# Factor Models for Asset Returns

Prof. Daniel P. Palomar

The Hong Kong University of Science and Technology (HKUST)

MAFS6010R- Portfolio Optimization with R
MSc in Financial Mathematics
Fall 2018-19, HKUST, Hong Kong

# Outline

# Outline

# Why Factor Models?

- To decompose risk and return into explainable and unexplainable components.
- Generate estimates of abnormal returns.
- Describe the covariance structure of returns
  - without factors: the covariance matrix for $N$ stocks requires $N(N+1)/2$ parameters (e.g., $500(500+1)/2 = 125,250$)
  - with $K$ factors: the covariance matrix for $N$ stocks requires $N(K+1)$ parameters with $K \ll N$ (e.g., $500(3+1) = 2,000$)
- Predict returns in specified stress scenarios.
- Provide a framework for portfolio risk analysis.

# Types of Factor Models

- Factor models decompose the asset returns into two parts: low-dimensional factors and idiosyncratic residual noise.[1]
- Three types:
  1. Macroeconomic factor models
     - factors are observable economic and financial time series
     - no systematic approach to choose factors
  2. Fundamental factor models
     - factors are created from observable asset characteristics
     - no systematic approach to define the characteristics
  3. Statistical factor models
     - factors are unobservable and extracted from asset returns
     - more systematic, but factors have no clear interpretation

[1] R. S. Tsay, *Analysis of Financial Time Series*. John Wiley & Sons, 2005.

# Outline

# Linear Factor Model

**Data:**

- $N$ assets/instruments/indexes: $i = 1, \ldots, N$
- $T$ time periods: $t = 1, \ldots, T$
- $N$-variate random vector for returns at $t$: $\mathbf{x}_t = (x_{1,t}, \ldots, x_{N,t})^T$.

**Factor model for asset $i$:**

$$x_{i,t} = \alpha_i + \beta_{1,i} f_{1,t} + \cdots + \beta_{K,i} f_{K,t} + \epsilon_{i,t}, \quad t = 1, \ldots, T.$$

- $K$: the number of factors
- $\alpha_i$: intercept of asset $i$
- $\mathbf{f}_t = (f_{1,t}, \ldots, f_{K,t})^T$: common factors (same for all assets $i$)
- $\boldsymbol{\beta}_i = (\beta_{1,i}, \ldots, \beta_{K,i})^T$: factor loading of asset $i$ (independing ot $t$)
- $\epsilon_{i,t}$: residual idiosyncratic term for asset $i$ at time $t$

# Cross-Sectional Regression

**Factor model:**

$$\mathbf{x}_t = \boldsymbol{\alpha} + \mathbf{B}\mathbf{f}_t + \boldsymbol{\epsilon}_t, \quad t = 1, \ldots, T$$

where $\boldsymbol{\alpha} = \begin{bmatrix} \alpha_1 \\ \vdots \\ \alpha_N \end{bmatrix}$ $(N \times 1)$, $\mathbf{B} = \begin{bmatrix} \boldsymbol{\beta}_1^T \\ \vdots \\ \boldsymbol{\beta}_N^T \end{bmatrix}$ $(N \times K)$, $\boldsymbol{\epsilon}_t = \begin{bmatrix} \epsilon_{1,t} \\ \vdots \\ \epsilon_{N,t} \end{bmatrix}$ $(N \times 1)$.

- $\boldsymbol{\alpha}$ and $\mathbf{B}$ are independent of $t$
- the factors $\{\mathbf{f}_t\}$ ($K$-variate) are stationary with

$$E[\mathbf{f}_t] = \boldsymbol{\mu}_f$$
$$Cov[\mathbf{f}_t] = E[(\mathbf{f}_t - \boldsymbol{\mu}_f)(\mathbf{f}_t - \boldsymbol{\mu}_f)^T] = \boldsymbol{\Sigma}_f$$

- the residuals $\{\boldsymbol{\epsilon}_t\}$ ($N$-variate) are white noise with

$$E[\boldsymbol{\epsilon}_t] = \mathbf{0}$$
$$Cov[\boldsymbol{\epsilon}_t, \boldsymbol{\epsilon}_s] = E[\boldsymbol{\epsilon}_t \boldsymbol{\epsilon}_s^T] = \boldsymbol{\Psi}\delta_{ts}, \quad \boldsymbol{\Psi} = \text{diag}\left(\sigma_1^2, \ldots, \sigma_N^2\right)$$

- the two processes $\{\mathbf{f}_t\}$ and $\{\boldsymbol{\epsilon}_t\}$ are uncorrelated

# Linear Factor Model

**Summary of parameters:**

- $\boldsymbol{\alpha}$: $(N \times 1)$ intercept for $N$ assets
- $\mathbf{B}$: $(N \times K)$ factor loading matrix
- $\boldsymbol{\mu}_f$: $(K \times 1)$ mean vector of $K$ common factors
- $\boldsymbol{\Sigma}_f$: $(K \times K)$ covariance matrix of $K$ common factors
- $\boldsymbol{\Psi} = \text{diag}\left(\sigma_1^2, \ldots, \sigma_N^2\right)$: $N$ asset-specific variances

**Properties of linear factor model:**

- the stochastic process $\{\mathbf{x}_t\}$ is a stationary multivariate time series with
- conditional moments

$$E\left[\mathbf{x}_t \mid \mathbf{f}_t\right] = \boldsymbol{\alpha} + \mathbf{B}\mathbf{f}_t$$
$$Cov\left[\mathbf{x}_t \mid \mathbf{f}_t\right] = \boldsymbol{\Psi}$$

- unconditional moments

$$E\left[\mathbf{x}_t\right] = \boldsymbol{\alpha} + \mathbf{B}\boldsymbol{\mu}_f$$
$$Cov\left[\mathbf{x}_t\right] = \mathbf{B}\boldsymbol{\Sigma}_f\mathbf{B}^T + \boldsymbol{\Psi}.$$

# Time-Series Regression

**Factor model:**

$$\mathbf{x}_i = \mathbf{1}_T \alpha_i + \mathbf{F}\boldsymbol{\beta}_i + \boldsymbol{\epsilon}_i, \quad i = 1, \dots, N$$

where $\mathbf{x}_i = \begin{bmatrix} x_{i,1} \\ \vdots \\ x_{i,T} \end{bmatrix}$ $(T \times 1)$, $\mathbf{1}_T = \begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix}$ $(T \times 1)$, $\mathbf{F} = \begin{bmatrix} \mathbf{f}_1^T \\ \vdots \\ \mathbf{f}_T^T \end{bmatrix}$ $(T \times K)$, $\boldsymbol{\epsilon}_i = \begin{bmatrix} \epsilon_{i,1} \\ \vdots \\ \epsilon_{i,T} \end{bmatrix}$ $(T \times 1)$.

