Factor Models for Asset Returns

Eric Zivot
University of Washington
BlackRock Alternative Advisors

March 14, 2011

Outline

1. Introduction
2. Factor Model Specification
3. Macroeconomic factor models
4. Fundamental factor models
5. Statistical factor models

Introduction

Factor models for asset returns are used to

- Decompose risk and return into explanable 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

1. Macroeconomic factor model
   (a) Factors are observable economic and financial time series
2. Fundamental factor model
   (a) Factors are created from observerable asset characteristics
3. Statistical factor model
   (a) Factors are unobservable and extracted from asset returns

Factor Model Specification

The three types of multifactor models for asset returns have the general form

$$\begin{aligned} R_{it} &= \alpha_i + \beta_{1i} f_{1t} + \beta_{2i} f_{2t} + \cdots + \beta_{Ki} f_{Kt} + \varepsilon_{it} \\ &= \alpha_i + \boldsymbol{\beta}_i' \mathbf{f}_t + \varepsilon_{it} \end{aligned} \quad (1)$$

- R_{it} is the simple return (real or in excess of the risk-free rate) on asset i ($i = 1, \ldots, N$) in time period t ($t = 1, \ldots, T$),
- f_{kt} is the k^{th} *common factor* ($k = 1, \ldots, K$),
- β_{ki} is the *factor loading* or *factor beta* for asset i on the k^{th} factor,
- ε_{it} is the asset *specific factor*.

Assumptions

1. The factor realizations, $\mathbf{f}_t$, are stationary with unconditional moments

$$\begin{aligned} E[\mathbf{f}_t] &= \boldsymbol{\mu}_f \\ cov(\mathbf{f}_t) &= E[(\mathbf{f}_t - \boldsymbol{\mu}_f)(\mathbf{f}_t - \boldsymbol{\mu}_f)'] = \boldsymbol{\Omega}_f \end{aligned}$$

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

$$cov(f_{kt}, \varepsilon_{it}) = 0, \text{ for all } k, \ i \text{ and } t.$$

3. Error terms ε_{it} are serially uncorrelated and contemporaneously uncorrelated across assets

$$\begin{aligned} cov(\varepsilon_{it}, \varepsilon_{js}) &= \sigma_i^2 \text{ for all } i = j \text{ and } t = s \\ &= 0, \text{ otherwise} \end{aligned}$$

Notation

Vectors with a subscript t represent the cross-section of all assets

$$\underset{(N\times 1)}{\mathbf{R}_t} = \begin{pmatrix} R_{1t} \\ \vdots \\ R_{Nt} \end{pmatrix}, \ t = 1, \ldots, T$$

Vectors with a subscript i represent the time series of a given asset

$$\underset{(T\times 1)}{\mathbf{R}_i} = \begin{pmatrix} R_{i1} \\ \vdots \\ R_{iT} \end{pmatrix}, \ i = 1, \ldots, N$$

Matrix of all assets over all time periods (columns = assets, rows = time period)

$$\underset{(T\times N)}{\mathbf{R}} = \begin{pmatrix} R_{11} & \cdots & R_{N1} \\ \vdots & \ddots & \vdots \\ R_{1T} & \cdots & R_{NT} \end{pmatrix}$$

Cross Section Regression

The multifactor model (1) may be rewritten as a *cross-sectional* regression model at time t by stacking the equations for each asset to give

$$\begin{aligned} \underset{(N\times 1)}{\mathbf{R}_t} &= \underset{(N\times 1)}{\boldsymbol{\alpha}} + \underset{(N\times K)}{\mathbf{B}}\underset{(K\times 1)}{\mathbf{f}_t} + \underset{(N\times 1)}{\boldsymbol{\varepsilon}_t},\ t = 1,\ldots,T \qquad (2) \\ \underset{(N\times K)}{\mathbf{B}} &= \begin{bmatrix} \boldsymbol{\beta}_1' \\ \vdots \\ \boldsymbol{\beta}_N' \end{bmatrix} = \begin{bmatrix} \beta_{11} & \cdots & \beta_{1K} \\ \vdots & \ddots & \vdots \\ \beta_{N1} & \cdots & \beta_{NK} \end{bmatrix} \\ E[\boldsymbol{\varepsilon}_t\boldsymbol{\varepsilon}_t'|\mathbf{f}_t] &= \mathbf{D} = diag(\sigma_1^2,\ldots,\sigma_N^2) \end{aligned}$$

Note: Cross-sectional heteroskedasticity

Time Series Regression

The multifactor model (1) may also be rewritten as a *time-series* regression model for asset i by stacking observations for a given asset i to give

$$\begin{aligned} \underset{(T\times 1)}{\mathbf{R}_i} &= \underset{(T\times 1)}{\mathbf{1}_T}\underset{(1\times 1)}{\alpha_i} + \underset{(T\times K)}{\mathbf{F}}\underset{(K\times 1)}{\boldsymbol{\beta}_i} + \underset{(T\times 1)}{\boldsymbol{\varepsilon}_i},\ i = 1,\ldots,N \qquad (3) \\ \underset{(T\times K)}{\mathbf{F}} &= \begin{bmatrix} \mathbf{f}_1' \\ \vdots \\ \mathbf{f}_T' \end{bmatrix} = \begin{bmatrix} f_{11} & \cdots & f_{Kt} \\ \vdots & \ddots & \vdots \\ f_{1T} & \cdots & f_{KT} \end{bmatrix} \\ E[\boldsymbol{\varepsilon}_i\boldsymbol{\varepsilon}_i'] &= \sigma_i^2\mathbf{I}_T \end{aligned}$$

