

Factor Models and Analysis

Christopher Ting

<http://cting.x10host.com/CUHKSZ/CUHKSZ.html>

<http://www.mysmu.edu/faculty/christophert/>
✉: christophert@smu.edu.sg

July 14, 2018






Broad Lesson Plan

- 1 Introduction
- 2 Factor Model Specification
- 3 Macroeconomic Factor Models
- 4 Fundamental Factor Models
- 5 Takeaways

What is Factor Analysis?

- Are there relationships among many dependent variables?
- How many different factors are needed to explain the pattern of relationships among these variables?
- What is the nature of these factors?
- How well do the hypothesized factors explain the observed data?
- How much purely random or unique variance does each observed variable include?
- Goal: To discover something about the nature of the independent variables that affect dependent variables, even though those independent variables are not measured directly.

Functions of Factor Models

-  Decompose risk and return into explainable and unexplainable components
-  Generate estimates of abnormal return
-  Describe the covariance structure of returns
-  Predict returns in specified stress scenarios
-  Provide a framework for portfolio risk analysis

Three Types of Factor Models

- 👉 Macroeconomic factor model
 - Factors are observable economic and financial time series.
- 👉 Fundamental factor model
 - Factors are created from observable asset characteristics.
- 👉 Statistical factor model (PCA)
 - Factors are unobservable and extracted from asset returns.

General Form

▢ R_{it} : simple return on asset i ($i = 1, \dots, N$) in time period t ($t = 1, \dots, T$).

▢ f_{kt} : k -th **common factors** ($k = 1, \dots, M$)

▢ β_{ki} : **Factor loading** or **factor beta** for asset i on the k -th factor

▢ ϵ_{it} : asset's **specific factor**

$$\begin{aligned} R_{it} &= \alpha_i + \beta_{1i}f_{1t} + \beta_{2i}f_{2t} + \cdots + \beta_{Ki}f_{Mt} + \epsilon_{it} \\ &= \alpha_i + \boldsymbol{\beta}'_i \mathbf{f}_t + \epsilon_{it} \end{aligned} \tag{1}$$

Assumptions

- ▢ The factor realizations, \mathbf{f}_t , are stationary with unconditional moments

$$\mathbb{E}(\mathbf{f}_t) = \boldsymbol{\mu}_f$$

$$\mathbb{V}(\mathbf{f}_t) = \boldsymbol{\Omega}_f = \mathbb{E}((\mathbf{f}_t - \boldsymbol{\mu}_f)(\mathbf{f}_t - \boldsymbol{\mu}_f)')$$

- ▢ Asset specific error terms, ϵ_{it} , are uncorrelated with each of the common factors, f_{kt} ,

$$\mathbb{C}(f_{kt}, \epsilon_{it}) = 0, \quad \text{for all } k, i, \text{ and } t.$$

- ▢ Error terms ϵ_{it} are serially uncorrelated and contemporaneously uncorrelated across assets

$$\mathbb{C}(\epsilon_{it}, \epsilon_{js}) = \begin{cases} \sigma_i^2, & \text{for all } i = j \text{ and } t = s; \\ 0, & \text{otherwise.} \end{cases}$$

Notations

- 📁 Vectors with a subscript t : **cross-section** of all assets

$$\mathbf{R}_t = \begin{bmatrix} R_{1t} \\ \vdots \\ R_{Nt} \end{bmatrix}$$

- 📁 Vectors with a subscript i : **time series** of a given asset

$$\mathbf{R}_i = \begin{bmatrix} R_{i1} \\ \vdots \\ R_{iT} \end{bmatrix}$$

- 📁 $T \times N$ matrix of all assets over all time periods

$$\mathbf{R} = \begin{bmatrix} R_{11} & \cdots & R_{N1} \\ \vdots & \ddots & \vdots \\ R_{1T} & \cdots & R_{NT} \end{bmatrix}$$

Alpha

- Expected return of asset i

$$\mathbb{E}(R_{it}) = \alpha_i + \beta'_i \mathbb{E}(\mathbf{f}_t)$$

- $\beta'_i \mathbb{E}(\mathbf{f}_t)$ = explained expected return due to systematic risk factors

- $\alpha_i =: \mathbb{E}(R_{it}) - \beta'_i \mathbb{E}(\mathbf{f}_t)$ is the unexplained expected return.