- $\boldsymbol{\epsilon}_i$ is the ($T$-variate) vector of white noise with

$$E\left[\boldsymbol{\epsilon}_i\right] = \mathbf{0}$$
$$Cov\left[\boldsymbol{\epsilon}_i\right] = \sigma_i^2 \mathbf{I}_T.$$

# Multivariate Regression

**Factor model (using compact matrix notation):**

$$\mathbf{X}^T = \boldsymbol{\alpha}\mathbf{1}^T + \mathbf{B}\mathbf{F}^T + \mathbf{E}^T$$

where $\mathbf{X} = \begin{bmatrix} \mathbf{x}_1^T \\ \vdots \\ \mathbf{x}_T^T \end{bmatrix}$ $(T \times N)$, $\mathbf{F} = \begin{bmatrix} \mathbf{f}_1^T \\ \vdots \\ \mathbf{f}_T^T \end{bmatrix}$ $(T \times K)$, $\mathbf{E} = \begin{bmatrix} \boldsymbol{\epsilon}_1^T \\ \vdots \\ \boldsymbol{\epsilon}_T^T \end{bmatrix}$ $(T \times N)$.

# Expected Return ($\alpha$–$\beta$) Decomposition

Recall the factor model for asset $i$:

$$x_{i,t} = \alpha_i + \boldsymbol{\beta}_i^T \mathbf{f}_t + \epsilon_{i,t}, \;\; t = 1, \ldots, T.$$

Take the expectation:

$$E[x_{i,t}] = \alpha_i + \boldsymbol{\beta}_i^T \boldsymbol{\mu}_f$$

where

- $\boldsymbol{\beta}_i^T \boldsymbol{\mu}_f$ is the explained expected return due to systematic risk factors
- $\alpha_i = E[x_{i,t}] - \boldsymbol{\beta}_i^T \boldsymbol{\mu}_f$ is the unexplained exptected return (abnormal return).

According to the CAPM model, the alpha should be zero, but is it?

# Portfolio Analysis

Let $\mathbf{w} = (w_1, \ldots, w_N)^T$ be a vector of portfolio weights ($w_i$ is the fraction of wealth in asset $i$). The portfolio return is

$$r_{p,t} = \mathbf{w}^T \mathbf{r}_t = \sum_{i=1}^{N} w_i r_{i,t}, \quad t = 1, \ldots, T.$$

**Portfolio factor model:**

$$\mathbf{r}_t = \boldsymbol{\alpha} + \mathbf{B}\mathbf{f}_t + \boldsymbol{\epsilon}_t \qquad \Rightarrow$$
$$r_{p,t} = \mathbf{w}^T \boldsymbol{\alpha} + \mathbf{w}^T \mathbf{B}\mathbf{f}_t + \mathbf{w}^T \boldsymbol{\epsilon}_t = \alpha_p + \boldsymbol{\beta}_p^T \mathbf{f}_t + \epsilon_{p,t}$$

where

$$\alpha_p = \mathbf{w}^T \boldsymbol{\alpha}$$
$$\boldsymbol{\beta}_p^T = \mathbf{w}^T \mathbf{B}$$
$$\epsilon_{p,t} = \mathbf{w}^T \boldsymbol{\epsilon}_t$$

and

$$var(r_{p,t}) = \mathbf{w}^T \left( \mathbf{B} \boldsymbol{\Sigma}_f \mathbf{B}^T + \boldsymbol{\Psi} \right) \mathbf{w}.$$

# Outline

# Macroeconomic Factor Models

Recall the factor model for asset $i$:

$$x_{i,t} = \alpha_i + \boldsymbol{\beta}_i^T \mathbf{f}_t + \epsilon_{i,t}, \quad t = 1, \ldots, T.$$

In this model, the factors $\{\mathbf{f}_t\}$ are observed economic/financial time series.

Econometric problems:

- choice of factors
- estimation of mean vector and covariance matrix of factors $\boldsymbol{\mu}_f$ and $\boldsymbol{\Sigma}_f$ from observed history of factors
- estimation of factor betas $\boldsymbol{\beta}_i$'s and residual variances $\sigma_i^2$'s using time series regression techniques

**Single factor model of Sharpe (1964) (aka CAPM):**

$$x_{i,t} = \alpha_i + \beta_i R_{M,t} + \epsilon_{i,t}, \quad t = 1, \ldots, T$$

where

- $R_{M,t}$ is the return of the market (in excess of the risk-free asset rate): **market risk factor** (typically a value weighted index like the S&P 500)
- $x_{i,t}$ is the return of asset $i$ (in excess of the risk-free rate)
- $K = 1$ and the single factor is $f_{1,t} = R_{M,t}$
- the unconditional cross-sectional covariance matrix of the assets is

$$Cov\left[\mathbf{x}_t\right] = \mathbf{\Sigma} = \sigma_M^2 \boldsymbol{\beta}\boldsymbol{\beta}^T + \mathbf{\Psi}$$

  where
  - $\sigma_M^2 = var\left(R_{M,t}\right)$
  - $\boldsymbol{\beta} = \left(\beta_1, \ldots, \beta_N\right)^T$
  - $\mathbf{\Psi} = \text{diag}\left(\sigma_1^2, \ldots, \sigma_N^2\right)$

# Macroeconomic Factor Models

**Estimation of single factor model:**

$$\mathbf{x}_i = \mathbf{1}_T \hat{\alpha}_i + \hat{\beta}_i \mathbf{r}_M + \hat{\boldsymbol{\epsilon}}_i, \quad i = 1, \ldots, N$$

where $\mathbf{r}_M = (R_{M,1}, \ldots, R_{M,T})$ with estimates:

- $\hat{\beta}_i = \widehat{cov}\left(x_{i,t}, R_{M,t}\right) / \widehat{var}\left(R_{M,t}\right)$
- $\hat{\alpha}_i = \bar{x}_i - \hat{\beta}_i \bar{r}_M$
- $\hat{\sigma}_i^2 = \frac{1}{T-2} \hat{\boldsymbol{\epsilon}}_i^T \hat{\boldsymbol{\epsilon}}_i$, $\hat{\boldsymbol{\Psi}} = \mathrm{diag}\left(\hat{\sigma}_1^2, \ldots, \hat{\sigma}_N^2\right)$