Note: Time series homoskedasticity

Multivariate Regression

Collecting data from $i = 1, \ldots, N$ allows the model (3) to be expressed as the multivariate regression

$$[\mathbf{R}_1, \ldots, \mathbf{R}_N] = \mathbf{1}_T[\alpha_1, \ldots, \alpha_N] + \mathbf{F}[\boldsymbol{\beta}_1, \ldots, \boldsymbol{\beta}_N] + [\varepsilon_1, \ldots, \varepsilon_N]$$

or

$$\begin{aligned}
\underset{(T\times N)}{\mathbf{R}} &= \underset{(T\times 1)}{\mathbf{1}_T}\underset{(1\times T)}{\boldsymbol{\alpha}'} + \underset{(T\times K)}{\mathbf{F}}\underset{(K\times N)}{\mathbf{B}'} + \underset{(T\times N)}{\mathbf{E}} \\
&= \mathbf{X}\boldsymbol{\Gamma}' + \mathbf{E} \\
\underset{(T\times(K+1))}{\mathbf{X}} &= [\mathbf{1}_T \vdots \mathbf{F}], \quad \underset{((K+1)\times N)}{\boldsymbol{\Gamma}'} = \begin{bmatrix} \boldsymbol{\alpha}' \\ \mathbf{B}' \end{bmatrix},
\end{aligned}$$

Alternatively, collecting data from $t = 1, \ldots, T$ allows the model (2) to be expressed as the multivariate regression

$$[\mathbf{R}_1, \ldots, \mathbf{R}_T] = [\boldsymbol{\alpha}, \ldots, \boldsymbol{\alpha}] + \mathbf{B}[\mathbf{f}_1, \ldots, \mathbf{f}_T] + [\varepsilon_1, \ldots, \varepsilon_T]$$

or

$$\begin{aligned}
\underset{(N\times T)}{\mathbf{R}'} &= \underset{(N\times 1)}{\boldsymbol{\alpha}}\underset{(1\times T)}{\mathbf{1}'_T} + \underset{(N\times K)}{\mathbf{B}}\underset{(K\times T)}{\mathbf{F}'} + \underset{(N\times T)}{\mathbf{E}'} \\
&= \boldsymbol{\Gamma}\mathbf{X}' + \mathbf{E}' \\
\underset{((K+1)\times T)}{\mathbf{X}'} &= \begin{bmatrix} \mathbf{1}'_T \\ \mathbf{F}' \end{bmatrix}, \quad \underset{(N\times(K+1))}{\boldsymbol{\Gamma}} = [\boldsymbol{\alpha} \vdots \mathbf{B}],
\end{aligned}$$

Expected Return ($\alpha - \beta$) Decomposition

$$E[R_{it}] = \alpha_i + \boldsymbol{\beta}_i' E[\mathbf{f}_t]$$

- $\boldsymbol{\beta}_i' E[\mathbf{f}_t] =$ explained expected return due to systematic risk factors
- $\alpha_i = E[R_{it}] - \boldsymbol{\beta}_i' E[\mathbf{f}_t] =$ unexplained expected return (abnormal return)

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

Covariance Structure

Using the cross-section regression

$$\underset{(N\times 1)}{\mathbf{R}_t} = \underset{(N\times 1)}{\boldsymbol{\alpha}} + \underset{(N\times K)}{\mathbf{B}} \underset{(K\times 1)}{\mathbf{f}_t} + \underset{(N\times 1)}{\boldsymbol{\varepsilon}_t}, \; t = 1, \ldots, T$$

and the assumptions of the multifactor model, the $(N \times N)$ covariance matrix of asset returns has the form

$$cov(\mathbf{R}_t) = \boldsymbol{\Omega}_{FM} = \mathbf{B}\boldsymbol{\Omega}_f\mathbf{B}' + \mathbf{D} \qquad (4)$$

Note, (4) implies that

$$\begin{aligned} var(R_{it}) &= \boldsymbol{\beta}_i'\boldsymbol{\Omega}_f\boldsymbol{\beta}_i + \sigma_i^2 \\ cov(R_{it}, R_{jt}) &= \boldsymbol{\beta}_i'\boldsymbol{\Omega}_f\boldsymbol{\beta}_j \end{aligned}$$

Portfolio Analysis

Let $\mathbf{w} = (w_1, \ldots, w_n)$ be a vector of portfolio weights (w_i = fraction of wealth in asset i). If $\mathbf{R}_t$ is the $(N \times 1)$ vector of simple returns then

$$R_{p,t} = \mathbf{w}'\mathbf{R}_t = \sum_{i=1}^{N} w_i R_{it}$$

Portfolio Factor Model

$$\begin{aligned}
\mathbf{R}_t &= \boldsymbol{\alpha} + \mathbf{B}\mathbf{f}_t + \boldsymbol{\varepsilon}_t \Rightarrow \\
R_{p,t} &= \mathbf{w}'\boldsymbol{\alpha} + \mathbf{w}'\mathbf{B}\mathbf{f}_t + \mathbf{w}'\boldsymbol{\varepsilon}_t = \alpha_p + \boldsymbol{\beta}_p'\mathbf{f}_t + \varepsilon_{p,t} \\
\alpha_p &= \mathbf{w}'\boldsymbol{\alpha},\ \boldsymbol{\beta}_p' = \mathbf{w}'\mathbf{B},\ \varepsilon_{p,t} = \mathbf{w}'\boldsymbol{\varepsilon}_t \\
var(R_{p,t}) &= \boldsymbol{\beta}_p'\Omega_f\boldsymbol{\beta}_p + var(\varepsilon_{p,t}) = \mathbf{w}'\mathbf{B}\boldsymbol{\Omega}_f\mathbf{B}'\mathbf{w} + \mathbf{w}'\mathbf{D}\mathbf{w}
\end{aligned}$$

