# Computational Finance - Lesson 7

15/11/2024

ginevra.angelini@polimi.it

# Factor Models

# Factor Models

- *Factor models* are a fundamental concept in portfolio allocation theory. They provide a framework for understanding the sources of risk and return in a portfolio.

- A factor model posits that the returns of a portfolio (or individual assets) can be explained by a set of **systematic factors**, in addition to idiosyncratic or specific risks.

- In the context of finance and portfolio management, two main types of factor models are commonly used:

  1. ***Single-Factor Models:*** They assumes that a single systematic factor influences the returns of all assets in the portfolio (i.e. market returns in the CAPM model)

  2. ***Multi-Factors Models:*** They extend the single-factor model by considering multiple systematic factors that can affect asset returns

# Single - Factor Models

- In a *single-factor model*, the relationship between the return of an asset $r_a$ and the systematic factor $F$ can be represented by the following equation:

$$r_a = \alpha_a + \beta_a F + \varepsilon_a$$

Where:

- $\alpha_a$ represents the return not explained by the factor and is attributed to factors specific to the asset. It reflects the **excess return** attributable to factors beyond the chosen systematic factor.
- $\beta_a$ is the **sensitivity**, or factor loading, of asset *a* to the factor. It indicates how much the asset's return is expected to change for a given change in the factor.
- $F$ is the **value of the common factor**, which in a single-factor model is typically the return of the market.
- $\varepsilon_a$ represents the **idiosyncratic return of asset *a***, which is not explained by the factor $F$ and is not explained by any systematic factors considered in the model.

- The main **limitation** of single-factor models is that they assume that a single factor is sufficient to explain most of the variation in asset returns. In reality, there may be multiple factors influencing returns.

# Multi-Factors Models

- In a *multi-factors model*, the relationship between the return of an asset $r_a$ and multiple systematic factors $F_1, F_2, .., F_k$ can be represented by the following equation:

$$r_a = \alpha_a + \beta_{a,1} F_1 + \beta_{a,2} F_2 + .. + \beta_{a,k} F_k + \varepsilon_a$$

  Where $\beta_{a,1}$, $\beta_{a,2}$,.., $\beta_{a,k}$ are the **sensitivities**, or factor loadings, of asset $a$ to each of the k factors, $F_1, F_2, .., F_k$. They indicate how much the asset's return is expected to change for a given change in each respective factor.

- When constructing a multi-factor model, it is difficult to decide how many and which factors to include.

- Factors could represent **different dimensions of risk**, such as market risk, interest rate risk, economic conditions, industry-specific factors, etc. Common factors include market return, interest rates, inflation rates, and other macroeconomic variables.

- Multi-factor models also help *explain the weight of the different factors* used in the models, indicating which factor has more of an impact on the price of an asset.

# Multi-Factor Models

- Multi-factor models are used in portfolio construction for several important reasons:

  - **Capture Diverse Sources of Risk:** these models allow for the consideration of various systematic factors beyond just the market return. These factors can represent different sources of risk, allowing investors to have a more comprehensive understanding of the sources of risk in their portfolios.

  - **Diversification Benefits:** Different factors may exhibit low or negative correlations with each other. By diversifying across multiple factors, investors can potentially reduce overall portfolio risk.

  - **Better Performance Attribution:** It's easier to understand which factors are driving returns.

  - **Portfolio Optimization:** Incorporating multiple factors allows for more sophisticated portfolio optimization techniques.

  - **Adaptation to Market Conditions:** Different factors may dominate in different market environments.

# Multi-Factor Models

- Multi-factor models can be divided into three categories:

  - **Macroeconomic models:** Macroeconomic models compare a security's return to such factors as employment, inflation, and interest rates.

  - **Fundamental models:** Fundamental models analyze the relationship between a security's return and its underlying financials, such as earnings, market capitalization, and debt levels.

  - **Statistical models:** Statistical models posit that observed returns can be decomposed into common factors and idiosyncratic components. They don’t need exogenous data, they only need security returns as inputs.

