



Principal Component Analysis

34.1 The General Procedure

In addition to the autoregressive models described above, which are used for instance in the form of GARCH models when modeling volatility, a further technique of time series analysis, called *principal component analysis* (abbreviated as *PCA*), is widely applied in the financial world. This technique is employed in the analysis of term structure evolutions, for instance. A first approach to describe the *stochastic* dynamic of an interest rate term structure could be to define a risk factor for each vertex of the curve, i.e. the zero rate or the forward rate at that vertex. However, because of the large number of vertices, this approach would be quite calculation intensive. Instead, it is possible to reduce the number of stochastic variables to just a few (1 or 2, sometimes more) driving factors without loss of too much information. This approach has its justification in principal component analysis. Principal component analysis is a statistical technique which extracts the statistical components from the time series which are most relevant for the dynamics of the process in order of their importance. Applying this method to interest rates (i.e. the vertices of the interest rate curve) shows that often more than 90% of the term structure's dynamics can be ascribed to the one or two most important components.

Two other well-known statistical methods used in time series analysis are *factor analysis* [32] and *cointegration* [86]. Principal component analysis sets itself apart from these other methods by the ease with which the results

can be interpreted. An extensive description of principal component analysis can be found in [116], for example.

Since the time dependence of the data is not analyzed in PCA, the remarks made in Chap. 32 hold for this type of time series analysis as well: principal component analysis is a suitable approach for *stationary* time series only. Time series exhibiting non-stationary behavior like trends, etc. should be first freed of their non-stationary components by, for example, estimating trends and subsequently subtracting them from the time series or by taking differences of the data with respect to time, in other words, by performing a pre-treatment as described in Chap. 35. A reasonable principal component analysis of non-stationary time series is not to be achieved without first removing these potential non-stationary components.

In principal component analysis, the time series of not just one but *several* stochastic process are considered which may be strongly correlated. Principal component analysis is therefore an example of *multivariate* time series analysis. In financial applications, each process represents a risk factor. The situation is thus similar to that in Eq. 21.21 in Sect. 21.5 where n risk factors were considered as well. In the case of PCA, it is not a matter of the stochastic differential equations of the risk factors, but of their historic evolution. For each risk factor, there exists an associated time series of data for times t_k , $k = 1, \dots, T$. For example, the data of the relative changes (as in Eq. 31.1)

$$X_i(t_k) = \ln \left(\frac{S_i(t_k + \delta t)}{S_i(t_k)} \right) \approx \frac{S_i(t_k + \delta t) - S_i(t_k)}{S_i(t_k)}$$

$$i = 1, \dots, n, \quad k = 1, \dots, T .$$

These time series are assumed to be stationary, i.e., pre-treatment of the data has been performed already and the non-stationary components have been removed, if this has proved to be necessary. As a typical example of such a group of n risk factors, we can consider the relative changes of the interest rates at the vertices of an interest rate curve. The variances and pair wise correlations of these n risk factors can be arranged in a covariance matrix $\delta \Sigma$ as in Eq. 21.22. The entries of the covariance matrix can be determined from historical data as described in Sect. 31.1.

The dimension of this matrix is obviously equal to the number of risk factors (equivalently, the number of time series) and is thus equal to n . The central idea of principal component analysis is now to reduce the dimension of the problem on the basis of its statistical structure. The reduction is achieved by transforming the variables X_i into new uncorrelated variables Y_i appearing in

order of the size of their variance. This results, circumstances permitting, in a significant simplification of the subsequent analysis. It is in many cases possible to neglect statistical dependences of higher order in these new coordinates and to consider the time series of the new (transformed) variables as independent. As will be shown in the construction of the transformations, the new variables generally have different variances. In this case, it is possible to neglect those variables with relatively small variance.