- Equilibrium asset pricing models impose the restriction $\alpha_i = 0$ (no abnormal return) for all assets $i = 1, \dots, N$.

Covariance

Under cross-section regression,

$$\mathbf{R}_t = \boldsymbol{\alpha} + \mathbf{B}\mathbf{f}_t + \boldsymbol{\epsilon}_t, \quad t = 1, 2, \dots, T,$$

where \mathbf{R}_t is a $N \times 1$ vector, and \mathbf{B} is a $N \times M$ matrix.

The $N \times N$ covariance matrix under the assumptions of multi-factor model is,

$$\mathbb{C}(\mathbf{R}_t) = \mathbf{B}\boldsymbol{\Omega}_f\mathbf{B}' + \mathbf{D},$$

where the diagonal matrix is

$$\mathbf{D} := \mathbb{E}(\boldsymbol{\epsilon}_t\boldsymbol{\epsilon}_t' | \mathbf{f}_t).$$

That is, with $\sigma_i^2 := D_{ii}$,

$$\mathbb{V}(R_{it}) = \boldsymbol{\beta}_i'\boldsymbol{\Omega}_f\boldsymbol{\beta}_i + \sigma_i^2$$

$$\mathbb{C}(R_{it}, R_{jt}) = \boldsymbol{\beta}_i'\boldsymbol{\Omega}_f\boldsymbol{\beta}_j$$

An Example

📄 Consider a multi-factor of price change of an asset:

$$\Delta p_i = \mu_i + \beta_{i1}\Delta f_1 + \cdots + \beta_{iM}\Delta f_M + \epsilon_i.$$

📄 In this model

- μ_i : mean price change of the i -th asset
- Δf_j : change of the j -th factor
- β_{ij} : factor loading of the i -th asset on the j -th factor
- ϵ_i : idiosyncratic error of the i -th asset

📄 In vector-matrix form,

$$\Delta \mathbf{p} = \boldsymbol{\mu} + \mathbf{B}\Delta \mathbf{f} + \boldsymbol{\epsilon}.$$

Covariance of the Example

- We can relate the covariance matrix Ω of the individual asset price changes to the covariance matrix Ω_f of the factor changes:

$$\begin{aligned}
 \Omega &= \mathbb{E}((\Delta \mathbf{p} - \boldsymbol{\mu})(\Delta \mathbf{p} - \boldsymbol{\mu})') \\
 &= \mathbb{E}((\mathbf{B}\Delta \mathbf{f} + \boldsymbol{\epsilon})(\mathbf{B}\Delta \mathbf{f} + \boldsymbol{\epsilon})') \\
 &= \mathbf{B} \mathbb{E}(\Delta \mathbf{f}\Delta \mathbf{f}')\mathbf{B}' + \mathbf{B} \mathbb{E}(\Delta \mathbf{f}\boldsymbol{\epsilon}') + \mathbb{E}(\boldsymbol{\epsilon}\Delta \mathbf{f}')\mathbf{B}' + \mathbb{E}(\boldsymbol{\epsilon}\boldsymbol{\epsilon}') \\
 &= \mathbf{B} \mathbb{E}(\Delta \mathbf{f}\Delta \mathbf{f}')\mathbf{B}' + \mathbf{D} \\
 &= \mathbf{B}\Omega_f\mathbf{B}' + \mathbf{D}
 \end{aligned}$$

Portfolio Analysis

Let $\mathbf{w} = [w_1 \ w_2 \ \cdots \ w_N]$ be a vector of portfolio weights.