# Macroeconomic Factor Models

- Macroeconomic factor models use observable economic time series as the factors. These factors include: *inflation, economic growth, interest rates, exchange rates*.

- Macroeconomic factor models assume that the random return of each security responds linearly to the macroeconomic shocks.

- The betas $\beta_i$ represent the sensitivities to the Macroeconomic factors. In particular,
  - the **sign of a beta** indicates the direction of the relationship between the asset's returns and the corresponding macroeconomic factor
  - the **magnitude of a beta** indicates the sensitivity of the asset's returns to changes in the macroeconomic factor. A higher magnitude (absolute value) implies a stronger sensitivity.

- Limitations of the models:
  - They are based on historical financial data.
  - They may not capture all relevant factors that influence asset values or returns.
  - Macroeconomic relationships may not always be linear and interactions between factors may be more complex than can be captured by a simple linear model
  - Correlation does not necessarily imply causation

# Fundamental Factor Models

- These models rely on information about a **company's financial health**, operations, industry position, and economic environment to assess the intrinsic value of an asset.

- Among the fundamental factors are book-value-to-price ratio, market capitalization, price-to-earnings ratio, and financial leverage, dividend yield, and industry classification.

- These models are based on economic theory and the belief that a company's financial performance and position are crucial in determining the value or return of its securities.

- Fundamental models are often used for **valuation purposes**, helping investors estimate the intrinsic value of a security. This information can be compared to the market price to make investment decisions.

- **Limitations** of the models:
  - They are based on historical financial data.
  - They may not capture all relevant factors that influence asset values or returns.
  - Relationships may not always be linear and interactions between factors may be more complex than can be captured by a simple linear model
  - Correlation does not necessarily imply causation

# Statistical Factor Models

- Statistical factor models use various maximum likelihood and principal-components-based factor analysis procedures on security returns to identify the pervasive factors in returns.

- There are two primary methods that are used to create statistical factor models:

    - ***Factor Models***, in which factors are portfolios that explain the *covariance* in asset returns

    - ***Principal Component Models***, in which factors are portfolios that explain the *variance* in asset returns

- The main **advantage** is that we only need data on the asset's prices. From the prices we compute the returns on which we estimate the statistical model. That is very different from macroeconomic and fundamental models that need a lot of macroeconomic and fundamental data.

- The main **limitation** is that an economic interpretation of the results is difficult in the sense that it is not possible to interpret the underlying drivers of the statistical factors

# Principal Component Analysis

- One of the most famous method to build a statistical factor model is the **Principal Component Analysis (PCA)**.