The estimated single factor model covariance matrix is

$$\hat{\boldsymbol{\Sigma}} = \hat{\sigma}_M^2 \hat{\boldsymbol{\beta}} \hat{\boldsymbol{\beta}}^T + \hat{\boldsymbol{\Psi}}$$

# Macroeconomic Factor Models: Market Neutrality

Recall the single factor model:

$$\mathbf{x}_t = \boldsymbol{\alpha} + \boldsymbol{\beta} R_{M,t} + \boldsymbol{\epsilon}_t, \quad t = 1, \ldots, T$$

When designing a portfolio $\mathbf{w}$, it is common to have a market-neutral constraint:

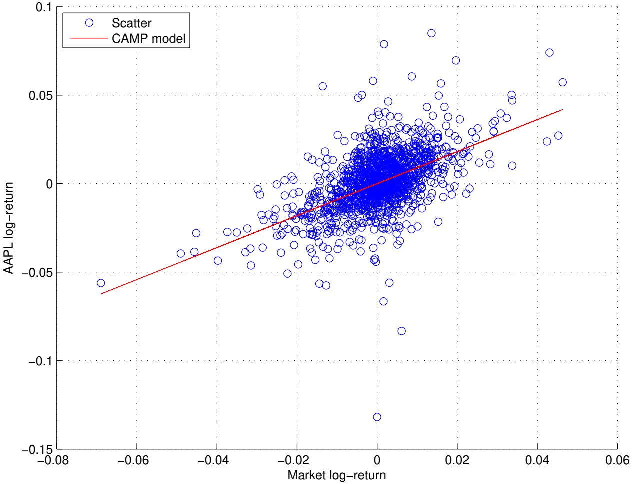
$$\boldsymbol{\beta}^T \mathbf{w} = 0.$$

This is to avoid exposure to the market risk. The resulting risk is given by

$$\mathbf{w}^T \boldsymbol{\Psi} \mathbf{w}.$$

# Capital Asset Pricing Model: AAPL vs SP500

- AAPL regressed against the SP500 index (using risk-free rate $r_f = 2\%/252$):



# $R^2$

The $R^2$ from the time series regression is a measure of the proportion of "market" risk and $1 - R^2$ is a measure of asset specific risk ($\sigma_i^2$ is a measure of the typical size of asset specific risk).

Given the variance decomposition:

$$var\left(x_{i,t}\right) = \beta_i^2 var\left(R_{M,t}\right) + var\left(\epsilon_{i,t}\right) = \beta_i^2 \sigma_M^2 + \sigma_i^2$$

$R^2$ can be estimated as

$$R^2 = \frac{\hat{\beta}_i^2 \hat{\sigma}_M^2}{\widehat{var}\left(x_{i,t}\right)}$$

Extensions:

- robust regression techniques to estimate $\beta_i$, $\sigma_i^2$, and $\sigma_M^2$
- factor model not constant over time, obtaining time-varying $\beta_{i,t}$, $\sigma_{i,t}^2$, and $\sigma_{M,t}^2$
  - $\beta_{i,t}$ can be estimated via rolling regression or Kalman filter techniques
  - $\sigma_{i,t}^2$ and $\sigma_{M,t}^2$ (i.e., conditional heteroskedasticity) can be captured via GARCH models or exponential weights

# Macroeconomic Factor Models

**General multifactor model:**

$$x_{i,t} = \alpha_i + \boldsymbol{\beta}_i^T \mathbf{f}_t + \epsilon_{i,t}, \quad t = 1, \ldots, T.$$

where the factors $\{\mathbf{f}_t\}$ represent macro-economic variables such as[2]

- market risk
- price indices (CPI, PPI, commodities) / inflation
- industrial production (GDP)
- money growth
- interest rates
- housing starts
- unemployment

In practice, there are many factors and in most cases they are very expensive to obtain. Typically, investment funds have to pay substantial subscription fees to have access to them (not available to small investors).

[2] Chen, Ross, Roll (1986). "Economic Forces and the Stock Market"

**Estimation of multifactor model ($K > 1$):**

$$\begin{aligned} \mathbf{x}_i &= \mathbf{1}_T \alpha_i + \mathbf{F}\boldsymbol{\beta}_i + \boldsymbol{\epsilon}_i, \quad i = 1, \ldots, N \\ &= \tilde{\mathbf{F}}\boldsymbol{\gamma}_i + \boldsymbol{\epsilon}_i \end{aligned}$$

where $\tilde{\mathbf{F}} = \begin{bmatrix} \mathbf{1}_T & \mathbf{F} \end{bmatrix}$ and $\boldsymbol{\gamma}_i = \begin{bmatrix} \alpha_i \\ \boldsymbol{\beta}_i \end{bmatrix}$.

Estimates:

- $\hat{\boldsymbol{\gamma}}_i = (\tilde{\mathbf{F}}^T \tilde{\mathbf{F}})^{-1} \tilde{\mathbf{F}}^T \mathbf{x}_i$, $\hat{\mathbf{B}} = \begin{bmatrix} \hat{\boldsymbol{\beta}}_1 & \cdots & \hat{\boldsymbol{\beta}}_N \end{bmatrix}^T$
- $\hat{\boldsymbol{\epsilon}}_i = \mathbf{x}_i - \tilde{\mathbf{F}}\hat{\boldsymbol{\gamma}}_i$
- $\hat{\sigma}_i^2 = \frac{1}{T-K-1} \hat{\boldsymbol{\epsilon}}_i^T \hat{\boldsymbol{\epsilon}}_i$, $\hat{\boldsymbol{\Psi}} = \text{diag}\left(\hat{\sigma}_1^2, \ldots, \hat{\sigma}_N^2\right)$
- $\hat{\boldsymbol{\Sigma}}_f = \frac{1}{T-1} \sum_{t=1}^{T} \left(\mathbf{f}_t - \bar{\mathbf{f}}\right)\left(\mathbf{f}_t - \bar{\mathbf{f}}\right)^T$, $\bar{\mathbf{f}} = \frac{1}{T} \sum_{t=1}^{T} \mathbf{f}_t$

The estimated multifactor model covariance matrix is

$$\hat{\boldsymbol{\Sigma}} = \hat{\mathbf{B}} \hat{\boldsymbol{\Sigma}}_f \hat{\mathbf{B}}^T + \hat{\boldsymbol{\Psi}}$$