Hence, given N simple returns R_{it} , the return of the portfolio, R_{pt} , is

$$R_{pt} = \mathbf{w}'\mathbf{R}_t = \sum_{i=1}^N w_i R_{it}.$$

Portfolio Factor Model: Since $\mathbf{R}_t = \boldsymbol{\alpha} + \mathbf{B}\mathbf{f}_t + \boldsymbol{\epsilon}_t$, we get

$$R_{pt} = \mathbf{w}'\boldsymbol{\alpha} + \mathbf{w}'\mathbf{B}\mathbf{f}_t + \mathbf{w}'\boldsymbol{\epsilon}_t =: \alpha_p + \boldsymbol{\beta}'_p\mathbf{f}_t + \epsilon_{pt}.$$

The variance of R_{pt} is

$$\mathbb{V}(R_{pt}) = \boldsymbol{\beta}'_p\boldsymbol{\Omega}_f\boldsymbol{\beta}_p + \mathbb{V}(\epsilon_{pt}) = \mathbf{w}'\mathbf{B}\boldsymbol{\Omega}_f\mathbf{B}'\mathbf{w} + \mathbf{w}'\mathbf{D}\mathbf{w}.$$

Active versus Static Portfolio

- Active portfolios have weights that change over time due to active asset allocation decisions.
- Static portfolios have weights that are fixed over time (e.g. equally weighted portfolio).
- Factor models can be used to analyze the risk of both active and static portfolios.

Econometric Problems

- ✧ The model is

$$R_{it} = \alpha_i + \beta_i' \mathbf{f}_t + \epsilon_{it}.$$

- ✧ Choose factors \mathbf{f}_t (observed economic and financial time series).
- ✧ Estimate factor betas, β_i , and residual variances, σ_i^2 , using time series regression techniques.
- ✧ Estimate factor covariance matrix, Ω_f , from observed history of factors.

Sharpe's Single Factor Model

- Let R_{mt} denote the return or excess return (relative to the risk-free rate) on a market portfolio at time t .
- A macroeconomic factor model with a single market factor:

$$R_{it} = \alpha_i + \beta_i R_{mt} + \epsilon_{it},$$

for $i = 1, 2, \dots, N$, and $t = 1, 2, \dots, T$.

- Risk-adjusted expected return** and **abnormal return**

$$\mathbb{E}(R_{it}) = \beta_i \mathbb{E}(R_{mt})$$

$$\alpha_i := \mathbb{E}(R_{it}) - \beta_i \mathbb{E}(R_{mt})$$

Estimation of Sharpe's Single Factor Model

- Because R_{mt} is observable, the parameters β_i and σ_i^2 of the single factor model for each asset can be estimated using time series regression (i.e., ordinary least squares).
- The “slope” estimate is, for $i = 1, 2, \dots, N$,

$$\hat{\beta}_i = \frac{\hat{\sigma}_{im}}{\hat{\sigma}_m^2}.$$

- The “intercept” estimate is, for $i = 1, 2, \dots, N$,

$$\hat{\alpha}_i = \bar{R}_i - \hat{\beta}_i \bar{R}_m.$$

- The variance of the residuals is

$$\hat{\sigma}_i^2 = \frac{1}{T-2} \hat{\epsilon}_i' \hat{\epsilon}_i.$$

General Multi-Factor Model

- Model specifies M observable macro-variables

$$R_{it} = \alpha_i + \beta_i' \mathbf{f}_t + \epsilon_{it}.$$

- Sometimes the macroeconomic factors are standardized to have mean zero and a common scale.
- The factors must be stationary (not trending).
- Sometimes the factors are made orthogonal.

Estimation: Multi-Factor Model

- The factor realizations f_t are observable and we can organize it as

$$F = \begin{bmatrix} f'_1 \\ f'_2 \\ \vdots \\ f'_T \end{bmatrix}.$$

- For multi-factor model with a factor matrix F , it is

$$R_i = \alpha_i \mathbf{1}_T + F\beta_i + \epsilon_i, \quad \text{for } i = 1, 2, \dots, N.$$

- We define $X := [\mathbf{1}_T \ F]$.