This transformation of the random variables X_1, \dots, X_n into the new variables Y_1, \dots, Y_n is linear and thus we can write $Y_k = \sum_{i=1}^n \alpha_{ki} X_i$, or equivalently, in the vector notation used in Sect. 21.5

$$\mathbf{Y} = \boldsymbol{\alpha}\mathbf{X} \quad \text{mit} \quad \mathbf{X} = \begin{pmatrix} X_1 \\ \vdots \\ X_n \end{pmatrix}, \mathbf{Y} = \begin{pmatrix} Y_1 \\ \vdots \\ Y_n \end{pmatrix} \tag{34.1}$$

$$\boldsymbol{\alpha} = \begin{pmatrix} \alpha_{11} & \cdots & \alpha_{1n} \\ \vdots & \ddots & \vdots \\ \alpha_{n1} & \cdots & \alpha_{nn} \end{pmatrix} = \begin{pmatrix} (\boldsymbol{\alpha}^1)^T \\ \vdots \\ (\boldsymbol{\alpha}^n)^T \end{pmatrix} .$$

In the last step, the rows of the transformation matrix have been written in terms of the vectors $\boldsymbol{\alpha}$ defined as follows:

$$(\boldsymbol{\alpha}^k)^T := (\alpha_{k1} \cdots \alpha_{kn}) \implies \boldsymbol{\alpha}^k = \begin{pmatrix} \alpha_{k1} \\ \vdots \\ \alpha_{kn} \end{pmatrix} .$$

This implies that the components α_i^k of these vectors and the components α_{ki} of the matrix are related as follows:

$$\alpha_i^k = \alpha_{ki} = (\boldsymbol{\alpha}^k)^T_i \quad \forall k, i = 1, \dots, n .$$

The transformation Eq. 34.1 is thus given by

$$Y_k = (\boldsymbol{\alpha}^k)^T \mathbf{X} = \sum_{i=1}^n \alpha_{ki} X_i, \quad k = 1, \dots, n . \tag{34.2}$$

The coefficients of the matrix α are now chosen such that the transformed variables Y_k possess the following properties:

We require every “transformation vector” α_k (each row of the matrix α) to have a norm of 1, i.e.,

$$(\alpha^k)^T \alpha^k = \sum_{i=1}^n \alpha_i^k \alpha_i^k = \sum_{i=1}^n (\alpha_{ki})^2 \stackrel{!}{=} 1 \quad \forall k = 1, \dots, n. \quad (34.3)$$

We now select the components α_i^1 of α^1 in such a way that the variance of the first transformed variable $Y_1 = (\alpha^1)^T \mathbf{X}$ is as large as possible (maximal) while satisfying Eq. 34.3. This is an optimization problem subject to the constraint $(\alpha^1)^T \alpha^1 = 1$. It will be shown explicitly in the material below how problems of this type are treated.

Having determined α^1 (and consequently Y_1), we proceed by determining the components α_i^2 of α^2 such that the variance of the second transformed variable $Y_2 = (\alpha^2)^T \mathbf{X}$ is as large as possible (maximal), again subject to the condition that Eq. 34.3 holds. *In addition*, Y_2 must be uncorrelated with the vector Y_1 already determined. This additional condition can be expressed as

$$\text{cov}(Y_2, Y_1) \stackrel{!}{=} 0.$$

Using Eq. 34.2, we can write this covariance as

$$\begin{aligned} \text{cov}(Y_2, Y_1) &= \text{cov} \left((\alpha^2)^T \mathbf{X}, (\alpha^1)^T \mathbf{X} \right) \\ &= \text{cov} \left(\sum_{i=1}^n \alpha_i^2 X_i, \sum_{j=1}^n \alpha_j^1 X_j \right) \\ &= \sum_{i=1}^n \sum_{j=1}^n \alpha_{2i} \alpha_{1j} \underbrace{\text{cov}(X_i, X_j)}_{\delta \Sigma_{ij}} = \sum_{i=1}^n \sum_{j=1}^n (\alpha^2)_i^T \delta \Sigma_{ij} \alpha_j^1 \\ &= (\alpha^2)^T \delta \Sigma \alpha^1. \end{aligned}$$

This implies that, in addition to the constraint expressed in Eq. 34.3, α^2 must satisfy the condition

$$(\alpha^2)^T \delta \Sigma \alpha^1 \stackrel{!}{=} 0 \quad (34.4)$$

Thus, the determination of α^2 (and thus of Y_2) involves an optimization problem with *two* constraints.