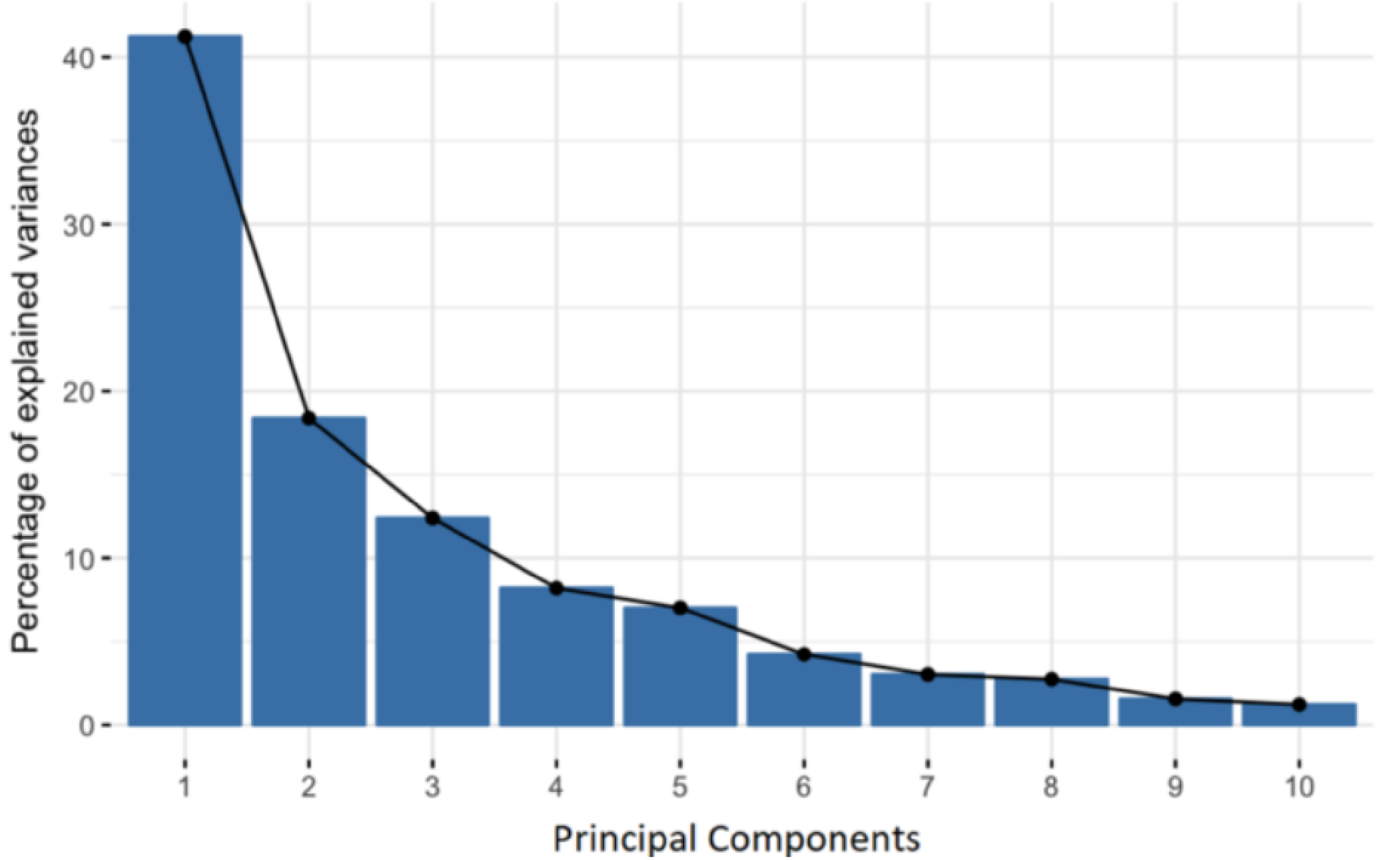
- PCA is a popular technique for analyzing large datasets containing a high number of features per observation that increase the interpretability of data while *preserving the maximum amount of information*.

- Formally PCA *is a dimensionality reduction method* that is often used to reduce the dimensionality of large data sets, by transforming a large set of variables into a smaller one that still contains **most of the information** in the large set.

- This is accomplished by **linearly transforming** the data into a new coordinate system where (most of) the variation in the data can be described with fewer dimensions than the initial data.

- In particular PCA is an **orthogonal linear transformation** that transforms the data to a new coordinate system (i.e. *set of orthogonal axes, called principal components*) such that the greatest variance by some scalar projection of the data comes to lie on the first coordinate (called the first principal component), the second greatest variance on the second coordinate, and so on.

# Principal Component Analysis




- In other words principal components are new variables that are constructed as **linear combinations** of the initial variables.

