drop sharply close to zero between the eigenvalue T and the eigenvalue $T + 1$. This feature is observed in real matrices, too. The above procedure works if the lowest non-vanishing eigenvalue can be clearly distinguished from the numerical noise created by rounding errors. (This also means a good conditioning, since the highest eigenvalue has the trace as upper bound.)

The following algorithm provides a stable and extremely fast decomposition $\Sigma = \mathbf{A} \mathbf{A}'$:

1. From the N risk factors choose $k > T$ risk factors that will define the sub-matrix \mathbf{F} . Ideally, the corresponding risk factors are chosen from different markets, such that they are not highly correlated.

2. Load the first k rows of the matrix and divide it into blocks \mathbf{F} and \mathbf{G}' as described above. (The other rows follow from linear dependence.)

3. Decompose $\mathbf{F} = \mathbf{Q} \mathbf{D} \mathbf{Q}'$ with an orthogonal $k \times k$ matrix \mathbf{Q} and a diagonal matrix $\mathbf{D} = \text{diag}(\lambda_1, \dots, \lambda_k)$, with rank-ordered eigenvalues. There are $q < k$ non-vanishing eigenvalues. If the eigenvalue q does not clearly exceed the eigenvalue $q + 1$, the matrix \mathbf{F} has to be chosen using more risk factors or less correlated risk factors.

4. Set $\mathbf{a} = \mathbf{Q} \mathbf{D}^{\frac{1}{2}}$ and build the pseudo-inverse

$$\mathbf{b} = \text{diag}\left(\frac{1}{\sqrt{\lambda_1}}, \dots, \frac{1}{\sqrt{\lambda_q}}, 0, \dots, 0\right) \mathbf{Q}'.$$

5. Set $\mathbf{A} = \begin{pmatrix} \mathbf{a} \\ \mathbf{G} \mathbf{b} \end{pmatrix}$ and delete all columns of norm zero in order to get a sparse matrix. In most cases this is already the desired decomposition. Exceptions can be caught by the following two steps.

6. This step is optional, but is a fast way to capture the main corrections. Check whether the diagonal elements of the correlation matrix are reproduced correctly by multiplying each row of \mathbf{A} by its transpose. Find all indices for which this is not true. If there are any, include the corresponding risk factors in the matrix \mathbf{F} and repeat steps 1–6.

7. Check the correctness of the decomposition by multiplication of \mathbf{A} with its transpose and comparison with the correlation matrix. Should there be any deviations, include the corresponding risk factors in the matrix \mathbf{F} and repeat steps 1–7.

Step 6 takes care of certain manual corrections that lead to isolated subspaces that are not captured by the initial choice of \mathbf{F} . Note that, in such

a case, the trace of \mathbf{F} will usually not equal the number of risk factors, as would be expected for a correlation matrix. Therefore, checking only the diagonal elements (which is a very quick check) will usually reveal the rows that span extra dimensions. Note that step 7 is more time-consuming, but takes care of all exceptions. It guarantees a safe decomposition. Finally, \mathbf{A} can be used to find the eigenvalue decomposition by SVD as described above for return input, i.e. using the time series directly.

Application to Data Quality Management

Data quality is essential for a reliable market risk determination. Therefore, great effort is spent cleaning up historical market data. The most difficult part is checking the data quality of correlations due to the large number of these which are present. Factor analysis provides a convenient method for this task and has already been used to address similar questions, see for example Plerou *et al.* (1999).

Suppose we have a correlation matrix, or the time series of historical returns as in our data sample, and would like to check the quality of the roughly 12.7 million correlations. This can be done in the following way. We start with an eigenvalue decomposition. After this, we estimate the number of factors which are of importance as explained above. In our example these are around 37, which correspond to $37 \cdot 5,039 = 186.443$ eigenvector components. These eigenvectors will then be checked for consistency with our expectations that risk factors within the same market share a similar behavior, e.g. all equities will be correlated.