Active and Static Portfolios

- 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

Macroeconomic Factor Models

$$
\begin{aligned}
R_{it} &= \alpha_i + \boldsymbol{\beta}_i' \mathbf{f}_t + \varepsilon_{it} \\
\mathbf{f}_t &= \text{observed economic/financial time series}
\end{aligned}
$$

Econometric problems:

- Choice of factors
- Estimate factor betas, $\boldsymbol{\beta}_i$, and residual variances, σ_i^2, using time series regression techniques.
- Estimate factor covariance matrix, $\boldsymbol{\Omega}_f$, from observed history of factors

Shape's Single Factor Model

Sharpe's single factor model is a macroeconomic factor model with a single market factor:

$$
R_{it} = \alpha_i + \beta_i R_{Mt} + \varepsilon_{it},\ i = 1, \ldots, N; t = 1, \ldots, T \tag{5}
$$

where R_{Mt} denotes the return or excess return (relative to the risk-free rate) on a market index (typically a value weighted index like the S&P 500 index) in time period t.

Risk-adusted expected return and abnormal return

$$
\begin{aligned}
E[R_{it}] &= \beta_i E[R_{Mt}] \\
\alpha_i &= E[R_{it}] - \beta_i E[R_{Mt}]
\end{aligned}
$$

Covariance matrix of assets

$$\mathbf{\Omega}_{FM} = \sigma_M^2 \boldsymbol{\beta}\boldsymbol{\beta}' + \mathbf{D} \tag{6}$$

where

$$\begin{aligned} \sigma_M^2 &= var(R_{Mt}) \\ \boldsymbol{\beta} &= (\beta_1, \ldots, \beta_N)' \\ \mathbf{D} &= diag(\sigma_1^2, \ldots, \sigma_N^2), \\ \sigma_i^2 &= var(\varepsilon_{it}) \end{aligned}$$

Estimation

Because R_{Mt} is observable, the parameters β_i and σ_i^2 of the single factor model (5) for each asset can be estimated using time series regression (i.e., ordinary least squares) giving

$$\begin{aligned} \mathbf{R}_i &= \widehat{\alpha}_i \mathbf{1}_T + \mathbf{R}_M \widehat{\beta}_i + \widehat{\boldsymbol{\varepsilon}}_i, \ i = 1, \ldots, N \\ \widehat{\beta}_i &= \widehat{cov}(R_{it}, R_{Mt}) / \widehat{var}(R_{Mt}) = \hat{\sigma}_{iM} / \hat{\sigma}_M^2 \\ \widehat{\alpha}_i &= \bar{R}_i - \hat{\beta}_i \bar{R}_M \\ \widehat{\sigma}_i^2 &= \frac{1}{T-2} \widehat{\boldsymbol{\varepsilon}}_i' \widehat{\boldsymbol{\varepsilon}}_i \end{aligned}$$

The estimated single factor model covariance matrix is

$$\widehat{\mathbf{\Omega}}_{FM} = \widehat{\sigma}_M^2 \widehat{\boldsymbol{\beta}} \widehat{\boldsymbol{\beta}}' + \widehat{\mathbf{D}}$$

Remarks

1. Computational efficiency may be obtained by using multivariate regression. The coefficients α_i and β_i and the residual variances σ_i^2 may be computed in one step in the multivariate regression model

$$\mathbf{R} = \mathbf{X}\mathbf{\Gamma}' + \mathbf{E}$$

The multivariate OLS estimator of $\mathbf{\Gamma}'$ is

$$\widehat{\mathbf{\Gamma}}' = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{R}'.$$

The estimate of the residual covariance matrix is

$$\widehat{\mathbf{\Sigma}} = \frac{1}{T-2}\widehat{\mathbf{E}}'\widehat{\mathbf{E}}$$

where $\hat{\mathbf{E}} = \mathbf{R} - \mathbf{X}\hat{\mathbf{\Gamma}}'$ is the multivariate least squares residual matrix. The diagonal elements of $\widehat{\mathbf{\Sigma}}$ are the diagonal elements of $\widehat{\mathbf{D}}$.

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. Additionally, $\widehat{\sigma}_i$ is a measure of the typical size of asset specific risk. Given the variance decomposition

$$var(R_{it}) = \beta_i^2 var(R_{Mt}) + var(\varepsilon_{it}) = \beta_i^2\sigma_M^2 + \sigma_i^2$$

R^2 can be estimated using

$$R^2 = \frac{\hat{\beta}_i^2\hat{\sigma}_M^2}{\widehat{var}(R_{it})}$$

3. Robust regression techniques can be used to estimate β_i and σ_i^2. Also, a robust estimate of σ_M^2 could be computed.