# Interlude: LS Regression

Consider the LS regression:

$$\underset{\gamma}{\text{minimize}} \quad \|\mathbf{x} - \tilde{\mathbf{F}}\gamma\|^2$$

We set the gradient to zero

$$2\tilde{\mathbf{F}}^T\tilde{\mathbf{F}}\gamma - 2\tilde{\mathbf{F}}^T\mathbf{x} = \mathbf{0}$$

to find the estimator as the optimal solution

$$\hat{\gamma} = (\tilde{\mathbf{F}}^T\tilde{\mathbf{F}})^{-1}\tilde{\mathbf{F}}^T\mathbf{x}.$$

Now we can compute the residual of the LS regression as $\hat{\epsilon} = \mathbf{x} - \tilde{\mathbf{F}}\hat{\gamma}$ and its variance with the sample estimator

$$\hat{\sigma}^2 = \frac{1}{T}\|\hat{\epsilon}\|^2 = \frac{1}{T}\|\mathbf{x} - \tilde{\mathbf{F}}\hat{\gamma}\|^2.$$

# Interlude: Maximum Likelihood Estimation

The pdf for the residual under a Gaussian distribution is

$$p\left(\epsilon_t\right) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2\sigma^2}\epsilon_t^2}$$

and the log-likelihood of the parameters $\left(\boldsymbol{\gamma}, \sigma^2\right)$ given the $T$ uncorrelated observations in $\boldsymbol{\epsilon}$ is

$$L\left(\boldsymbol{\gamma}, \sigma^2; \boldsymbol{\epsilon}\right) = -\frac{T}{2}\log\left(2\pi\sigma^2\right) - \frac{1}{2\sigma^2}\|\mathbf{x} - \tilde{\mathbf{F}}\boldsymbol{\gamma}\|^2.$$

We can then formulate the MLE as the optimization problem

$$\underset{\boldsymbol{\gamma}, \sigma^2}{\text{minimize}} \quad \tfrac{T}{2}\log\left(\sigma^2\right) + \tfrac{1}{2\sigma^2}\|\mathbf{x} - \tilde{\mathbf{F}}\boldsymbol{\gamma}\|^2.$$

Setting the gradient with respect to $\boldsymbol{\gamma}$ and $1/\sigma^2$, respectively, we get

$$\frac{1}{\hat{\sigma}^2}\tilde{\mathbf{F}}^T\tilde{\mathbf{F}}\hat{\boldsymbol{\gamma}} - \frac{1}{\hat{\sigma}^2}\tilde{\mathbf{F}}^T\mathbf{x} = \mathbf{0} \quad \text{and} \quad -\frac{T}{2}\hat{\sigma}^2 + \frac{1}{2}\|\mathbf{x} - \tilde{\mathbf{F}}\hat{\boldsymbol{\gamma}}\|^2 = \mathbf{0}$$

leading to the estimators

$$\hat{\boldsymbol{\gamma}} = (\tilde{\mathbf{F}}^T\tilde{\mathbf{F}})^{-1}\tilde{\mathbf{F}}^T\mathbf{x} \quad \text{and} \quad \hat{\sigma}^2 = \frac{1}{T}\|\mathbf{x} - \tilde{\mathbf{F}}\hat{\boldsymbol{\gamma}}\|^2.$$

# Interlude: MLE is Biased

The MLE is consistent and asymptotically efficient and unbiased. However, it is biased for finite number of observations:

- the estimation of $\gamma$ is unbiased

$$E[\hat{\gamma}] = E\left[(\tilde{\mathbf{F}}^T\tilde{\mathbf{F}})^{-1}\tilde{\mathbf{F}}^T\mathbf{x}\right] = (\tilde{\mathbf{F}}^T\tilde{\mathbf{F}})^{-1}\tilde{\mathbf{F}}^T E[\mathbf{x}] = (\tilde{\mathbf{F}}^T\tilde{\mathbf{F}})^{-1}\tilde{\mathbf{F}}^T\tilde{\mathbf{F}}\gamma = \gamma$$

- the estimation of $\sigma^2$ is, however, biased. First, rewrite it as

$$\hat{\sigma}^2 = \frac{1}{T}\|\mathbf{x} - \tilde{\mathbf{F}}\hat{\gamma}\|^2 = \frac{1}{T}\|\mathbf{x} - \tilde{\mathbf{F}}(\tilde{\mathbf{F}}^T\tilde{\mathbf{F}})^{-1}\tilde{\mathbf{F}}^T\mathbf{x}\|^2 = \frac{1}{T}\|\mathbf{P}_{\tilde{\mathbf{F}}}^{\perp}\mathbf{x}\|^2$$

  where $\mathbf{P}_{\tilde{\mathbf{F}}}^{\perp} = \mathbf{I} - \tilde{\mathbf{F}}(\tilde{\mathbf{F}}^T\tilde{\mathbf{F}})^{-1}\tilde{\mathbf{F}}^T$ is the projection onto the subspace orthogonal to the one spanned by $\tilde{\mathbf{F}}$ (contains $T-(K+1)$ eigenvalues equal to one and $K+1$ zero eigenvalues). Now, the bias is

$$E\left[\hat{\sigma}^2\right] = \frac{1}{T}E\left[\mathbf{x}^T\mathbf{P}_{\tilde{\mathbf{F}}}^{\perp}\mathbf{x}\right] = \frac{1}{T}\mathrm{Tr}\left(\mathbf{P}_{\tilde{\mathbf{F}}}^{\perp}E\left[\mathbf{x}\mathbf{x}^T\right]\right) =$$

$$= \frac{1}{T}\mathrm{Tr}\left(\mathbf{P}_{\tilde{\mathbf{F}}}^{\perp}\left(\sigma^2\mathbf{I} + \tilde{\mathbf{F}}\gamma\gamma^T\tilde{\mathbf{F}}^T\right)\right) = \frac{\sigma^2}{T}\mathrm{Tr}(\mathbf{P}_{\tilde{\mathbf{F}}}^{\perp}) = \sigma^2\frac{T-(K+1)}{T}$$

# Interlude: MLE is Biased

Since we now know there is a bias in the MLE of the variance, we could consider correcting this bias to obtain an unbiased estimator (Bessel't correction):

$$\hat{\sigma}^2_{unbiased} = \frac{T}{T-(K+1)} \times \hat{\sigma}^2 = \frac{1}{T-(K+1)} \|\mathbf{x} - \tilde{\mathbf{F}}\hat{\gamma}\|^2.$$

The interpretation is that $T-(K+1)$ represent the degrees of freedom.