To check this assumption the risks factors must be categorized. In our example we use four attributes per risk factor. These are:

1. Risk factor group to separate interest rate, equities, and foreign exchange
2. Country
3. Curve type/sector for interest rates/equities
4. Term/industry for interest rates/equities

All risk factors which have at least one or two attributes in common will be considered as a member of a set G_j . For example, all equities belong to one group, all Chinese equities to another group. Note that these groups are not disjoint. Potentially, these are sets where the risk factors within the set are strongly correlated.

The major eigenvectors should select those markets with the largest internal correlation. To verify this, we define the unit vector $\mathbf{u}_{G_j} = 1/\sqrt{|G_j|} \cdot \sum_{i \in G_j} \mathbf{u}_i$ which points in the positive direction given by the components of the set G_j , with \mathbf{u}_i being the standard basis of the Cartesian coordinate system and $|G_j|$ the cardinality of the set. For each eigenvector and each subspace spanned by G_j , the projection of the eigenvector on \mathbf{u}_{G_j} is:

$$w_k(G_j) = \mathbf{u}_{G_j} \cdot \mathbf{v}_k = \frac{1}{\sqrt{|G_j|}} \sum_{i \in G_j} v_{ki}$$

The maximal values $w_k(G_j) = \pm 1$ correspond to a full correlation in set G_j . No correlation within the set corresponds to $w_k(G_j) = 0$.

Once all values $w_k(G_j)$ for a given eigenvector have been calculated, these values are ordered by decreasing norm, e.g. $|w_k(\text{equities})| \geq |w_k(\text{US equities})| \geq |w_k(\text{IR})| \geq \dots$. Since the sets are not disjoint, we remove (starting from the top) those sets in the list which have risk factor elements already included in a set above. In this way we end up with a sequence of disjoint sets which are ordered according to their projection value. The result of this procedure is shown for the first and second eigenvectors out of the data sample in Figures 2 and 3.

By visual inspection of the factor loadings for the most important sets, artifacts can be identified and further analyzed. The first two eigenvectors reveal correlations in the equity market. The general behavior is reasonable apart from small artifacts in the British interest rates which are best seen in Figure 2. There, factor loadings for British interest rates are shown as orange dots. From left to right there is a government curve, an inflation curve, and a general yield curve. At the left short-term region of each curve, the loadings drop significantly. By inspection of the time series, this effect can be ascribed to a higher number of returns with zero values for short-term rates, which reduces correlations.

For equities the same effect is present in the data, as can be seen in Figure 3 for US equities. The bulk of equities have a factor loading around -0.02 , but there is a tail toward zero factor loadings. In Figure 4, the factor loading for each US equity is plotted against the number of zero returns in the time series. It is clearly seen that correlations diminish with increasing number of returns with zero value. This shows that the data quality of our test sample needs to be improved, e.g. by combining prices from different exchanges.

Figure 2: The factor loadings of the first principal component show an overall correlation in the equity market.

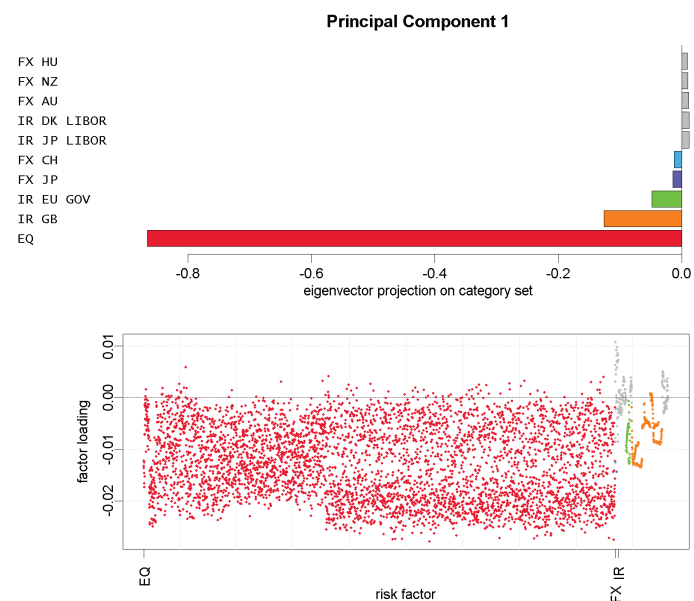


Figure 3: The second principal component indicates an anti-correlation between the US and European equity markets.