4. The single factor covariance matrix (6) is constant over time. This may not be a good assumption. There are several ways to allow (6) to vary over time. In general, β_i, σ_i^2 and σ_M^2 can be time varying. That is,

$$\beta_i = \beta_{it},\ \sigma_i^2 = \sigma_{it}^2,\ \sigma_M^2 = \sigma_{Mt}^2.$$

To capture time varying betas, rolling regression or Kalman filter techniques could be used. To capture conditional heteroskedasticity, GARCH models may be used for σ_{it}^2 and σ_{Mt}^2. One may also use exponential weights in computing estimates of β_{it}, σ_{it}^2 and σ_{Mt}^2. A time varying factor model covariance matrix is

$$\widehat{\mathbf{\Omega}}_{FM,t} = \widehat{\sigma}_{Mt}^2 \widehat{\boldsymbol{\beta}}_t \widehat{\boldsymbol{\beta}}_t' + \widehat{\mathbf{D}}_t,$$

General Multi-factor Model

Model specifies K observable macro-variables

$$R_{it} = \alpha_i + \boldsymbol{\beta}_i' \mathbf{f}_t + \varepsilon_{it}$$

- Chen, Roll and Ross (1986) provides a description of commonly used macroeconomic factors for equity. Lo (2008) discusses hedge funds.

- 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

Because the factor realizations are observable, the parameter matrices $\mathbf{B}$ and $\mathbf{D}$ of the model may be estimated using time series regression:

$$\begin{aligned}
\mathbf{R}_i &= \widehat{\alpha}_i \mathbf{1}_T + \mathbf{F}\widehat{\boldsymbol{\beta}}_i + \widehat{\varepsilon}_i = \mathbf{X}\hat{\gamma} + \widehat{\varepsilon}_i,\ i = 1, \ldots, N \\
\mathbf{X} &= [\mathbf{1}_T \vdots \mathbf{F}],\ \hat{\gamma} = (\hat{\alpha}_i, \hat{\boldsymbol{\beta}}_i')' = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{R}_i \\
\widehat{\sigma}_i^2 &= \frac{1}{T - K - 1}\widehat{\varepsilon}_i'\widehat{\varepsilon}_i
\end{aligned}$$

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

$$\widehat{\boldsymbol{\Omega}}_f = \frac{1}{T-1}\sum_{t=1}^{T}(\mathbf{f}_t - \bar{\mathbf{f}})(\mathbf{f}_t - \bar{\mathbf{f}})',\ \bar{\mathbf{f}} = \frac{1}{T}\sum_{t=1}^{T}\mathbf{f}_t$$

The estimated multifactor model covariance matrix is then

$$\widehat{\boldsymbol{\Omega}}_{FM} = \widehat{\mathbf{B}}\widehat{\boldsymbol{\Omega}}_f\widehat{\mathbf{B}}' + \widehat{\mathbf{D}} \tag{7}$$

Remarks

1. As with the single factor model, robust regression may be used to compute $\boldsymbol{\beta}_i$ and σ_i^2. A robust covariance matrix estimator may also be used to compute and estimate of $\boldsymbol{\Omega}_f$.

2. $\boldsymbol{\Omega}_{FM}$ can be made time varying by allowing $\boldsymbol{\beta}_i$, $\boldsymbol{\Omega}_f$ and σ_i^2 $(i = 1, \ldots, N)$ to be time varying

Example: Estimation of Single Index Model in R using investment data from Berndt (1991).

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.

- Factor betas are constructed from observable asset characteristics (i.e., $\mathbf{B}$ is known)
- Factor realizations, $\mathbf{f}_t$, are estimated/constructed for each t given $\mathbf{B}$
- In practice, fundamental factor models are estimated in two ways.

BARRA Approach

- This approach was pioneered by Bar Rosenberg, founder of BARRA Inc., and is discussed at length in Grinold and Kahn (2000), Conner et al (2010), and Cariño et al (2010).

- In this approach, the observable asset specific fundamentals (or some transformation of them) are treated as the factor betas, $\boldsymbol{\beta}_i$, which are time invariant.

- The factor realizations at time t, $\mathbf{f}_t$, are unobservable. The econometric problem is then to estimate the factor realizations at time t given the factor betas. This is done by running T cross-section regressions.

Fama-French Approach

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

- For a given observed asset specific characteristic, e.g. size, they determined factor realizations using a two step process. First they sorted the cross-section of assets based on the values of the asset specific characteristic. Then they formed a hedge portfolio which is long in the top quintile of the sorted assets and short in the bottom quintile of the sorted assets. The observed return on this hedge portfolio at time t is the observed factor realization for the asset specific characteristic. This process is repeated for each asset specific characteristic.

- Given the observed factor realizations for $t = 1, \ldots, T$, the factor betas for each asset are estimated using N time series regressions.

BARRA-type Single Factor Model

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

$$\underset{(N\times 1)}{\mathbf{R}_t} = \underset{(N\times 1)(1\times 1)}{\boldsymbol{\beta} \quad f_t} + \underset{(N\times 1)}{\boldsymbol{\varepsilon}_t}, t = 1, \ldots, T$$

- $\boldsymbol{\beta}$ 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.

- $var(f_t) = \sigma_f^2$; $cov(f_t, \varepsilon_{it}) = 0$, for all i, t; $var(\varepsilon_{it}) = \sigma_i^2, i = 1, \ldots, N$.