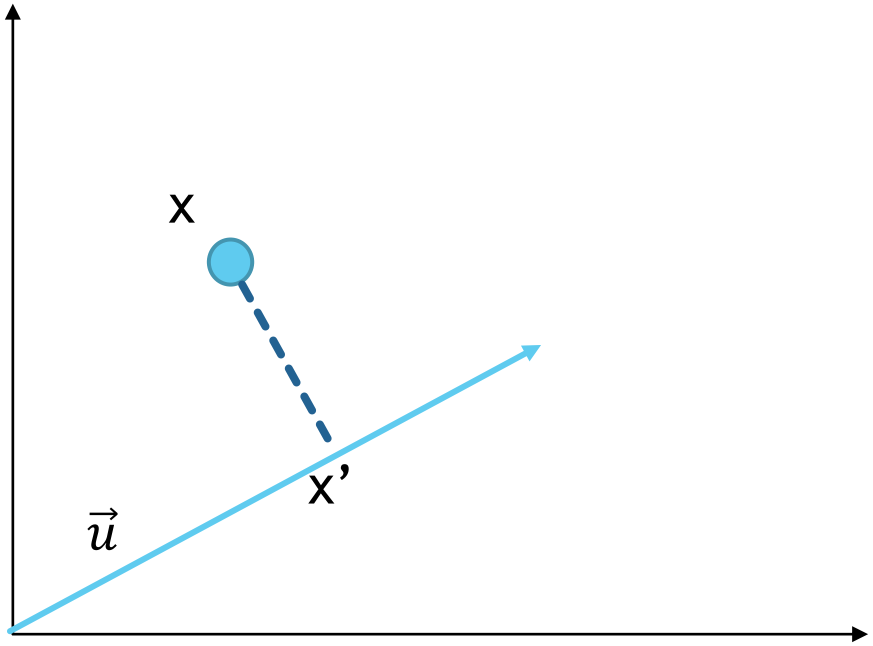
- These combinations are done in such a way that the new variables (i.e., **principal components**) are ***uncorrelated*** and most of the information within the initial variables is squeezed or compressed into the first components.

# Principal Component Analysis

- In order to pick the first principal component, we have to answer the following question: *" How we can arrange these points on a line that preserves as much information as possible?"*

- What information means in this context?

- When we project a point *x* on a unit vector *u*, we obtain x', such that $|x'| = (x^T u)$.




- Thus we can define the ***information preserved*** as the quantity $(x^T u)^2$, which is maximal when $x$ is parallel to $u$ and minimal when they are orthogonal.

# Principal Component Analysis

- Thus in order to **pick the first principal component**, i.e. find the component that maximizes the information, we have to solve the following problem:

$$\max \sum_i (x_i^T u)^2$$

Under the constraint, $u^T u = 1$

- We can simplify the objective function in this way:

$$\max \sum_i x_i^T u \; x_i^T u$$

$$\max \sum_i u^T x_i \, x_i^T u$$

# Principal Component Analysis

- If the data, *x*, are standardized, then the quantity $\sum_i x_i x_i^T$ times *n*, is the ***covariance matrix*** of the data. Thus the optimization problem becomes:

$$\max u^T C\, u$$

- To solve this problem we can use the lagrangian multipliers method. Thus:

$$f(u, \lambda) = u^T C\, u - \lambda(u^T u - 1)$$

$$\frac{\partial f(u, \lambda)}{\partial u} = 2Cu - 2\lambda u = 0$$

$$\boxed{Cu = \lambda u} \rightarrow$$

Equation of the **eigenvectors** and **eigenvalues** of C

# Principal Component Analysis

- If we multiply both sides for $u^T$

$$u^T C u = \lambda u^T u$$

$$u^T C u = \lambda$$

- This means *that the amount of information preserved after projection on a eigenvector is given by the corresponding eigenvalue*, so the best direction to pick is the eigenvector with the largest eigenvalue.