- The factor loading is solved by OLS to yield, for each i ,

$$\left(\hat{\alpha}_i \quad \hat{\beta}'_i \right)' = (X'X)^{-1} X'R_i.$$

Estimation: Multi-Factor Model (Cont'd)

- The variance of the residuals is, by setting $K := M + 1$,

$$\hat{\sigma}_i^2 = \frac{1}{T - K} \hat{\boldsymbol{\epsilon}}_i' \hat{\boldsymbol{\epsilon}}_i.$$

- The covariance matrix of the factor realizations may be estimated using the time series sample covariance matrix

$$\hat{\boldsymbol{\Omega}}_f = \frac{1}{T - 1} \sum_{t=1}^T (\mathbf{f}_t - \bar{\mathbf{f}}) (\mathbf{f}_t - \bar{\mathbf{f}})', \quad \text{where } \bar{\mathbf{f}} = \frac{1}{T} \sum_{t=1}^T \mathbf{f}_t.$$

- The estimated multi-factor model covariance matrix is then

$$\hat{\boldsymbol{\Omega}}_{\text{FM}} = \hat{\mathbf{B}} \hat{\boldsymbol{\Omega}}_f \hat{\mathbf{B}}' + \hat{\mathbf{D}}.$$

What is a Fundamental Factor Model?

- 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.
- Factor betas are constructed from observable asset characteristics (i.e., B is known)
- Factor realizations, f_t , are estimated/constructed for each t given B .
- In practice, fundamental factor models are estimated in two ways.
 - BARRA approach
 - Fama-French Approach

BARRA Approach

- ✧ Pioneered by Bar Rosenberg, founder of BARRA Inc, which was merged with **MSCI** in 2004.
- ✧ Observable asset specific fundamentals (or some transformation of them) are treated as the factor betas, β_i .
- ✧ Factor realizations at time t , f_t , are treated as unobservable.
- ✧ The econometric problem is then to estimate the factor realizations at time t given the factor betas.
- ✧ Estimation is done by running T cross-section regressions.

BARRA-Type Single Factor Model

- Consider a single factor model in the form of a cross-sectional regression at time t :

$$\mathbf{R}_t = \boldsymbol{\beta} f_t + \boldsymbol{\epsilon}_t, \quad t = 1, 2, \dots, T.$$

- $\boldsymbol{\beta}$ is an $N \times 1$ vector of observed values of an asset specific attribute (e.g., market capitalization, industry classification, style classification).
- f_t is an unobserved factor realization.
- Assumptions include
 - $\mathbb{V}(f_t) = \sigma_f^2$
 - $\mathbb{C}(f_t, \epsilon_{it}) = 0$, for all $i = 1, 2, \dots, N$.
 - $\mathbb{V}(\epsilon_{it}) = \sigma_i^2$, for all $i = 1, 2, \dots, N$.

Estimation: BARRA Approach

- For each time period $t = 1, 2, \dots, T$, the vector of factor betas, β , is treated as data and the factor realization f_t , is the parameter to be estimated.
- Since the error term ϵ_t is heteroskedastic, efficient estimation of f_t is done by **weighted least squares** (WLS), assuming the asset specific variances σ_i^2 are known.

$$\hat{f}_{t,\text{wls}} = (\beta' D^{-1} \beta)^{-1} \beta' D^{-1} R_t, \quad t = 1, 2, \dots, T,$$

where D is a diagonal matrix with $\sigma_1^2, \sigma_2^2, \dots, \sigma_N^2$ along the diagonal.

- σ_i^2 can be consistently estimated and a feasible WLS estimate can be computed.

Factor Mimicking Portfolio

- f_t in the BARRA approach has an interesting interpretation as the return on a portfolio $\mathbf{h} = [h_1 \ h_2 \ \cdots \ h_N]'$ that solves

$$\min_{\mathbf{h}} \mathbf{h}' \mathbf{D} \mathbf{h} \quad \text{subject to } \mathbf{h}' \boldsymbol{\beta} = 1.$$

- The portfolio \mathbf{h} minimizes asset return residual variance subject to having unit exposure to the attribute $\boldsymbol{\beta}$ and is given by

$$\mathbf{h}' = (\boldsymbol{\beta}' \mathbf{D}^{-1} \boldsymbol{\beta})^{-1} \boldsymbol{\beta}' \mathbf{D}^{-1}.$$

- The estimated factor realization is then the portfolio return

$$\hat{f}_{t,\text{wls}} = \mathbf{h}' \mathbf{R}_t.$$

- When the portfolio \mathbf{h} is normalized such that $\sum_i^N h_i = 1$, it is referred to as a **factor mimicking portfolio**.