Estimation

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

$$\begin{aligned} \hat{f}_{t,wls} &= (\boldsymbol{\beta}'\mathbf{D}^{-1}\boldsymbol{\beta})^{-1}\boldsymbol{\beta}'\mathbf{D}^{-1}\mathbf{R}_t, \ t = 1, \ldots, T \\ \mathbf{D} &= diag(\sigma_1^2, \ldots, \sigma_N^2) \end{aligned} \tag{8}$$

Note 1: σ_i^2 can be consistently estimated and a feasible WLS estimate can be computed

$$\begin{aligned} \hat{f}_{t,fwls} &= (\boldsymbol{\beta}'\hat{\mathbf{D}}^{-1}\boldsymbol{\beta})^{-1}\boldsymbol{\beta}'\hat{\mathbf{D}}^{-1}\mathbf{R}_t, \ t = 1, \ldots, T \\ \hat{\mathbf{D}} &= diag(\hat{\sigma}_1^2, \ldots, \hat{\sigma}_N^2) \end{aligned}$$

Note 2: Other weights besides $\hat{\sigma}_i^2$ could be used

Factor Mimicking Portfolio

The WLS estimate of f_t in (8) has an interesting interpretation as the return on a portfolio $\mathbf{h} = (h_1, \ldots, h_N)'$ that solves

$$\min_{\mathbf{h}} \frac{1}{2}\mathbf{h}'\mathbf{D}\mathbf{h} \text{ subject to } \mathbf{h}'\boldsymbol{\beta} = 1$$

The portfolio $\mathbf{h}$ minimizes asset return residual variance subject to having unit exposure to the attribute β 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

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

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 K mutually exclusive industries. The factor sensitivities β_{ik} in (1) for each asset are time invariant and of the form

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

and f_{kt} represents the factor realization for the k^{th} industry in time period t.

- The factor betas are dummy variables indicating whether a given asset is in a particular industry.

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

Industry Factor Model Regression

The industry factor model with K industries is summarized as

$$\begin{aligned} R_{it} &= \beta_{i1} f_{1t} + \cdots + \beta_{iK} f_{Kt} + \varepsilon_{it},\ i = 1, \ldots, N; t = 1, \ldots, T \\ \text{var}(\varepsilon_{it}) &= \sigma_i^2,\ i = 1, \ldots, N \\ \text{cov}(\varepsilon_{it}, f_{jt}) &= 0,\ j = 1, \ldots, K;\ i = 1, \ldots, N \\ \text{cov}(f_{it}, f_{jt}) &= \sigma_{ij}^f,\ i, j = 1, \ldots, K \end{aligned}$$

where

$$\begin{aligned} \beta_{ik} &= 1 \text{ if asset } i \text{ is in industry k } (k = 1, \ldots, K) \\ &= 0, \text{ otherwise} \end{aligned}$$

It is assumed that there are N_k firms in the kth industry such $\sum_{k=1}^{K} N_k = N$.

Estimation of Industry Factor Model Factors

Consider the cross-section regression at time t

$$\begin{aligned} \mathbf{R}_t &= \boldsymbol{\beta}_1 f_{1t} + \cdots + \boldsymbol{\beta}_K f_{Kt} + \boldsymbol{\varepsilon}_t, \\ &= \mathbf{B}\mathbf{f}_t + \boldsymbol{\varepsilon}_t \\ E[\boldsymbol{\varepsilon}_t \boldsymbol{\varepsilon}_t'] &= \mathbf{D}, \text{cov}(\mathbf{f}_t) = \boldsymbol{\Omega}_f \end{aligned}$$

Since the industries are mutually exclusive it follows that

$$\boldsymbol{\beta}_j' \boldsymbol{\beta}_k = N_k \text{ for } j = k,\ 0 \text{ otherwise}$$

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

$$\widehat{\mathbf{f}}_{t,\text{OLS}} = (\mathbf{B}'\mathbf{B})^{-1}\mathbf{B}'\mathbf{R}_t = \begin{pmatrix} \widehat{f}_{1t,\text{OLS}} \\ \vdots \\ \widehat{f}_{Kt,\text{OLS}} \end{pmatrix} = \begin{pmatrix} \frac{1}{N_1} \sum_{i=1}^{N_1} R_{it}^1 \\ \vdots \\ \frac{1}{N_K} \sum_{i=1}^{N_K} R_{it}^K \end{pmatrix}$$

Estimation of Factor Realization Covariance Matrix

Given $(\widehat{\mathbf{f}}_{1,\mathsf{OLS}}, \ldots, \widehat{\mathbf{f}}_{T,\mathsf{OLS}})$, the covariance matrix of the industry factors may be computed as the time series sample covariance

$$\begin{aligned}
\widehat{\boldsymbol{\Omega}}^F_{\mathsf{OLS}} &= \frac{1}{T-1}\sum_{t=1}^{T}(\widehat{\mathbf{f}}_{t,\mathsf{OLS}} - \overline{\mathbf{f}}_{\mathsf{OLS}})(\widehat{\mathbf{f}}_{t,\mathsf{OLS}} - \overline{\mathbf{f}}_{\mathsf{OLS}})', \\
\overline{\mathbf{f}}_{\mathsf{OLS}} &= \frac{1}{T}\sum_{t=1}^{T}\widehat{\mathbf{f}}_{t,\mathsf{OLS}}
\end{aligned}$$

Estimation of Residual Variances

The residual variances, $\mathrm{var}(\varepsilon_{it}) = \sigma_i^2$, can be estimated from the time series of residuals from the T cross-section regressions as follows. Let $\widehat{\boldsymbol{\varepsilon}}_{t,\mathsf{OLS}}$, $t = 1, \ldots, T$, denote the $(N \times 1)$ vector of OLS residuals, and let $\widehat{\varepsilon}_{it,\mathsf{OLS}}$ denote the i^{th} row of $\widehat{\boldsymbol{\varepsilon}}_{t,\mathsf{OLS}}$. Then σ_i^2 may be estimated using

$$\begin{aligned}
\widehat{\sigma}^2_{i,\mathsf{OLS}} &= \frac{1}{T-1}\sum_{t=1}^{T}(\widehat{\varepsilon}_{it,\mathsf{OLS}} - \overline{\varepsilon}_{i,\mathsf{OLS}})^2,\ i = 1, \ldots, N \\
\overline{\varepsilon}_{i,\mathsf{OLS}} &= \frac{1}{T}\sum_{t=1}^{T}\widehat{\varepsilon}_{it,\mathsf{OLS}}
\end{aligned}$$

Estimation of Industry Factor Model Asset Return Covariance Matrix

The covariance matrix of the N assets is estimated using

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

where $\widehat{\mathbf{D}}_{\mathsf{OLS}}$ is a diagonal matrix with $\widehat{\sigma}^2_{i,\mathsf{OLS}}$ along the diagonal.