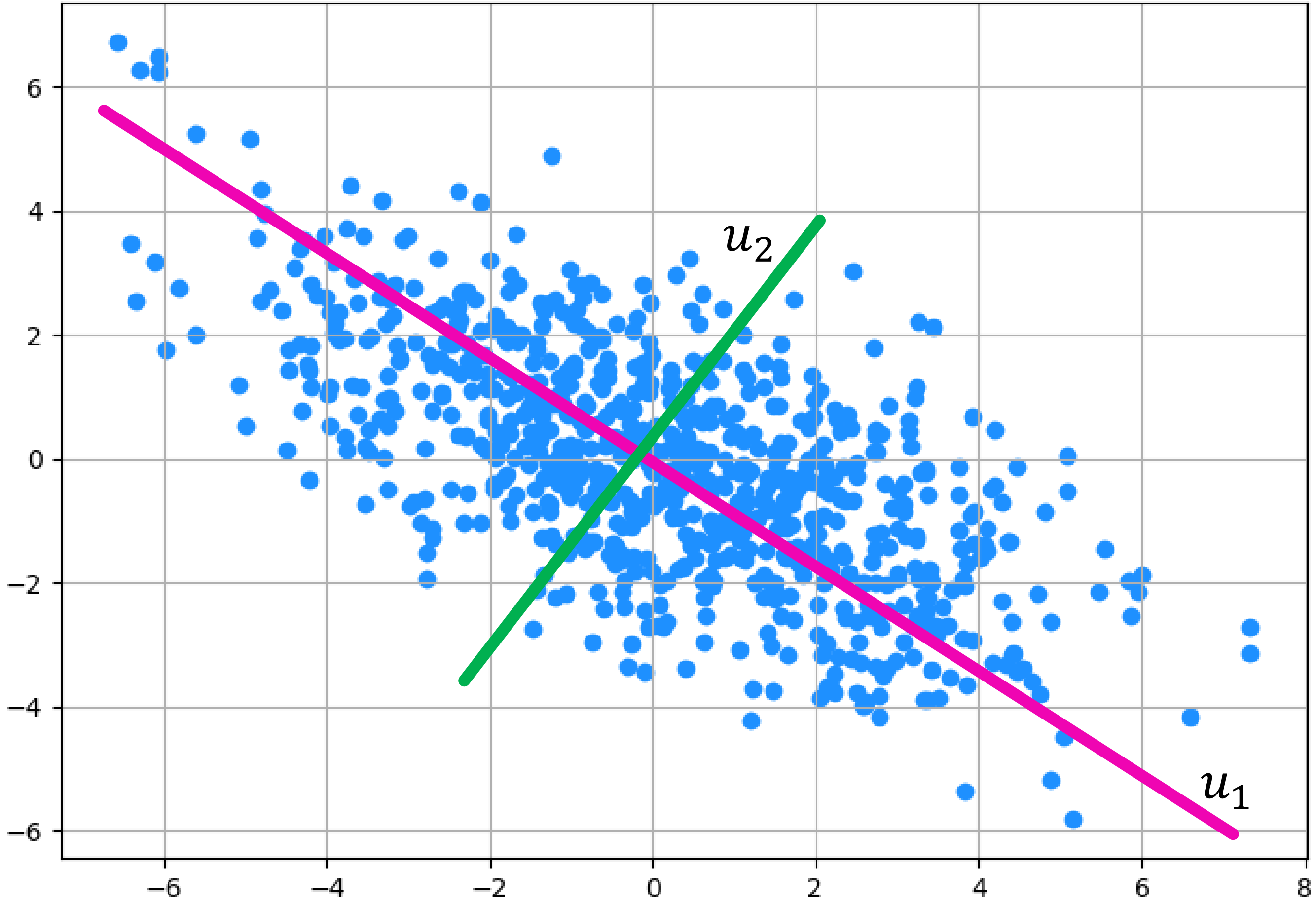
- Then in order to pick the second principal component, $u_2$, we have to add a *second constraint* such that $u_2 \perp u_1$ , meaning that the *second component is orthogonal to the first one* (contains information not contained in the first one).

- Thus the optimization problem, under the two constraints $u_2 \perp u_1$ and $u_2^T u_2 = 1$, is

$$\max u^T C\, u$$

- Following the same resolution process of the first component, we know that the second component is given by the second eigenvector of C.

