Dimension-Reduction Techniques

Semih Alay

Seminar „Quantitative Risk Management“

1st June 2018

Table of contents

What is the aim of dimension-reduction techniques?

What is the aim of dimension-reduction techniques?

- Financial risk management is very complex and high-dimensional
- Complex problems have to be modelled
- Analyze an amount of data by various variables, such as a person's age, profession or health
- $\rightarrow$ Multivariate statistical analysis

Course of the talk

- Factor models
- $\rightarrow$ Explain randomness of random vector by using a smaller set of common factors
- Principal component analysis (PCA)
- $\rightarrow$ Data rotation technique to reduce dimensionality of highly correlated data by finding a small set of uncorrelated linear combinations

Table of contents

Introduction in factor models

- Explain randomness in the components of a random vector X by representing it by a smaller set of common factors
- Large part of variation of equity returns can be explained in terms of smaller set of market index returns
- Drag of a car $\rightarrow$ relevant parameters such as length, width and height can be combined to size of a car

Remember

Definition

- $E(X) = \int_{-\infty}^{\infty} xf(x)dx = \mu$ is called mean of X
- $\sigma^2 = \int_{-\infty}^{\infty} (x-\mu)^2 f(x)dx$ is called variance of X
- $Cov(X,Y) = E[(X - E(X)) \cdot (Y - E(Y))]$ is called covariance of X and Y, notice $Cov(X,Y) = Cov(Y,X)$
- $Cov(X) = \begin{pmatrix} Var(x_1) & Cov(x_1,x_2) & \dots & Cov(x_1,x_n) \\ Cov(x_2,x_1) & Var(x_2) & \dots & Cov(x_2,x_n) \\ \vdots & \vdots & \ddots & \vdots \\ Cov(x_n,x_1) & Cov(x_n,x_2) & \dots & Var(x_n) \end{pmatrix}$

p-factor model

Definition

- Random vector X is said to follow a p-factor model if it can be written as
  $X = a + BF + \epsilon$, where

p-factor model

Definition

- Random vector X is said to follow a p-factor model if it can be written as
  $X = a + BF + \epsilon$, where
- $F = (F_, \ldots, F_p)^T$ is a random vector of common factors with $p < d$ and a positive definite covariance matrix
- $\epsilon = (\epsilon_1, \ldots, \epsilon_p)^T$ is a random vector of idiosyncratic error terms, which are uncorrelated and have mean 0
- $B \in \mathbb{R}^{d \times p}$ is a matrix of constant factor loadings and $a \in \mathbb{R}^d$ is a vector of constants
- $cov(F, \epsilon) = 0$, which means the errors and the common factors are uncorrelated

In case of normal distribution

- Assume X to be multivariate normally distributed and follows the p-factor model
- Let Ω be the covariance matrix of F and γ the one of ϵ
- $\rightarrow \Sigma = Cov(X) = B\Omega B + \gamma$
- Define $F^* = \Omega^{-1/2}(F - E(F))$ and $B^* = B\Omega^{1/2}$

In case of normal distribution

- Factor model of the form $X = \mu + B^* F^* + \epsilon$, where $\mu = E(X)$ and $\Sigma = B^*(B^*)^T + \gamma$

Theorem

Whenever a random vector X has a covariance matrix Σ that satisfies

$$\Sigma = BB^T + \gamma,$$

where $B \in \mathbb{R}^{d \times p}$ with $\text{rank}(B) = p < d$ and γ is a diagonal matrix, then the random vector X has a factor model representation for some p-dimensional factor vector F and a d-dimensional error vector ϵ.