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:

$$\begin{aligned}
\widehat{\mathbf{f}}_{t,\mathsf{GLS}} &= (\mathbf{B}'\widehat{\mathbf{D}}^{-1}_{\mathsf{OLS}}\mathbf{B})^{-1}\mathbf{B}'\widehat{\mathbf{D}}^{-1}_{\mathsf{OLS}}\mathbf{R}_t,\ t = 1, \ldots, T \\
\widehat{\mathbf{\Omega}}^{F}_{\mathsf{GLS}} &= \frac{1}{T-1}\sum_{t=1}^{T}(\widehat{\mathbf{f}}_{t,\mathsf{GLS}} - \overline{\mathbf{f}}_{\mathsf{GLS}})(\widehat{\mathbf{f}}_{t,\mathsf{GLS}} - \overline{\mathbf{f}}_{\mathsf{GLS}})' \\
\widehat{\sigma}^2_{i,\mathsf{GLS}} &= \frac{1}{T-1}\sum_{t=1}^{T}(\widehat{\varepsilon}_{it,\mathsf{GLS}} - \overline{\varepsilon}_{i,\mathsf{GLS}})^2,\ i = 1, \ldots, N \\
\widehat{\mathbf{\Omega}}_{\mathsf{GLS}} &= \mathbf{B}\widehat{\mathbf{\Omega}}^{F}_{\mathsf{GLS}}\mathbf{B}' + \widehat{\mathbf{D}}_{\mathsf{GLS}}
\end{aligned}$$

Statistical Factor Models for Returns

- In statistical factor models, the factor realizations $\mathbf{f}_t$ in (1) are not directly observable and must be extracted from the observable returns $\mathbf{R}_t$ using statistical methods. The primary methods are *factor analysis* and *principal components analysis*.

- Traditional factor analysis and principal component analysis are usually applied to extract the factor realizations if the number of time series observations, T, is greater than the number of assets, N.

- If $N > T$, then the sample covariance matrix of returns becomes singular which complicates traditional factor and principal components analysis. In this case, the method of *asymptotic principal component* analysis is more appropriate.

Sample Covariance Matrices

Traditional factor and principal component analysis is based on the $(N \times N)$ sample covariance matrix

$$\underset{(N\times N)}{\widehat{\mathbf{\Omega}}_N} = \frac{1}{T}\mathbf{R}'\mathbf{R}$$

where $\mathbf{R}$ is the $(N \times T)$ matrix of observed returns.

Asymptotic principal component analysis is based on the $(T \times T)$ covariance matrix

$$\underset{(T\times T)}{\widehat{\mathbf{\Omega}}_T} = \frac{1}{N}\mathbf{R}\mathbf{R}'.$$

Factor Analysis

Traditional factor analysis assumes a time invariant *orthogonal factor structure*

$$\begin{aligned}
\underset{(N\times 1)}{\mathbf{R}_t} &= \underset{(N\times 1)}{\boldsymbol{\mu}} + \underset{(N\times K)}{\mathbf{B}}\underset{(K\times 1)}{\mathbf{f}_t} + \underset{(N\times 1)}{\boldsymbol{\varepsilon}_t} \qquad (9)\\
cov(\mathbf{f}_t, \boldsymbol{\varepsilon}_s) &= 0, \text{ for all } t, s\\
E[\mathbf{f}_t] &= E[\boldsymbol{\varepsilon}_t] = 0\\
var(\mathbf{f}_t) &= \mathbf{I}_K\\
var(\boldsymbol{\varepsilon}_t) &= \mathbf{D}
\end{aligned}$$

where $\mathbf{D}$ is a diagonal matrix with σ_i^2 along the diagonal. Then, the return covariance matrix, $\boldsymbol{\Omega}$, may be decomposed as

$$\boldsymbol{\Omega} = \mathbf{B}\mathbf{B}' + \mathbf{D}$$

Hence, the K common factors $\mathbf{f}_t$ account for all of the cross covariances of asset returns.

Variance Decomposition

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

$$var(R_{it}) = \sum_{j=1}^{K} \beta_{ij}^2 + \sigma_i^2$$

- variance portion due to common factors, $\sum_{j=1}^{K} \beta_{ij}^2$, is called the *communality*,
- variance portion due to specific factors, σ_i^2, is called the *uniqueness*.

Non-Uniqueness of Factors and Loadings

The orthogonal factor model (9) does not uniquely identify the common factors $\mathbf{f}_t$ and factor loadings $\mathbf{B}$ since for any orthogonal matrix $\mathbf{H}$ such that $\mathbf{H}' = \mathbf{H}^{-1}$

$$\begin{aligned} \mathbf{R}_t &= \boldsymbol{\mu} + \mathbf{B}\mathbf{H}\mathbf{H}'\mathbf{f}_t + \varepsilon_t \\ &= \mu + \mathbf{B}^*\mathbf{f}_t^* + \varepsilon_t \end{aligned}$$

where $\mathbf{B}^* = \mathbf{BH}$, $\mathbf{f}_t^* = \mathbf{H}'\mathbf{f}_t$ and $var(\mathbf{f}_t^*) = \mathbf{I}_K$.