BARRA-Type Industry Factor Model

- Consider a stylized BARRA-type industry factor model with J mutually exclusive industries.
- The factor sensitivities β_{ij} in (1) for each asset are time invariant.
- The factor betas are dummy variables indicating whether a given asset is in a particular industry.

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

- The estimated value of f_{jt} will be equal to the weighted average excess return in time period t of the firms operating in industry j .

Industry Factor Model Regression

- The industry factor model with J industries is summarized as

$$R_{it} = \beta_{i1}f_{1t} + \cdots + \beta_{iJ}f_{Jt} + \epsilon_{it}$$

$$\mathbb{V}(\epsilon_{it}) = \sigma_i^2$$

$$\mathbb{C}(\epsilon_{it}, f_{jt}) = 0$$

$$\mathbb{C}(f_{it}, f_{jt}) = \sigma_{ij}$$

- It is understood that there are N_j firms in the j -th industry such that

$$\sum_{j=1}^J N_j = N.$$

Estimation of Industry Factor Model Factors

- Consider the cross-section regression at time t :

$$\begin{aligned} \mathbf{R}_t &= \beta_1 f_{1t} + \cdots + \beta_J f_{Jt} + \boldsymbol{\epsilon}_t \\ &= \mathbf{B} \mathbf{f}_t + \boldsymbol{\epsilon}_t \end{aligned}$$

$$\mathbb{E}(\boldsymbol{\epsilon}_t, \boldsymbol{\epsilon}_t') = \mathbf{D}$$

$$\mathbb{C}(\mathbf{f}_t) = \boldsymbol{\Omega}_f$$

- Since the industries are mutually exclusive it follows that

$$\beta_j' \beta_k = \begin{cases} N_k & \text{for } j = k \\ 0 & \text{otherwise} \end{cases}$$

Estimation of Industry Factor Model Factors (Cont'd)

- An unbiased but inefficient estimate of the factor realizations \mathbf{f}_t can be obtained by OLS:

$$\hat{\mathbf{f}}_{t,\text{OLS}} = (\mathbf{B}'\mathbf{B})^{-1}\mathbf{B}'\mathbf{R}_t = \begin{bmatrix} \hat{\mathbf{f}}_{1t,\text{OLS}} \\ \vdots \\ \hat{\mathbf{f}}_{Jt,\text{OLS}} \end{bmatrix} = \begin{bmatrix} \frac{1}{N_1} \sum_{i=1}^{N_1} R_{it}^1 \\ \vdots \\ \vdots \\ \frac{1}{N_J} \sum_{i=1}^{N_J} R_{it}^J \end{bmatrix}$$

Estimation of Factor Realization Covariance Matrix

- Given $\hat{\mathbf{f}}_{1,OLS}, \dots, \hat{\mathbf{f}}_{T,OLS}$, the covariance matrix of the industry factors may be computed as the time series sample covariance:

$$\bar{\mathbf{f}}_{OLS} = \frac{1}{T} \sum_{t=1}^T \hat{\mathbf{f}}_{t,OLS}$$

$$\hat{\Omega}_{OLS}^F = \frac{1}{T-1} \sum_{t=1}^T (\hat{\mathbf{f}}_{t,OLS} - \bar{\mathbf{f}}_{OLS})(\hat{\mathbf{f}}_{t,OLS} - \bar{\mathbf{f}}_{OLS})'$$

Estimation of Residual Variances

- The residual variances, $\mathbb{V}(\epsilon_{it}) = \sigma_i^2$, can be estimated from the time series of residuals from the T cross-section regressions.
- For each i , σ_i^2 may be estimated, for $i = 1, 2, \dots, N$,

$$\bar{\epsilon}_{i,\text{OLS}} = \frac{1}{T} \sum_{t=1}^T \hat{\epsilon}_{it,\text{OLS}}$$

$$\hat{\sigma}_{i,\text{OLS}}^2 = \frac{1}{T-1} \sum_{t=1}^T (\hat{\epsilon}_{it,\text{OLS}} - \bar{\epsilon}_{i,\text{OLS}})^2$$