The remaining variables Y_3, \dots, Y_n are determined successively subject to analogous conditions: in the k th step, α^k is chosen such that the variance of the k th transformed variable $Y_k = (\alpha^k)^T \mathbf{X}$ is as large as possible (maximal) subject to the condition that Eq. 34.3 as well as the *additional* $k - 1$ conditions hold, namely that Y_k be uncorrelated with *all* of the previously determined Y_i , $i = 1, \dots, k - 1$, i.e.,

$$\text{cov}(Y_k, Y_i) \stackrel{!}{=} 0 \quad \text{for all } i < k .$$

The formulation of these $k - 1$ conditions on α^i is analogous to Eq. 34.4

$$(\alpha^k)^T \delta \Sigma \alpha^i \stackrel{!}{=} 0 \quad \text{for all } i = 1, \dots, k - 1 . \tag{34.5}$$

Overall, k conditions must be satisfied in the k th step. This decreases the maximal possible variance of Y_k from step to step, since with each step, the maximum is taken over a smaller class of vectors. It is thus not surprising that Y_1 has the greatest variance among the Y_k and that the variances of the Y_k decrease rapidly with increasing k . Therefore, the first Y_k contribute most to the total variance of all Y_k . These transformed variables Y_k are therefore referred to as the *principal components* and the vectors α^k defined in the above construction as the *principal axes* of the system.

To provide the reader with a concrete example, we compute the first principal component Y_1 here. In order to do so, the variance of Y_1 conditional upon the satisfaction of Eq. 34.3 is maximized. According to Eq. A.12, the variance of Y_1 is

$$\begin{aligned} \text{var}(Y_1) &= \text{cov} \left((\alpha^1)^T \mathbf{X}, (\alpha^1)^T \mathbf{X} \right) = \text{cov} \left(\sum_{i=1}^n \alpha_i^1 X_i, \sum_{j=1}^n \alpha_j^1 X_j \right) \\ &= \sum_{i=1}^n \sum_{j=1}^n \alpha_{1i} \alpha_{1j} \underbrace{\text{cov}(X_i, X_j)}_{\delta \Sigma_{ij}} = \sum_{i=1}^n \sum_{j=1}^n (\alpha^1)_i^T \delta \Sigma_{ij} \alpha_j^1 \\ &= (\alpha^1)^T \delta \Sigma \alpha^1 . \end{aligned} \tag{34.6}$$

This is now maximized subject to the constraint $(\alpha^1)^T \alpha_1 = 1$. We can take constraints into account in an optimization problem using the method of *Lagrange multipliers*.¹ The method has already been discussed in Sect. 26.2.1. It requires the construction of the *Lagrange function* \mathcal{L} . This function is equal to the function to be maximized less a “zero” multiplied by a *Lagrange multiplier* λ . This zero is written in the form of the constraint to be satisfied, which in our case here is $0 = (\alpha^1)^T \alpha_1 - 1$. The Lagrange function is thus given by

$$\begin{aligned} \mathcal{L}(\lambda_1) &= \underbrace{(\alpha^1)^T \delta \Sigma \alpha^1}_{\text{zu maximieren}} - \lambda_1 \underbrace{\left[(\alpha^1)^T \alpha_1 - 1 \right]}_{\text{Nebenbedingung}} \quad (34.7) \\ &= \sum_{i,j=1}^n \alpha_{1i} \delta \Sigma_{ij} \alpha_{1j} - \lambda_1 \sum_{i=1}^n (\alpha_{1i})^2 + \lambda_1 . \end{aligned}$$

In order to find the optimal values α_{1i} subject to this constraint, we differentiate \mathcal{L} with respect to the parameter α_{1i} and set the resulting expression equal to zero; in other words, we locate the maximum of the Lagrange function.²

$$0 \stackrel{!}{=} \frac{\partial \mathcal{L}}{\partial \alpha_{1i}} = 2 \sum_j^n \delta \Sigma_{ij} \alpha_{1j} - 2 \lambda_1 \alpha_{1i} \quad \forall i = 1, \dots, n$$

from which it follows

$$\sum_j^n \delta \Sigma_{ij} \alpha_{1j} - \lambda_1 \alpha_{1i} = 0 \quad \forall i = 1, \dots, n$$

or in the compact matrix notation:

$$(\delta \Sigma - \lambda_1 \mathbf{1}) \alpha^1 = 0 , \quad (34.8)$$

where, as in Eq. 21.37, $\mathbf{1}$ denotes the n -dimensional identity matrix. This condition implies, however, that the covariance matrix applied to the vector α^1

¹An introduction in the technique of Lagrange multipliers for solving extreme value problems with boundary conditions can be found, for example, in [34].