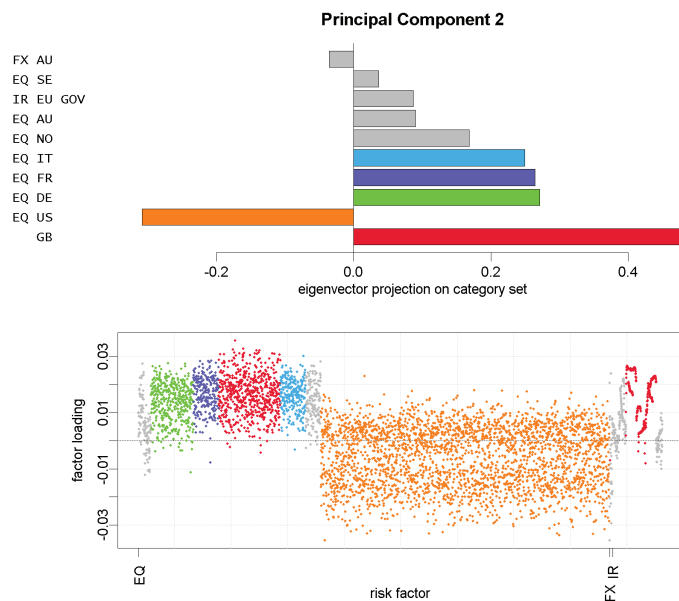
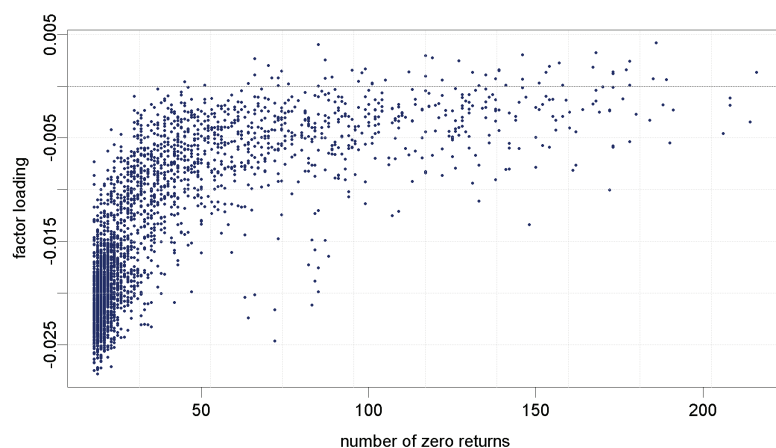


Figure 4: For US equities, the factor loadings of the first principal component are plotted against the number of returns with zero value in the time series. The correlations diminish with increasing number of zero returns, which shows the effect of poor data quality on the estimated correlations.



Analysis of Market Risk Concentration

In this section we will address how the analysis of risk concentration can be simplified by using a factor model. What do we mean by risk concentration? Primarily, risk concentration always refers to a set of categories to which risk is allocated. These categories could be risk factors, countries, portfolios,

or any other category of interest. So, risk concentration can be viewed with regard to different aspects.

Suppose our set of categories under consideration consist of risk factors. A risk concentration is apparent if there is a subset of risk factors which together carry a risk well above average. This requires that each risk factor in the subset carries a substantial risk and that the risk contributions per risk factor are correlated.

In the delta-normal VaR approach, risk concentration can be studied in the following way. Remember that the delta-normal VaR can be decomposed by eigenvectors of $\hat{\Sigma}_{ij} = \Delta_i \cdot \Sigma_{ij} \cdot \Delta_j$:

$$VaR^2 = c \cdot \sum_{k=1}^N \hat{\lambda}_k \cdot \left(\sum_{i=1}^N \hat{v}_{ki} \right)^2 = c \cdot \sum_{k=1}^N \left(\sum_{i=1}^N \hat{\beta}_{ik} \right)^2$$