Industry Factor Model: Covariance Matrix

- Let $\widehat{\mathbf{D}}_{\text{OLS}}$ be the diagonal matrix with $\widehat{\sigma}_{i,\text{OLS}}^2$ along the diagonal.
- The covariance matrix of the N assets is estimated by

$$\widehat{\mathbf{\Omega}}_{\text{OLS}} = \mathbf{B}\widehat{\mathbf{\Omega}}_{\text{OLS}}^F\mathbf{B}' + \widehat{\mathbf{D}}_{\text{OLS}}.$$

Weighted Least Squares Estimation

- ✧ The OLS estimation of the factor realizations \mathbf{f}_t is inefficient due to the cross-sectional heteroskedasticity in the asset returns.
- ✧ The estimates of the residual variances may be used as weights for weighted least squares (**feasible GLS**) estimation:

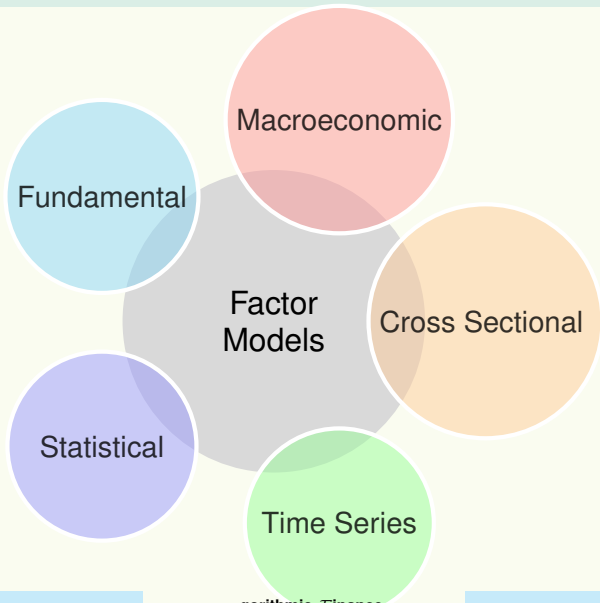
$$\hat{\mathbf{f}}_{t,\text{GLS}} = (\mathbf{B}' \hat{\mathbf{D}}_{\text{OLS}}^{-1} \mathbf{B})^{-1} \mathbf{B}' \hat{\mathbf{D}}_{\text{OLS}}^{-1} \mathbf{R}_t$$

$$\hat{\boldsymbol{\Omega}}_{\text{GLS}}^F = \frac{1}{T-1} \sum_{t=1}^T (\hat{\mathbf{f}}_{t,\text{GLS}} - \bar{\mathbf{f}}_{\text{GLS}}) (\hat{\mathbf{f}}_{t,\text{GLS}} - \bar{\mathbf{f}}_{\text{GLS}})'$$

$$\hat{\sigma}_{i,\text{GLS}}^2 = \frac{1}{T-1} \sum_{t=1}^T (\hat{\epsilon}_{it,\text{GLS}} - \bar{\epsilon}_{i,\text{GLS}})^2$$

$$\hat{\boldsymbol{\Omega}}_{\text{GLS}} = \mathbf{B} \hat{\boldsymbol{\Omega}}_{\text{GLS}}^F \mathbf{B}' + \hat{\mathbf{D}}_{\text{GLS}}.$$

Takeaways



References

- 🐟 Factor Models for Asset Returns, Eric Zivot (2011)
- 🐟 A Practitioner's Guide to Factor Models, The Research Foundation of the Institute of Chartered Financial Analysis (1994)
- 🐟 Using Macroeconomic Factors to Control Portfolio Risk, Edwin Burmeister, Richard Roll, and Stephen A. Ross (2003)
- 🐟 Introducing MSCI FaCS: A New Factor Classification Standard for Equity Portfolios, George Bonne, Leon Roisenberg, Raman Aylur Subramanian, Dimitris Melas (2018)