Because the factors and factor loadings are only identified up to an orthogonal transformation (rotation of coordinates), the interpretation of the factors may not be apparent until suitable rotation is chosen.

Estimation

Estimation using factor analysis consists of three steps:

- Estimation of the factor loading matrix $\mathbf{B}$ and the residual covariance matrix $\mathbf{D}$.

- Construction of the factor realizations $\mathbf{f}_t$.

- Rotation of coordinate system to enhance interpretation

Maximum Likelihood Estimation of $\mathbf{B}$ and $\mathbf{D}$

Maximum likelihood estimation of $\mathbf{B}$ and $\mathbf{D}$ is performed under the assumption that returns are jointly normally distributed and temporally iid.

Given estimates $\widehat{\mathbf{B}}$ and $\widehat{\mathbf{D}}$, an empirical version of the factor model (2) may be constructed as

$$\mathbf{R}_t - \widehat{\boldsymbol{\mu}} = \widehat{\mathbf{B}}\mathbf{f}_t + \widehat{\boldsymbol{\varepsilon}}_t \tag{10}$$

where $\widehat{\boldsymbol{\mu}}$ is the sample mean vector of $\mathbf{R}_t$. The error terms in (10) are heteroskedastic so that OLS estimation is inefficient.

Estimation of Factor Realizations $\hat{\mathbf{f}}_t$

Using (10), the factor realizations in a given time period t, $\mathbf{f}_t$, can be estimated using the cross-sectional feasible weighted least squares (FWLS) regression

$$\widehat{\mathbf{f}}_{t,fwls} = (\widehat{\mathbf{B}}'\widehat{\mathbf{D}}^{-1}\widehat{\mathbf{B}})^{-1}\widehat{\mathbf{B}}'\widehat{\mathbf{D}}^{-1}(\mathbf{R}_t - \widehat{\boldsymbol{\mu}}) \tag{11}$$

Performing this regression for $t = 1, \ldots, T$ times gives the time series of factor realizations $(\widehat{\mathbf{f}}_1, \ldots, \widehat{\mathbf{f}}_T)$.

The factor model estimated covariance matrix is then given by

$$\widehat{\boldsymbol{\Omega}}^F = \widehat{\mathbf{B}}\widehat{\mathbf{B}}' + \widehat{\mathbf{D}}$$

Tests for the Number of Factors

Using the maximum likelihood estimates of $\mathbf{B}$ and $\mathbf{D}$ based on a $K-$factor model and the sample covariance matrix $\widehat{\mathbf{\Omega}}$, a likelihood ratio test (modified for improved small sample performance) of the adequacy of K factors is of the form

$$LR(K) = -(T - 1 - \frac{1}{6}(2N + 5) - \frac{2}{3}K) \cdot \left(\ln|\widehat{\mathbf{\Omega}}| - \ln|\widehat{\mathbf{B}}\widehat{\mathbf{B}}' + \widehat{\mathbf{D}}|\right).$$

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

Remarks:

- Traditional factor analysis starts with a $\sqrt{T}$- consistent and asymptotically normal estimator of $\mathbf{\Omega}$, usually the sample covariance matrix $\widehat{\mathbf{\Omega}}$, and makes inference on K based on $\widehat{\mathbf{\Omega}}$. A likelihood ratio test is often used to select K under the assumption that ε_{it} is normally distributed (see below). However, when $N \to \infty$ consistent estimation of $\mathbf{\Omega}$, an $N \times N$ matrix, is not a well defined problem. Hence, if N is large relative to T, then traditional factor analysis may run into problems. Additionally, typical algorithms for factor analysis are not efficient for very large problems.

- Traditional factor analysis is only appropriate if ε_{it} is cross-sectionally uncorrelated, serially uncorrelated, and serially homoskedastic.

Principal Components

- Principal component analysis (PCA) is a dimension reduction technique used to explain the majority of the information in the sample covariance matrix of returns.

- With N assets there are N principal components, and these principal components are just linear combinations of the returns.

- The principal components are constructed and ordered so that the first principal component explains the largest portion of the sample covariance matrix of returns, the second principal component explains the next largest portion, and so on. The principal components are constructed to be orthogonal to each other and to be normalized to have unit length.

- In terms of a multifactor model, the K most important principal components are the factor realizations. The factor loadings on these observed factors can then be estimated using regression techniques.

Population Principal Components

Let $\mathbf{R}_t$ denote the $N \times 1$ vector of returns with $N \times N$ covariance matrix $\boldsymbol{\Omega} = E[(\mathbf{R}_t - E[\mathbf{R}_t])(\mathbf{R}_t - E[\mathbf{R}_t])']$. Consider creating factors from the linear combinations (portfolios) of returns

$$
\begin{aligned}
f_{1t} &= \mathbf{p}_1'\mathbf{R}_t = p_{11}R_{1t} + \cdots + p_{1N}R_{Nt} \\
f_{2t} &= \mathbf{p}_2'\mathbf{R}_t = p_{21}R_{1t} + \cdots + p_{2N}R_{Nt} \\
&\vdots \\
f_{Nt} &= \mathbf{p}_N'\mathbf{R}_t = p_{N1}R_{1t} + \cdots + p_{NN}R_{Nt}
\end{aligned}
$$

Note:

$$
var(f_{kt}) = \mathbf{p}_k'\boldsymbol{\Omega}\mathbf{p}_k,\ cov(f_{jt}, f_{kt}) = \mathbf{p}_j'\boldsymbol{\Omega}\mathbf{p}_k
$$

The principal components are those uncorrelated factors $f_{1t}, f_{2t}, \ldots, f_{Nt}$ whose variances are as large as possible.