(Note that if $\gamma$ was known perfectly, then the degrees of freedom would be $T$.)

# Outline

# Fundamental Factor Models

Fundamental factor models use observable asset specific characteristics (fundamentals) like industry classification, market capitalization, style classification (value, growth), etc., to determine the common risk factors $\{\mathbf{f}_t\}$.

- factor loading betas are constructed from observable asset characteristics (i.e., $\mathbf{B}$ is known)
- factor realizations $\{\mathbf{f}_t\}$ are then estimated/constructed for each $t$ given $\mathbf{B}$
- note that in macroeconomic factor models the process is the opposite, i.e., the factors $\{\mathbf{f}_t\}$ are given and $\mathbf{B}$ is estimated
- in practice, fundamental factor models are estimated in two ways: BARRA approach and Fama-French approach.

# BARRA Approach

The BARRA approach was pioneered by Bar Rosenberg, founder of BARRA Inc. (cf. Grinold and Kahn (2000), Conner et al. (2010), Cariño et al. (2010)).

Econometric problems:

- choice of betas
- estimation of factor realizations $\{\mathbf{f}_t\}$ for each $t$ given $\mathbf{B}$ (i.e., by running $T$ cross-sectional regressions)

# Fama-French Approach

This approach was introduced by Eugene Fama and Kenneth French (1992):

- For a given observed asset specific characteristics, e.g. size, determine factor realizations for each $t$ with the following two steps:
  1. sort the cross-section of assets based on that attribute
  2. form a hedge portfolio by longing in the top quintile and shorting in the bottom quintile of the sorted assets
- Define the common factor realizations with the return of $K$ of such hedge portfolios corresponding to the $K$ fundamental asset attributes.
- Then estimate the factor loadings using time series regressions (like in macroeconomic factor models).

# BARRA Industry Factor Model

Consider a stylized BARRA-type industry factor model with $K$ mutually exclusive industries. Define the $K$ factor loadings as

$$\beta_{i,k} = \begin{cases} 1 & \text{if asset } i \text{ is in industry } k \\ 0 & \text{otherwise} \end{cases}$$

The industry factor model is (note that $\boldsymbol{\alpha} = \mathbf{0}$)

$$\mathbf{x}_t = \mathbf{B}\mathbf{f}_t + \boldsymbol{\epsilon}_t, \quad t = 1, \ldots, T$$

LS estimation (inefficient due to heteroskedasticity in $\boldsymbol{\Psi}$):

$$\hat{\mathbf{f}}_t = (\mathbf{B}^T\mathbf{B})^{-1}\mathbf{B}^T\mathbf{x}_t, \quad t = 1, \ldots, T$$

but since $\mathbf{B}^T\mathbf{B} = \text{diag}\,(N_1, \ldots, N_K)$, where $N_k$ is the count of assets in industry $k$ ($\sum_{k=1}^K N_k = N$), then $\hat{\mathbf{f}}_t$ is a vector of industry averages!!

- The residual covariance matrix unbiased estimator is $\hat{\boldsymbol{\Psi}} = \text{diag}\left(\hat{\sigma}_1^2, \ldots, \hat{\sigma}_N^2\right)$ where $\hat{\boldsymbol{\epsilon}}_t = \mathbf{x}_t - \mathbf{B}\hat{\mathbf{f}}_t$ and

$$\hat{\sigma}_i^2 = \frac{1}{T-1}\sum_{t=1}^{T}\left(\hat{\epsilon}_{i,t} - \bar{\hat{\epsilon}}_i\right)\left(\hat{\epsilon}_{i,t} - \bar{\hat{\epsilon}}_i\right)^T \quad \text{and} \quad \bar{\hat{\epsilon}}_i = \frac{1}{T}\sum_{t=1}^{T}\hat{\epsilon}_{i,t}.$$

# Outline

# Statistical Factor Models: Factor Analysis

In statistical factor models, both the common-factors $\{\mathbf{f}_t\}$ and the factor loadings $\mathbf{B}$ are unknown. The primary methods for estimation of factor structure are

- Factor Analysis (via EM algorithm)
- Principal Component Analysis (PCA)

Both methods model the covariance matrix $\mathbf{\Sigma}$ by focusing on the sample covariance matrix $\hat{\mathbf{\Sigma}}_{\text{SCM}}$ computed as follows:

$$\mathbf{X}^T = \begin{bmatrix} \mathbf{x}_1 & \cdots & \mathbf{x}_T \end{bmatrix} \quad (N \times T)$$

$$\bar{\mathbf{X}}^T = \mathbf{X}^T \left( \mathbf{I} - \frac{1}{T}\mathbf{1}_T\mathbf{1}_T^T \right) \quad \text{(demeaned by row)}$$

$$\hat{\mathbf{\Sigma}}_{\text{SCM}} = \frac{1}{T-1}\bar{\mathbf{X}}^T\bar{\mathbf{X}}$$

# Factor Analysis Model

**Linear factor model as cross-sectional regression:**

$$\mathbf{x}_t = \boldsymbol{\alpha} + \mathbf{B}\mathbf{f}_t + \boldsymbol{\epsilon}_t, \;\; t = 1, \ldots, T$$

with $E[\mathbf{f}_t] = \boldsymbol{\mu}_f \;\; Cov\left[\mathbf{f}_t\right] = \boldsymbol{\Sigma}_f$.

**Invariance to linear transformations of $\mathbf{f}_t$:**

- The solution we seek for $\mathbf{B}$ and $\{\mathbf{f}_t\}$ is not unique (this problem was not there when only $\mathbf{B}$ or $\{\mathbf{f}_t\}$ had to be estimated).
- For any $K \times K$ invertible matrix $\mathbf{H}$ we can define $\tilde{\mathbf{f}}_t = \mathbf{H}\mathbf{f}_t$ and $\widetilde{\mathbf{B}} = \mathbf{B}\mathbf{H}^{-1}$.
- We can write the factor model as

$$\mathbf{x}_t = \boldsymbol{\alpha} + \mathbf{B}\mathbf{f}_t + \boldsymbol{\epsilon}_t = \boldsymbol{\alpha} + \tilde{\mathbf{B}}\tilde{\mathbf{f}}_t + \boldsymbol{\epsilon}_t$$

with

$$E[\tilde{\mathbf{f}}_t] = E\left[\mathbf{H}\mathbf{f}_t\right] = \mathbf{H}\boldsymbol{\mu}_f$$
$$Cov[\tilde{\mathbf{f}}_t] = Cov\left[\mathbf{H}\mathbf{f}_t\right] = \mathbf{H}\boldsymbol{\Sigma}_f\mathbf{H}^T.$$