# Principal Component Analysis




- The first principal component captures the most variation in the data
- The second principal component captures the maximum variance that is orthogonal to the first principal component
- This continues until a total of $p$ principal components have been calculated, equal to the original number of variables

# Principal Component Analysis

- Step-by-step explanation of PCA:

1. **Standardize the data:** the aim of this step is to standardize the range of the continuous initial variables so that each one of them contributes equally to the analysis.

$$X' = \frac{X - \mu}{\sigma}$$

   Standardization is necessary because PCA is quite sensitive regarding the variances of the initial variables (if there are large differences between the ranges of initial variables, those variables with larger ranges will dominate over those with small ranges) thus leading to biased results.

2. **Compute the Covariance Matrix:** the aim of this step is to understand how the variables of the input data set are varying from the mean with respect to each other, or in other words, to see if there is any relationship between them.

# Principal Component Analysis

The correlation matrix is computed as:

$$C' = \frac{1}{n} X'^{T} X'$$

3. **Eigenvectors and Eigenvalues of the Covariance Matrix:** Once you have the covariance matrix C you can compute its eigenvectors $(u_1, u_2, .., u_m)$ and eigenvalues $(\lambda_1, \lambda_2, .., \lambda_m)$. These eigenvectors represent the principal components and the eigenvalues indicate the amount of variance explained by each component, i.e it's importance.

4. **Sort Eigenvectors by Eigenvalues:** Arrange the eigenvectors in descending order of their corresponding eigenvalues. This way, the most significant components (those explaining the most variance) come first.

5. **Select Principal Components:** Decide how many principal components to retain based on how much variance you want to preserve. Typically, you might choose a number of components that cumulatively explain a high percentage (e.g., 95%) of the total variance.

6. **Project Data onto the New Feature Space:** Multiply the original data by the selected eigenvectors (principal components) to obtain the new feature space.

# Factor Models in Portfolio Optimization

- With a factor model, N asset returns can be expressed as a linear combination of k factor returns where $k \ll N$. Let $r_a, \mu_a$, and $\varepsilon_a$ be, respectively, the asset return, the mean of asset return and the idiosyncratic return related to each asset. It is known that:

$$r_a = \mu_a + Fr_f + \varepsilon_a$$

Where $r_f$ is the factor return and $F$ is the matrix of factor loadings (NxK).

- In the mean-variance framework, portfolio risk is then:

$$\begin{aligned}\sigma_P &= var(r_a^T w_a) = var\left((\mu_a + Fr_f + \varepsilon_a)^T w_a\right) \\ &= w_a^T\left(F\Sigma_f F + D\right)w_a \\ &= w_f^T \Sigma_f w_f + w_a^T D w_a\end{aligned}$$

Where $\Sigma_f$ is the **covariance of factor returns**, D is the **variance of idiosyncratic returns** and $\mu_f = F^T \mu_a$

# Factor Models in Portfolio Optimization

- Therefore, the mean-variance optimization problem is formulated as:

$$\max(\mu_a^T w_a \ - \sigma_P^2)$$

Under the following constraints:

- $0 \leq w_i \leq 1$
- $\sum w_i \ = 1$
- $\mu_f \ = \ F^T \mu_a$

# PCA in Portfolio Optimization: Pros & Cons

## Pros

- Dimensionality Reduction
  - In large portfolios (e.g., with hundreds of assets), PCA simplifies the data structure without losing too much information
- Improved Portfolio Diversification
  - Better diversification across independent risk factors, potentially enhancing risk-adjusted returns and reducing exposure to specific risks
- Mitigates Multicollinearity
  - In financial data, assets are often highly correlated, which can make the covariance matrix estimation unstable
- Improved Portfolio Adaptability to Changing Market Structures
  - PCA can reveal shifts in market dynamics as it responds to changes in the covariance structure

## Cons

- PCA is a Linear Method
  - Financial markets often have nonlinear interactions between assets
- Sensitivity to Market Conditions
  - The structure of market risks may change often. Frequent re-computation may be required, leading to high turnover and transaction costs
- Interpretation of Principal Components
  - Without clear economic meaning, it can be challenging for portfolio managers to interpret the sources of risk in terms of fundamental factors
- Instability in Low Sample Sizes
  - When applied to portfolios with limited historical data, PCA might produce spurious results due to overfitting