with $\hat{\beta}_{ik} = \sqrt{\hat{\lambda}_k} \cdot \hat{v}_{ki}$. Within this model, the risk contribution per risk factor i is fully determined by N factors $\hat{\beta}_{ik}$. Or, in other words, the risk contribution of each risk factor is given by a point in \mathbb{R}^N , where the position vector has components $\hat{\beta}_{ik}$. For an example see Figure 6 later, where the risk factor points are shown as projections on the first three dimensions. For a combination of risk factors, the total risk is obtained by the norm of the sum of the position vectors up to the factor \sqrt{c} . Therefore, clusters of points which are sufficiently separated from the point of coordinate origin correspond to risk concentration.

Based on these considerations, we propose the following algorithm:

1. From a principal component analysis, the factors $\hat{\beta}_{ik} = \sqrt{\hat{\lambda}_k} \cdot \hat{v}_{ki}$ are determined.

2. We exclude risk factors j with small risk contributions, i.e. $\sqrt{\sum_{k=1}^N \hat{\beta}_{jk}^2} < \alpha \cdot \text{mean}_i \left(\sqrt{\sum_{k=1}^N \hat{\beta}_{ik}^2} \right)$.

3. To reduce the complexity, only the first m principal components $k = 1, \dots, m$ will be considered.

4. For the points given by $\hat{\beta}_{ik}$ in \mathbb{R}^m , the cluster algorithm k -Means by Hartigan and Wong (1979) (we use the implementation in the statistical computing software R) is run to assign the factors to clusters. The number n_c of clusters will be predefined and specifies the granularity in which the user likes to see the risk concentration. The k -Means algorithm optimizes the location of the cluster center in such a way that the average squared distance of cluster points to the cluster center is minimal. Each point is assigned to the cluster with the nearest cluster center.

In practice, there are often a substantial number of factors with small risk. The k -Means algorithm would find a cluster for these factors close to the origin of coordinate, but the cluster size would not necessarily be small, since it depends on the position of other clusters. This could spoil the assignment of points to clusters. Therefore, we exclude in our algorithm as in step 2 factors with small risk accounts in the k -Means algorithm.

We will illustrate the algorithm on our market data sample. We consider as example an investment of:

- 5 million € long in US financial equities
- 5 million € short in US technology equities

- 10 million € long in European government bonds
- 10 million € short in British government bonds
- 10 million € long in British pound futures

Each investment is allocated equally to the risk factors in the corresponding group. For the exclusion criterion (2) we choose $\alpha = 0.01$. For the cluster algorithm, we have to decide how many principal components to take into account. For this example, we restrict ourselves to the first three principal components which have a communality of almost 60 percent. To include more principal components does not change the result significantly. In addition, we have to fix the number of clusters. An idea of the granularity of the data is obtained by considering how fast the within-cluster sum of squares decreases with increasing number of clusters. This is shown in Figure 5. It indicates that between four and six clusters seems to be a reasonable choice. Obviously, the interpretation of the figure is to some degree a matter of taste. For practical applications we propose to fix the number of clusters to the granularity to which the risk should be decomposed.

In this very simple example, risk concentration has its origin in the four investment classes: US technology equities, US financial equities, EU interest, and GB interest. The FX risk is hedged, so no contribution is expected in FX risk factors. To see if we recover the four investment classes, we run the cluster algorithm for four clusters. The location of the cluster in the subspace of the first three principal components is shown in Figure 6. Figure 7 shows the volatility and correlations for clusters found and clusters according to the investment groups. The agreement is reasonable. But it should be stressed that the simple recovery of investment concentration as observed in this example is not the aim of the method. In fact, investment areas may merge to a common cluster or are split into different clusters depending on the correlations of underlying risk factors. The advantage of our algorithm is that there is no bias arising from specific predefined investment categories.

Figure 5: Within-cluster sum of squares as a function of the number of clusters. Up to four to six clusters; additional clusters lower the within-cluster sum of squares substantially.