# Factor Analysis Model

**Standard formulation of factor analysis:**

- Orthogonal factors: $\boldsymbol{\Sigma}_f = \mathbf{I}_K$
  This is achieved by choosing $\mathbf{H} = \boldsymbol{\Lambda}^{-1/2}\boldsymbol{\Gamma}^T$, where $\boldsymbol{\Sigma}_f = \boldsymbol{\Gamma}\boldsymbol{\Lambda}\boldsymbol{\Gamma}^T$ is the spectral/eigen decomposition with orthogonal $K \times K$ matrix $\boldsymbol{\Gamma}$ and diagonal matrix $\boldsymbol{\Lambda} = \operatorname{diag}(\lambda_1, \ldots, \lambda_K)$ with $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_K$.
- Zero-mean factors: $\boldsymbol{\mu}_f = \mathbf{0}$
  This is achieved by adjusting $\boldsymbol{\alpha}$ to incorporate the mean contribution from the factors: $\tilde{\boldsymbol{\alpha}} = \boldsymbol{\alpha} + \mathbf{B}\boldsymbol{\mu}_f$.

Under these assumptions, the unconditional covariance matrix of the observations is

$$\boldsymbol{\Sigma} = \mathbf{B}\mathbf{B}^T + \boldsymbol{\Psi}.$$

# Factor Analysis: Variance Decomposition

Recall the covariance matrix for the returns

$$\boldsymbol{\Sigma} = \mathbf{B}\mathbf{B}^T + \boldsymbol{\Psi}.$$

For a given asset $i$, the return variance may then be expressed as

$$var\left(x_{i,t}\right) = \sum_{k=1}^{K} \beta_{ik}^2 + \sigma_i^2$$

- variance portion due to common factors, $\sum_{k=1}^{K} \beta_{ik}^2$, is called the *communality*
- variance portion due to specific factors, $\sigma_i^2$, is called the *uniqueness*
- another quantity of interest is the factor's marginal contribution to active risk (FMCAR).

**Maximum likelihood estimation:**

Consider the model

$$\mathbf{x}_t = \boldsymbol{\alpha} + \mathbf{B}\mathbf{f}_t + \boldsymbol{\epsilon}_t, \quad t = 1, \ldots, T$$

where

- $\boldsymbol{\alpha}$ and $\mathbf{B}$ are vector/matrix constants
- all random variables are Gaussian/Normal:
  - $\mathbf{f}_t$ i.i.d. $N(\mathbf{0}, \mathbf{I})$
  - $\boldsymbol{\epsilon}_t$ i.i.d. $N(\mathbf{0}, \boldsymbol{\Psi})$ with $\boldsymbol{\Psi} = \text{diag}\left(\sigma_1^2, \ldots, \sigma_N^2\right)$
  - $\mathbf{x}_t$ i.i.d. $N\left(\boldsymbol{\alpha}, \boldsymbol{\Sigma} = \mathbf{B}\mathbf{B}^T + \boldsymbol{\Psi}\right)$

**Probability density function (pdf):**

$$\begin{aligned} p\left(\mathbf{x}_1, \ldots, \mathbf{x}_T \mid \boldsymbol{\alpha}, \boldsymbol{\Sigma}\right) &= \prod_{t=1}^{T} p\left(\mathbf{x}_t \mid \boldsymbol{\alpha}, \boldsymbol{\Sigma}\right) \\ &= \prod_{t=1}^{T} (2\pi)^{-\frac{N}{2}} \left|\boldsymbol{\Sigma}\right|^{-\frac{1}{2}} \exp\left(-\frac{1}{2}\left(\mathbf{x}_t - \boldsymbol{\alpha}\right)^T \boldsymbol{\Sigma}^{-1}\left(\mathbf{x}_t - \boldsymbol{\alpha}\right)\right) \end{aligned}$$

# Factor Analysis: MLE

**Likelihood of the factor model:**

The log-likelihood of the parameters $(\boldsymbol{\alpha}, \boldsymbol{\Sigma})$ given the $T$ i.i.d. observations is

$$\begin{aligned} L(\boldsymbol{\alpha}, \boldsymbol{\Sigma}) &= \log p(\mathbf{x}_1, \ldots, \mathbf{x}_T \mid \boldsymbol{\alpha}, \boldsymbol{\Sigma}) \\ &= -\frac{TN}{2}\log(2\pi) - \frac{T}{2}\log|\boldsymbol{\Sigma}| - \frac{1}{2}\sum_{t=1}^{T}(\mathbf{x}_t - \boldsymbol{\alpha})^T \boldsymbol{\Sigma}^{-1}(\mathbf{x}_t - \boldsymbol{\alpha}) \end{aligned}$$

**Maximum likelihood estimation (MLE):**

$$\begin{array}{ll} \underset{\boldsymbol{\alpha}, \boldsymbol{\Sigma}, \mathbf{B}, \boldsymbol{\Psi}}{\text{minimize}} & \frac{T}{2}\log|\boldsymbol{\Sigma}| + \frac{1}{2}\sum_{t=1}^{T}(\mathbf{x}_t - \boldsymbol{\alpha})^T \boldsymbol{\Sigma}^{-1}(\mathbf{x}_t - \boldsymbol{\alpha}) \\ \text{subject to} & \boldsymbol{\Sigma} = \mathbf{B}\mathbf{B}^T + \boldsymbol{\Psi} \end{array}$$

Note that without the constraint, the solution would be

$$\hat{\boldsymbol{\alpha}} = \frac{1}{T}\sum_{t=1}^{T}\mathbf{x}_t \quad \text{and} \quad \hat{\boldsymbol{\Sigma}} = \frac{1}{T}\sum_{t=1}^{T}(\mathbf{x}_t - \hat{\boldsymbol{\alpha}})(\mathbf{x}_t - \hat{\boldsymbol{\alpha}})^T.$$