²Since the Lagrange function differs only by *zero* from the value to be maximized (the variance), this equals the maximal value we are looking for. Of course, this is only true, if the difference is indeed equal to zero, i.e., if the boundary condition is fulfilled. This is the short form explanation of this method.

has no other effect than to multiply α^1 by a *number*: $\delta \Sigma \alpha^1 = \lambda_1 \alpha^1$. Equation 34.8 is therefore the *eigenvalue* equation of the covariance matrix (compare this to Eq. 22.22 in Sect. 22.3.2). The Lagrange multiplier λ_1 must in consequence be an eigenvalue of the covariance matrix and α^1 is the associated *eigenvector*. As discussed in detail in Sect. 22.3.2, the eigenvalue equation in 34.8 has a non-trivial solution $\alpha^1 \neq \mathbf{0}$ if and only if the matrix $(\delta \Sigma - \lambda_1 \mathbf{1})$ is singular. For this to be the case, its determinant must be equal to zero

$$\det(\delta \Sigma - \lambda_1 \mathbf{1}) = 0 .$$

The eigenvalue λ_1 is the solution to this determinant equation. Having determined λ_1 , it can be substituted into Eq. 34.8 to calculate the eigenvector α^1 .

The eigenvalue λ_1 has another important intuitive interpretation which can be seen immediately if we multiply both sides of Eq. 34.8 on the left by $(\alpha^1)^T$:

$$\begin{aligned} \lambda_1 \alpha^1 &= \delta \Sigma \alpha^1 \\ (\alpha^1)^T \lambda_1 \alpha^1 &= (\alpha^1)^T \delta \Sigma \alpha^1 \\ \lambda_1 \underbrace{(\alpha^1)^T \alpha^1}_1 &= \underbrace{(\alpha^1)^T \delta \Sigma \alpha^1}_{\text{Var}(Y_1)} . \end{aligned}$$

Therefore, λ_1 is the variance of the principal component Y_1 .

Analogously (i.e., maximization of the variances of the new variables Y_k), we proceed with the remaining principal components. As already mentioned above, one more constraint (see Eq. 34.5) must be taken into account with each further step. Analogous to Eq. 34.7, for each constraint a zero (in the form of the constraint multiplied by a Lagrange multipliers $\lambda_i, i = 1, \dots, k$) is subtracted from the variance of Y_k and the resulting Lagrange function \mathcal{L} is maximized. In the computation of the k th principal component, k Lagrange multipliers appear in the expression for the Lagrange function. The complexity of the corresponding computation increases substantially with each new constraint. We thus refrain from explicitly performing the computation here. We wish to remark, however, on the following properties holding for all principal components as was shown for Y_1 :

The Lagrange multipliers λ_k can be shown to be the eigenvalues of the covariance matrix $\delta \Sigma = \text{cov}(X, X)$ appearing in decreasing order. The row vectors of the transformation matrix α , i.e., the α^k , are the associated eigenvectors.

This result paves the way for the implementation of the principal component analysis: we do not take the optimization problem as the starting point for our analysis, but rather we compute the eigenvalues and eigenvectors of the covariance matrix directly. In a second step, the eigenvalues λ_k and their associated eigenvectors α^k are ordered according to the size of the eigenvalues, thereby constructing the transformation matrix from the ordered row vectors α^k as defined in Eq. 34.1:

$$\alpha = \begin{pmatrix} (\alpha^1)^T \\ \vdots \\ (\alpha^n)^T \end{pmatrix} .$$

Now, the original data X_i in the time series of the n risk factors are transformed into the new variables Y_i by applying the matrix α ; this is done for each observation in the time series

$$Y_j(t_k) = \sum_{i=1}^n \alpha_{ji} X_i(t_k) \quad \text{for all } j = 1, \dots, n \quad \text{and all } k = 1, \dots, T .$$

Finally, those components Y_k of the transformed data are neglected if the associated eigenvalues (and thus the variances³ of the new random variables Y_k) are small. It is often the case that the number of time series in the new variables Y_k necessary for further investigation can in this way be reduced to just two or three.