Three kinds of factor models

- Assume data $X_1, \ldots, X_n \in \mathbb{R}^d$ representing risk factor changes at different times $t = 1, \ldots, n$
- Each risk factor change X_t should follow a p-factor model
- $\rightarrow X_t = a + BF_t + \epsilon_t$, for $t = 1, \ldots, n$ with p-dimensional common factor vectors $F_t = (F_{t,1}, \ldots, F_{t,p})^T$, error vectors ϵ_t, a d-dimensional vector of constants a and loading matrix $B \in \mathbb{R}^{d \times p}$
- This model is an idealization where data would be explained perfectly by a factor model $\rightarrow$ seldomly the case in reality
- Therefore, find an approximating factor model that deals with the main sources of variability in the data

Macroeconomic factor models

- Assume that appropriate factors F_t are observable and we collect time-series data $F_1, \dots, F_n \in \mathbb{R}^p$
- Called macroeconomic because they are mostly used in finance and economics to observe macroeconomic factors as changes in inflation or interest rates

Application in Sharpe's single-index model

- F_t are observations of the return on a market index
- X_t are equity returns which are explained in terms of the market return
- B and a have to be determined by time-series regression techniques

Fundamental factor models

- Assume that the factors F_t are unobserved but the loading matrix B is given
- These models get their name by applications in modelling equity returns where stocks are classified according to their fundamental attributes, such as country or industry sector
- F_t have to be estimated from data X_t using cross-sectional regression at each time point t

Fundamental factor models

- If each risk factor change $X_{t,i}$ is identified with a unique value of a fundamental, e.g. a unique country, the loading Matrix B will only consist 0 and 1
- If one risk factor change $X_{t,i}$ depends to more than one country, e.g. a stock of a multinational company, then B can contain weighted values which sum up to 1
- Moreover there may be the case, where it is necessary to use time-dependent loading matrices B_t. This happens when the fundamental values change, e.g. the sales market of a company

Statistical factor models

- Most common model are statistical factor models
- Factors F_t and the loading matrix B are both not given
- We can estimate F_t and B from the data $X_1, \dots, X_n$ by using statistical techniques
- On the one hand, this approach can be very powerful in explaining the variability in data
- On the other hand, it is not secured that the observed data have an obvious interpretation

Factor finding in statistical factor models

- Classical statistical factor analysis, but is not used that commonly
- Principal component analysis (PCA) (next section)

Table of contents

Introduction in PCA

- Principal component analysis (PCA) is useful to reduce the dimensionality of highly correlated data
- Finding a small number of uncorrelated linear combinations that account for the most of the variance of the original data
- Example: Three factors F_1, F_2, F_3 and three variables length, width and pace and let the loading matrix be given by

Example

Example

factor	F_1	F_2	F_3
length(l)	0,862	0,481	-0,159
width(w)	0,977	0,083	0,198
pace(p)	-0,679	0,730	0,082

- For example, factor F_1 is given by the linear combination $F_1 = 0,862 \cdot l + 0,977 \cdot w - 0,679 \cdot p$
- Effects of the variables l and w are decisive, whereas the effect of p is small
- Influence of factor F_3 is not very huge and does not give many information about the data $\rightarrow$ eliminate

Spectral decompostion theorem of algebra

- Main mathematical result behind PCA
- Any symmetric matrix $A \in \mathbb{R}^{d \times d}$ (i.e. $A = A^T$) can be written as $A = \Gamma \Lambda \Gamma^T$
- Where $\Lambda = diag(\lambda_1, \ldots, \lambda_d)$ is the diagonal matrix consisting of eigenvalues of A and without loss of generality the eigenvalues λ_i are ordered decreasingly
- Γ is an orthogonal matrix (i.e. $\Gamma \Gamma^T = \Gamma^T \Gamma = I_d$), whose columns are standardized eigenvectors of A (i.e. eigenvectors with length 1)

Principal components transform

- Covariance matrix Σ of a random vector X is symmetric
- Positive semidefiniteness of Σ ensures that $\lambda_j \geq 0$ for all j
- Suppose the random vector X has the mean vector μ and covariance matrix Σ
- Apply spectral decomposition theorem to $\Sigma \rightarrow \Sigma = \Gamma \Lambda \Gamma^T$

Principal components transform and its properties

- The principal components transform of X is then given by $Y = \Gamma^T(X - \mu)$
- Interpreted as a rotation and recentring of X
- The jth component of the new vector is called jth principal component of X and is given by $Y_j = \gamma_j^T(X - \mu)$
- Where γ_j is the eigenvector of Σ corresponding to the jth ordered eigenvalue and is called also jth vector of loadings
- For the rotated vector Y the characteristics $E(Y) = 0$ and $Cov(Y) = \Gamma^T \Sigma \Gamma = \Gamma^T \Gamma \Lambda \Gamma^T \Gamma = \Lambda$ (thus Γ is orthogonal) hold

Maximizing of the variance

- Principal components of Y are uncorrelated and have variances $var(Y_j) = \lambda_j$ for all j
- Components are ordered by variance, from largest to smallest $\rightarrow$ maximizing of the variance
- The first principal component is the standardized linear combination of X, which has maximal variance among all such combinations
- $Var(\gamma_1^T X) = max[Var(a^T X)|a^T a = 1]$.