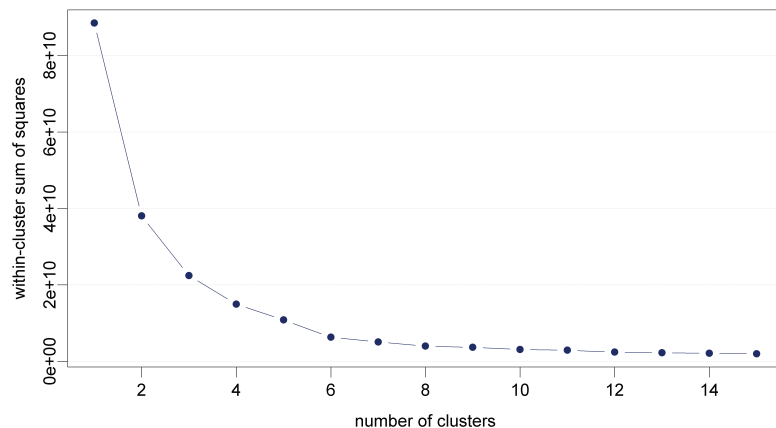


Figure 6: Result of the cluster search in the subspace spanned by the first three principal components. Each point represents the β values for one risk factor. The clusters found are indicated by various colors. For interest rates, the term structure is visible due to the effect that the risk increases with duration. For short terms, the interest rate risk factors merge with the equity clusters. Risk factors excluded in the cluster search are close to the coordinate origin and are therefore hidden between the two equity clusters.

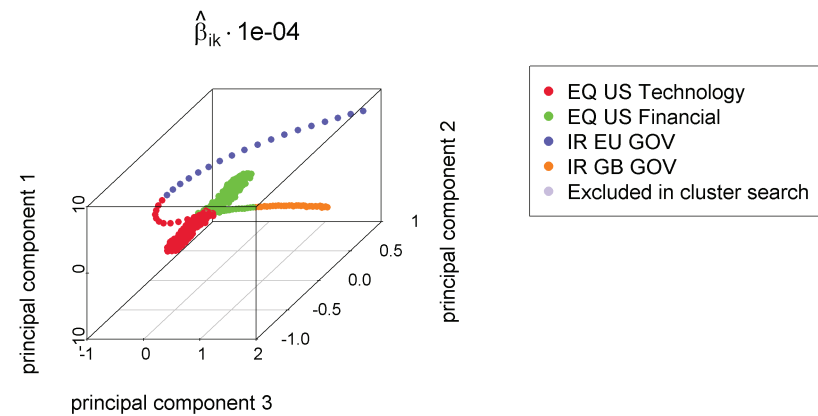
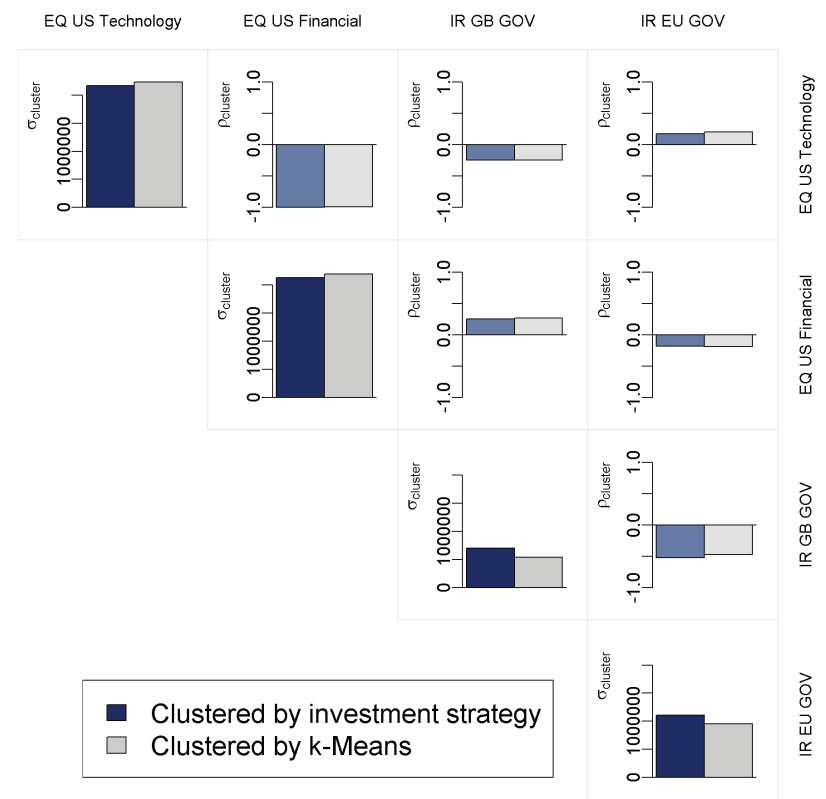