In many software packages, for example SAS, Matlab, or IDL, principal component analysis is provided as one of the few multivariate time series analysis procedures or at least modules exist allowing for its construction. The steps listed above are automated in these packages; only the original time series are needed as input. The covariance matrix is estimated from this data. The output is then obtained in the form of the eigenvalues of the covariance matrix, the associated eigenvectors, i.e., the transformation matrix, and the time series of the principal components. The reduction in dimension is achieved in that the data is reduced by $n - l - 1$ variables Y_{l+1}, \dots, Y_n . Here, l is chosen conditional upon the relative size of the eigenvalues.

³The k th eigenvalue is in fact equal to the variance of the new random variable Y_k .

34.2 Principal Component Analysis of the German Term Structure

In this section, we conduct a principal component analysis on real interest rate data. The aim of our investigation is to determine the typical drivers of the interest rate term structure and the relative contribution of each individual driver to the total dynamics of the interest rate curve. The results of this section are relevant in several different respects. Firstly, the investigation introduced here will provide a plausible explanation for why it is for many problems sufficient to model the entire interest rate curve in very low dimensional spaces as is done in modern one, two, and three-factor term structure models although the term structure is clearly constructed with a larger number of vertices. Secondly, *stress scenarios* for typical and statistically independent movements in the interest rate curve can be identified.

The time series investigated are the yields for ten vertices with terms between one and ten years for the German term structure in monthly time steps spanning over a period of ten years. The data were subjected to a principal component analysis as described in the above section *without* having first been pre-treated by taking time-differences.⁴ We find that the variances λ_i of the principal components Y_i , arranged in decreasing order, decrease *very* quickly. The proportion of the variances λ_1 and λ_2 of the first and second principal components with respect to the total variance $\sum_{i=1}^{10} \lambda_i$ is

$$\frac{\lambda_1}{\sum_{i=1}^{10} \lambda_i} \approx 96\% , \quad \frac{\lambda_2}{\sum_{i=1}^{10} \lambda_i} \approx 3\% ,$$

respectively. The proportion of the third and the fourth principal components is only

$$\frac{\lambda_3}{\sum_{i=1}^{10} \lambda_i} \approx 1\% , \quad \frac{\lambda_4}{\sum_{i=1}^{10} \lambda_i} \ll 1\% ,$$

⁴It has already been pointed out that principle component analysis assumes that the data in the time series are stationary. For the following investigation, this assumption is made keeping in mind that the results of the investigation should convey only a qualitative impression of term structure dynamics. Using the data directly (i.e. without a pre-treatment such as taking time differences, etc.) will simplify the *interpretation* of the results substantially.