# Factor Analysis: MLE

**MLE solution:**

$$\begin{array}{ll}\underset{\boldsymbol{\alpha},\boldsymbol{\Sigma},\mathbf{B},\boldsymbol{\Psi}}{\text{minimize}} & \frac{T}{2}\log\left|\boldsymbol{\Sigma}\right| + \frac{1}{2}\sum_{t=1}^{T}\left(\mathbf{x}_t - \boldsymbol{\alpha}\right)^T\boldsymbol{\Sigma}^{-1}\left(\mathbf{x}_t - \boldsymbol{\alpha}\right)\\ \text{subject to} & \boldsymbol{\Sigma} = \mathbf{B}\mathbf{B}^T + \boldsymbol{\Psi}\end{array}$$

- Employ the Expectation-Maximization (EM) algorithm to compute $\hat{\boldsymbol{\alpha}}$, $\hat{\mathbf{B}}$, and $\hat{\boldsymbol{\Psi}}$.
- Estimate factor realization $\{\mathbf{f}_t\}$ using, for example, the GLS estimator:

$$\hat{\mathbf{f}}_t = (\hat{\mathbf{B}}^T\hat{\boldsymbol{\Psi}}^{-1}\hat{\mathbf{B}})^{-1}\hat{\mathbf{B}}^T\hat{\boldsymbol{\Psi}}^{-1}\mathbf{x}_t, \;\; t = 1, \dots, T.$$

The number of factors $K$ can be estimated with a variety of methods such as the likelihood ratio (LR) test, Akaike information criterion (AIC), etc. For example:

$$LR(K) = -(T - 1 - \frac{1}{6}(2N+5) - \frac{2}{3}K)\left(\log|\hat{\boldsymbol{\Sigma}}_{\text{SCM}}| - \log|\hat{\mathbf{B}}\hat{\mathbf{B}}^T + \hat{\boldsymbol{\Psi}}|\right)$$

which is asymptotically chi-square with $\frac{1}{2}((N-K)^2 - N - K)$ degrees of freedom.

# Principal Component Analysis

**PCA:** is a dimension reduction technique used to explain the majority of the information in the covariance matrix.

- $N$-variate random variable $\mathbf{x}$ with $E[\mathbf{x}] = \boldsymbol{\alpha}$ and $Cov[\mathbf{x}] = \boldsymbol{\Sigma}$.
- Spectral/eigen decomposition $\boldsymbol{\Sigma} = \boldsymbol{\Gamma}\boldsymbol{\Lambda}\boldsymbol{\Gamma}^T$ where
  - $\boldsymbol{\Lambda} = \mathrm{diag}(\lambda_1, \ldots, \lambda_K)$ with $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_K$
  - $\boldsymbol{\Gamma}$ orthogonal $K \times K$ matrix: $\boldsymbol{\Gamma}^T\boldsymbol{\Gamma} = \mathbf{I}_N$
- Principal component variables: $\mathbf{p} = \boldsymbol{\Gamma}^T(\mathbf{x} - \boldsymbol{\alpha})$ with

$$E[\mathbf{p}] = E\left[\boldsymbol{\Gamma}^T(\mathbf{x} - \boldsymbol{\alpha})\right] = \boldsymbol{\Gamma}^T(E[\mathbf{x}] - \boldsymbol{\alpha}) = \mathbf{0}$$

$$Cov[\mathbf{p}] = Cov\left[\boldsymbol{\Gamma}^T(\mathbf{x} - \boldsymbol{\alpha})\right] = \boldsymbol{\Gamma}^T Cov[\mathbf{x}]\,\boldsymbol{\Gamma} = \boldsymbol{\Gamma}^T\boldsymbol{\Sigma}\boldsymbol{\Gamma} = \boldsymbol{\Lambda}.$$

- $\mathbf{p}$ is a vector of zero-mean, uncorrelated random variables, in order of importance (i.e., the first components explain the largest portion of the sample covariance matrix)
- In terms of multifactor model, the $K$ most important principal components are the factor realizations.

# Factor Models: Principal Component Analysis

**Factor model from PCA:**

- From PCA, we can write the random vector $\mathbf{x}$ as

$$\mathbf{x} = \boldsymbol{\alpha} + \boldsymbol{\Gamma}\mathbf{p}$$

where $E[\mathbf{p}] = \mathbf{0}$ and $Cov[\mathbf{p}] = \boldsymbol{\Lambda} = \text{diag}(\boldsymbol{\Lambda}_1, \boldsymbol{\Lambda}_2)$.

- Partition $\boldsymbol{\Gamma} = \begin{bmatrix} \boldsymbol{\Gamma}_1 & \boldsymbol{\Gamma}_2 \end{bmatrix}$ where $\boldsymbol{\Gamma}_1$ corresponds to the $K$ largest eigenvalues of $\boldsymbol{\Sigma}$.
- Partition $\mathbf{p} = \begin{bmatrix} \mathbf{p}_1 \\ \mathbf{p}_2 \end{bmatrix}$ where $\mathbf{p}_1$ contains the first $K$ elements
- Then we can write

$$\mathbf{x} = \boldsymbol{\alpha} + \boldsymbol{\Gamma}_1\mathbf{p}_1 + \boldsymbol{\Gamma}_2\mathbf{p}_2 = \boldsymbol{\alpha} + \mathbf{B}\mathbf{f} + \boldsymbol{\epsilon}$$

where

$$\mathbf{B} = \boldsymbol{\Gamma}_1, \quad \mathbf{f} = \mathbf{p}_1 \quad \text{and} \quad \boldsymbol{\epsilon} = \boldsymbol{\Gamma}_2\mathbf{p}_2.$$

This is like a factor model except that $Cov[\boldsymbol{\epsilon}] = \boldsymbol{\Gamma}_2\boldsymbol{\Lambda}_2\boldsymbol{\Gamma}_2^T$, where $\boldsymbol{\Lambda}_2$ is a diagonal matrix of last $N - K$ eigenvalues but $Cov[\boldsymbol{\epsilon}]$ is not diagonal...