Figure 7: Daily absolute volatilities and correlations of the total investment of the four clusters. Compared are values from the cluster search with clusters given by the investment strategy.



Conclusion

We have proposed procedures which simplify the decomposition of large correlation matrices and the analysis of correlation effects in market risk management. The methods take advantage of the fact that correlation coefficients are estimated from time series with a short horizon compared to the number of risk factors. The correlation information which can be extracted from these time series is much less than the dimension which the full correlation matrix suggests. For the analysis, correlation information is further structured by means of factor analysis. This allows us to tackle more profound issues, as we have demonstrated with the analysis of risk concentration.

Acknowledgment

The authors would like to thank J. Dittrich, whose support has been of great value.

FOOTNOTE

1. The assumption of a normal distribution should be seen as the market standard. It is known that the distributions in reality show fat tails; these have to be taken care of by supplementary methods like stress testing.

REFERENCES

- Akaike, H. 1974. A new look at the statistical model identification. *IEEE Transactions on Automatic Control* 19:716–723.
- Akaike, H. 1987. Factor analysis and AIC. *Psychometrika* 52(3): 317–332.
- Bai, J., and Ng, S. 2008. Large dimensional factor analysis. *Foundations and Trends® in Econometrics* 3(2):89–163.
- Briner, B.G., and Connor, G. 2008. How much structure is best? A comparison of market model, factor model and unstructured equity covariance matrices. *The Journal of Risk* 10 (4): 3–30.
- CEBS. 2009. CEBS Guidelines on aspects of the management of concentration risk under the supervisory review process. CEBS Consultation Paper 31.
- Hartigan, J. A., and Wong, M. A. 1979. A K-means clustering algorithm. *Applied Statistics* 28, 100–108.
- Laloux, L., Cizeau, P., Bouchaud, J.P., and Potters, M. 1999. Noise dressing of financial correlation matrices. *Physical Review Letters* 83 (7): 1467. Random matrix theory. *Risk* 12 (3): 69.
- Laloux, L., Cizeau, P., Bouchaud, J.P., and Potters, M. 2000. Random matrix theory and financial correlations. *International Journal of Theoretical and Applied Finance* 3 (3): 391–397.
- Lintner, J. 1965. The valuation of risk assets and the selection of risky investments in stock portfolios and capital budgets. *Review of Economics and Statistics* 47 (1): 13–37.
- Marčenko, V.A., and Pastur, L.A. 1967. Distribution of eigenvalues for some sets of random matrices. *Mathematics of the USSR-Sbornik* 1(4): 457–483.
- McDonald, R., and Marsh, H. 1990. Choosing a multivariate model: noncentrality and goodness of fit. *Psychological Bulletin* 107 (2): 247–255.
- Mossin, J. 1966. Equilibrium in a capital asset market. *Econometrica* 34 (4), 768–783.
- Plerou, V., Gopikrishnan, P., Rosenow, B., Amaral, L.A.N., and Stanley, H.E. 1999. Universal and non-universal properties of cross-correlations in financial time series. *Physical Review Letters* 83, 1471.
- Schwarz, G. 1978. Estimating the dimension of a model. *Annals of Statistics* 6:461–464.
- Sharpe, W. 1964. Capital asset prices: a theory of market equilibrium under conditions of risk. *Journal of Finance* 19 (3), 425–442.
- Sugiura, N. 1978. Further analysis of the data by Akaike's information criterion and the finite correction. *Communications in Statistics A – Theory and Methods* 7: 13–26.