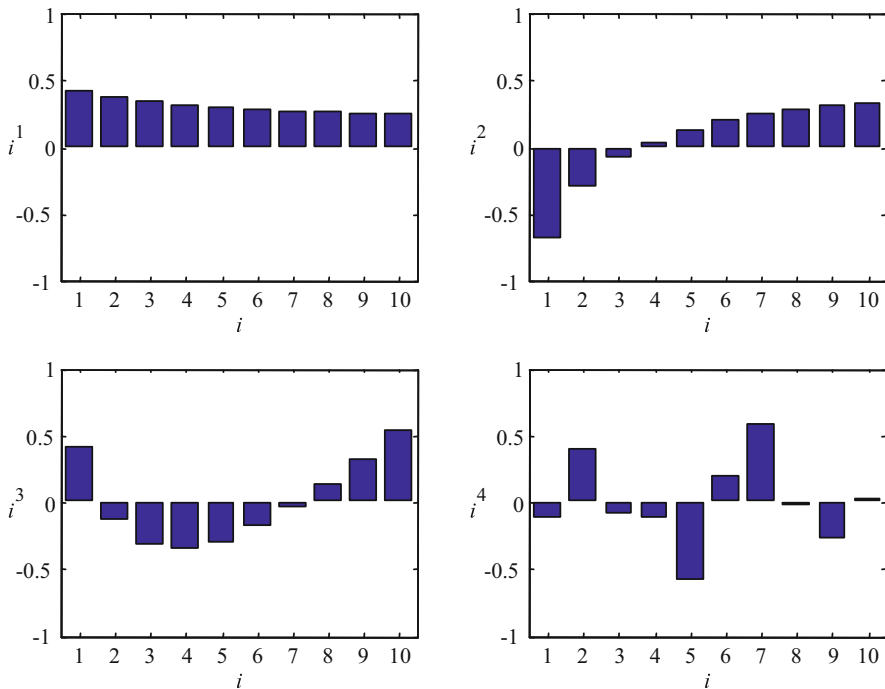


Fig. 34.1 The components of the first four eigenvectors $\alpha^1, \dots, \alpha^4$. For each of these eigenvectors the components $\alpha_1^i, \dots, \alpha_{10}^i$ are shown. The simple structure facilitates an intuitive interpretation of the associated principal components Y_1, \dots, Y_4

respectively. The variances of the further principal components disappear almost completely, lying well below one tenth of one percent. Obviously, the dynamics of the term structure can be described by just a very *few* variables. Indeed, this result provides an excellent motivation for modeling the term structure in spaces of small dimension (or even in *one*-dimensional spaces as in Chap. 14). In Fig. 34.1, the first four of ten eigenvectors are presented. These eigenvectors can be interpreted quite easily.

The first eigenvector α^1 weights the interest rates of all terms approximately equally. The eigenvector normalization allows the first principal component Y_1 to be interpreted as the mean interest rate level. The fluctuations of this principal component Y_1 thus represent *parallel shifts of the entire term structure*.

The second eigenvector α^2 weights the short term interest rates negatively and the long term rates positively. Considering the inverse transformation of the principal components to the original interest rate vectors, we can draw conclusions as to the interpretation of the second principal component Y_2 : adding the second component to the first has the effect of adjusting the mean

interest rate level by a mean slope. A change in this principal component Y_2 thus changes the mean *slope* of the term structure, or in other words, effects a *rotation of the term structure*.

The third eigenvector α^3 can be interpreted analogously. This vector weights the short and long-term rates positively, interest rates for the terms of intermediate length negatively. The addition of the associated principal component Y_3 to the first two thus effects a change in the mean *curvature of the term structure*. The fourth principal axis α^4 shows a periodic change in sign. With the associated principal component, periodic structures in the term structure can be represented such as those described in [159], for example.

Several practical conclusions for the analysis of scenarios commonly used in risk management can be drawn from the principal component analysis. Many of the common scenarios used to model a change in the term structure can be described in terms of the above decomposition. The most frequently used scenario is the *parallel shift*, which involves an increase or decrease in the entire term structure by a constant number of basis points. A further scenario, called the *twist*, involves a change in the slope of the term structure. This scenario is commonly realized through the addition (or subtraction) of, for example, m basis points to the interest rate corresponding to a term of m years; this is done for all terms in the term structure. Yet another scenario found in risk management is called *hump*. This scenario describes an increase in the short and long term rates and decrease in those for terms of intermediate length or vice versa. These three scenarios, the parallel shift, the twist and the hump, very often deduced on the basis of subjective experience, in fact correspond exactly to the first three principal components of the term structure. The scenarios mentioned here thus represent, from the statistical point of view, the most significant movements in the term structure. From the construction of the principal components, we can assume that these movements are approximately independent of one another and thus a simple representation of the dynamics of the term structure exists.

Finally, the analysis leads to the following results: first, practically all typical fluctuations in the term structure can be described by a combination of the above three scenarios. Second, making use of the eigenvalues, i.e., the variances of the principal components, confidence levels can be determined for the associated random variables. For example, information on the probability of the actual occurrence of one of these scenarios could be computed. We could proceed one step further: since the covariance matrix for the principal

components is diagonal, the value at risk can be differentiated with respect to movements in the individual principal components. The threat of VaR losses can be traced back to different types of movements in the term structure and interpreted accordingly. In addition, the values at risk from the three named stress scenarios can be taken to be uncorrelated and the value at risk can be computed by simply taking the sum of their squares.