Maximizing of the variance

- For $j = 2, \ldots, d$, the jth principal component is the standardized linear combination of X with maximal variance among all of these linear combinations that are orthogonal to the first $j-1$ linear combinations
- The last principal component vector has minimum variance among standardized linear combinations of X
- Due to $\sum_{j=1}^{d} Var(Y_j) = \sum_{j=1}^{d} \lambda_j = trace(\Sigma) = \sum_{j=1}^{d} Var(X_j)$ we can use principal components to explain the variance of X
- $\sum_{j=1}^{k} \lambda_j / \sum_{j=1}^{d} \lambda_j$ represents the amount of the variance that is explained by the first k principal components

Example

- Applying PCA to our known example shows us the distribution of the total variance to the principal components
factor	eigenvalue λ_j	proportion on total variance	sum
F_1	2,16	71,97	71,97
F_2	0,77	25,67	97,64
F_3	0,07	2,36	100,00
- The proportion of factor F_3 on the total variance is very small
- Eliminate factor F_3 and only consider factors F_1 and F_2

Factor Modelling by PCA

- Use PCA for the construction of factors to use them in factor modelling
- By inverting our principal component transform we get $X = \mu + \Gamma Y = \mu + \Gamma_1 Y_1 + \Gamma_2 Y_2$
- Where Y is divided into two vectors $Y_1 \in \mathbb{R}^k$ and $Y_2 \in \mathbb{R}^{d-k}$
- Y_1 contains the first k principal components and Y_2 the further ones

Factor Modelling by PCA

- Γ is partitioned into two matrices $\Gamma_1 \in \mathbb{R}^{d \times k}$ and $\Gamma_2 \in \mathbb{R}^{d \times (d-k)}$
- By choosing k so that the first k principal components explain a large part of the total variance of X, we can focus on these k principal components and ignore the further ones from $k+1$ to d
- Set the error vector $\epsilon = \Gamma_2 Y_2$ and we obtain $X = \mu + \Gamma_1 Y_1 + \epsilon$
- In the form of the basic factor model, where Y_1 replaces the factors F and Γ_1 is replacing the loading matrix B

Factor estimation process

- Assume that we have a time series of multivariate data observation $X_1, \ldots, X_n$ with identical distribution, unknown mean vector μ and covariance matrix $\Sigma = \Gamma \Lambda \Gamma^T$
- To construct sample principal components we need to estimate the unknown parameters μ and Σ
- Estimate μ by the sample mean vector $\bar{X}$ and Σ by the sample covariance matrix $S_X = \frac{1}{n} \sum_{t=1}^{n} (X_t - \bar{X})(X_t - \bar{X})^T$
- Apply the spectral decompostion to get $S_X = GLG^T$

Definitions

- Define sample principal components as $Y_t = G^T(X_t - \bar{X})$ for $t = 1, \ldots, n$
- jth sample principal component as a component of Y_t as $Y_{t,j} = g_j^T(X_t - \bar{X})$ where g_j is the jth column of G, which is the eigenvector of S_X coresponding to the jth largest eigenvalue
- L is the sample covariance matrix of the rotated vectors $Y_1, \ldots, Y_n$
- The rotated vectors have no correlation between components and the components are ordered decreasingly by their sample variances

Conclusion

- Dimension-reduction techniques are very important
- Reduce the complexity of a problem by modelling it
- PCA powerful device to explain the variance of a random vector X and to decide which components are relevant and which not
- Lose some precison $\leftrightarrow$ reduce the dimension of a problem and thus its complexity
- Sometimes no obvious interpretation of data by PCA

Thank you for your attention!