# Factor Models: Why PCA?

**But why is PCA the desired solution to the statistical factor model?**

- The idea is to obtain the factors from the returns themselves:

$$\mathbf{x}_t = \boldsymbol{\alpha} + \mathbf{B}\mathbf{f}_t + \boldsymbol{\epsilon}_t, \quad t = 1, \ldots, T$$
$$\mathbf{f}_t = \mathbf{C}^T\mathbf{x}_t + \mathbf{d}$$

  or, more compactly,

$$\mathbf{x}_t = \boldsymbol{\alpha} + \mathbf{B}\left(\mathbf{C}^T\mathbf{x}_t + \mathbf{d}\right) + \boldsymbol{\epsilon}_t, \quad t = 1, \ldots, T$$

- The problem formulation is

$$\underset{\boldsymbol{\alpha}, \mathbf{B}, \mathbf{C}, \mathbf{d}}{\text{minimize}} \quad \tfrac{1}{T}\textstyle\sum_{t=1}^{T} \|\mathbf{x}_t - \boldsymbol{\alpha} + \mathbf{B}\left(\mathbf{C}^T\mathbf{x}_t + \mathbf{d}\right)\|^2$$

- Solution is involved to derive and is given by: $\mathbf{f}_t = \boldsymbol{\Gamma}_1^T(\mathbf{x}_t - \boldsymbol{\alpha})$ or, if normalized factors are desired, $\mathbf{f}_t = \boldsymbol{\Lambda}_1^{-\frac{1}{2}}\boldsymbol{\Gamma}_1^T(\mathbf{x}_t - \boldsymbol{\alpha})$.

# Factor Models via PCA

1. Compute sample estimates

$$\hat{\boldsymbol{\alpha}} = \frac{1}{T}\sum_{t=1}^{T}\mathbf{x}_t \quad \text{and} \quad \hat{\boldsymbol{\Sigma}} = \frac{1}{T}\sum_{t=1}^{T}\left(\mathbf{x}_t - \hat{\boldsymbol{\alpha}}\right)\left(\mathbf{x}_t - \hat{\boldsymbol{\alpha}}\right)^T.$$

2. Compute spectral decomposition:

$$\hat{\boldsymbol{\Sigma}} = \hat{\boldsymbol{\Gamma}}\hat{\boldsymbol{\Lambda}}\hat{\boldsymbol{\Gamma}}^T, \quad \hat{\boldsymbol{\Gamma}} = \left[\begin{array}{cc} \hat{\boldsymbol{\Gamma}}_1 & \hat{\boldsymbol{\Gamma}}_2 \end{array}\right]$$

3. Form the factor loadings, factor realizations, and residuals (so that $\mathbf{x}_t = \hat{\boldsymbol{\alpha}} + \hat{\mathbf{B}}\hat{\mathbf{f}}_t + \hat{\boldsymbol{\epsilon}}_t$):

$$\hat{\mathbf{B}} = \hat{\boldsymbol{\Gamma}}_1\hat{\boldsymbol{\Lambda}}_1^{\frac{1}{2}}, \quad \hat{\mathbf{f}}_t = \hat{\boldsymbol{\Lambda}}_1^{-\frac{1}{2}}\hat{\boldsymbol{\Gamma}}_1^T\left(\mathbf{x}_t - \hat{\boldsymbol{\alpha}}\right), \quad \hat{\boldsymbol{\epsilon}}_t = \mathbf{x}_t - \hat{\boldsymbol{\alpha}} - \hat{\mathbf{B}}\hat{\mathbf{f}}_t$$

4. Covariance matrices:

$$\hat{\boldsymbol{\Sigma}}_{\mathbf{f}} = \frac{1}{T}\sum_{t=1}^{T}\hat{\mathbf{f}}_t\hat{\mathbf{f}}_t^T = \mathbf{I}_K, \quad \hat{\boldsymbol{\Psi}} = \frac{1}{T}\sum_{t=1}^{T}\hat{\boldsymbol{\epsilon}}_t\hat{\boldsymbol{\epsilon}}_t^T = \hat{\boldsymbol{\Gamma}}_2\hat{\boldsymbol{\Lambda}}_2\hat{\boldsymbol{\Gamma}}_2 \text{ (not diagonal!)}$$

$$\hat{\boldsymbol{\Sigma}} = \hat{\boldsymbol{\Gamma}}_1\hat{\boldsymbol{\Lambda}}_1\hat{\boldsymbol{\Gamma}}_1 + \hat{\boldsymbol{\Gamma}}_2\hat{\boldsymbol{\Lambda}}_2\hat{\boldsymbol{\Gamma}}_2 = \hat{\mathbf{B}}\hat{\boldsymbol{\Sigma}}_{\mathbf{f}}\hat{\mathbf{B}}^T + \hat{\boldsymbol{\Psi}}.$$

# Principal Factor Method

Since the previous method does not lead to a diagonal covariance matrix for the residual, let's refine the method with the following iterative approach

1. PCA:
   - sample mean: $\hat{\alpha} = \bar{\mathbf{x}} = \frac{1}{T}\mathbf{X}^T\mathbf{1}_T$
   - demeaned matrix: $\bar{\mathbf{X}}^T = \mathbf{X}^T - \bar{\mathbf{x}}\mathbf{1}_T^T$
   - sample covariance matrix: $\hat{\boldsymbol{\Sigma}} = \frac{1}{T-1}\bar{\mathbf{X}}^T\bar{\mathbf{X}}$
   - eigen-decomposition: $\hat{\boldsymbol{\Sigma}} = \hat{\boldsymbol{\Gamma}}_0\hat{\boldsymbol{\Lambda}}_0\hat{\boldsymbol{\Gamma}}_0^T$
   - set index $s = 0$
2. Estimates:
   - $\hat{\mathbf{B}}_{(s)} = \hat{\boldsymbol{\Gamma}}_{(s-1)}\hat{\boldsymbol{\Lambda}}_{(s-1)}^{\frac{1}{2}}$
   - $\hat{\boldsymbol{\Psi}}_{(s)} = \mathrm{diag}(\hat{\boldsymbol{\Sigma}} - \hat{\mathbf{B}}_{(s)}\hat{\mathbf{B}}_{(s)}^T)$
   - $\hat{\boldsymbol{\Sigma}}_{(s)} = \hat{\mathbf{B}}_{(s)}\hat{\mathbf{B}}_{(s)}^T + \hat{\boldsymbol{\Psi}}_{(s)}$
3. Update the eigen-decomposition as $\hat{\boldsymbol{\Sigma}} - \hat{\boldsymbol{\Psi}}_{(s)} = \hat{\boldsymbol{\Gamma}}_{(s)}\hat{\boldsymbol{\Lambda}}_{(s)}\hat{\boldsymbol{\Gamma}}_{(s)}^T$
4. Update $s \leftarrow s + 1$ and repeat Steps 2-3 generating a sequence of estimates $(\hat{\mathbf{B}}_{(s)}, \hat{\boldsymbol{\Psi}}_{(s)}, \hat{\boldsymbol{\Sigma}}_{(s)})$ until convergence.

# Thanks

For more information visit:

https://www.danielppalomar.com