Extracting Principal Components

The first population principal component is $\mathbf{p}_1^{*\prime}\mathbf{R}_t$ where the $(N \times 1)$ vector $\mathbf{p}_1^*$ solves

$$
\max_{p_1}\ \mathbf{p}_1'\boldsymbol{\Omega}\mathbf{p}_1 \text{ s.t. } \mathbf{p}_1'\mathbf{p}_1 = 1.
$$

The solution $\mathbf{p}_1^*$ is the eigenvector associated with the largest eigenvalue of $\boldsymbol{\Omega}$.

The second principal component is $\mathbf{p}_2^{*\prime}\mathbf{R}_t$ where the $(N \times 1)$ vector $\mathbf{p}_2^*$ solves

$$
\max_{p_2}\ \mathbf{p}_2'\boldsymbol{\Omega}\mathbf{p}_2 \text{ s.t. } \mathbf{p}_2'\mathbf{p}_2 = 1 \text{ and } \mathbf{p}_1^{*\prime}\mathbf{p}_2 = 0
$$

The solution $\mathbf{p}_2^*$ is the eigenvector associated with the second largest eigenvalue of $\boldsymbol{\Omega}$. This process is repeated until all N principal components are computed.

Spectral (Eigenvalue) Decomposition of Ω

$$\begin{aligned}
\mathbf{\Omega} &= \mathbf{P}\mathbf{\Lambda}\mathbf{P}' \\
\underset{(N\times N)}{\mathbf{P}} &= [\mathbf{p}_1^* \vdots \mathbf{p}_2^* \vdots \cdots \vdots \mathbf{p}_N^*],\ \mathbf{P}' = \mathbf{P}^{-1} \\
\mathbf{\Lambda} &= diag(\lambda_1, \ldots, \lambda_N),\ \lambda_1 > \lambda_2 > \cdots > \lambda_N
\end{aligned}$$

- $\mathbf{P}$ is the orthonormal maxtrix of eigen-vectors

- $\mathbf{\Lambda}$ is the diaganol matrix of ordered eigen-values

Variance Decomposition

$$\sum_{i=1}^{N} var(R_{it}) = \sum_{i=1}^{N} var(f_{it}) = \sum_{i=1}^{N} \lambda_i$$

where λ_i are the ordered eigenvalues of $var(\mathbf{R}_t) = \mathbf{\Omega}$. Therefore, the ratio

$$\frac{\lambda_i}{\sum_{i=1}^{N} \lambda_i}$$

gives the proportion of the total variance $\sum_{i=1}^{N} var(R_{it})$ attributed to the ith principal component factor return, and the ratio

$$\frac{\sum_{i=1}^{K} \lambda_i}{\sum_{i=1}^{N} \lambda_i}$$

gives the cumulative variance explained. Examination of these ratios help in determining the number of factors to use to explain the covariance structure of returns.

Sample Principal Components and Estimated Factors

Sample principal components are computed from the spectral decompositon of the $N \times N$ sample covariance matrix $\hat{\mathbf{\Omega}}_N$ when $N < T$:

$$\begin{aligned}
\hat{\mathbf{\Omega}}_N &= \hat{\mathbf{P}}\hat{\mathbf{\Lambda}}\hat{\mathbf{P}}' \\
\underset{(N\times N)}{\hat{\mathbf{P}}} &= [\hat{\mathbf{p}}_1^* \vdots \hat{\mathbf{p}}_2^* \vdots \cdots \vdots \hat{\mathbf{p}}_N^*],\ \hat{\mathbf{P}}' = \hat{\mathbf{P}}^{-1} \\
\hat{\mathbf{\Lambda}} &= diag(\hat{\lambda}_1, \ldots, \hat{\lambda}_N),\ \hat{\lambda}_1 > \hat{\lambda}_2 > \cdots > \hat{\lambda}_N
\end{aligned}$$

The estimated factor realizations are simply the first K sample principal components

$$\begin{aligned}
\widehat{f}_{kt} &= \hat{\mathbf{p}}_k^{*\prime}\mathbf{R}_t,\ k = 1, \ldots, K. \\
\hat{\mathbf{f}}_t &= (\hat{f}_{1t}, \ldots, \hat{f}_{Kt})'
\end{aligned} \tag{12}$$

The factor loadings for each asset, $\boldsymbol{\beta}_i$, and the residual variances, $var(\varepsilon_{it}) = \sigma_i^2$ can be estimated via OLS from the time series regression

$$R_{it} = \alpha_i + \boldsymbol{\beta}_i'\hat{\mathbf{f}}_t + \varepsilon_{it},\ t = 1, \ldots, T \tag{13}$$

giving $\widehat{\boldsymbol{\beta}}_i$ and $\widehat{\sigma}_i^2$ for $i = 1, \ldots, N$. The factor model covariance matrix of returns is then

$$\widehat{\mathbf{\Omega}}_{FM} = \widehat{\mathbf{B}}\widehat{\mathbf{\Omega}}_f\widehat{\mathbf{B}}' + \widehat{\mathbf{D}} \tag{14}$$

where

$$\widehat{\mathbf{B}} = \begin{pmatrix} \widehat{\boldsymbol{\beta}}_1' \\ \vdots \\ \widehat{\boldsymbol{\beta}}_N' \end{pmatrix},\ \widehat{\mathbf{D}} = \begin{pmatrix} \widehat{\sigma}_1^2 & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & \cdots & \widehat{\sigma}_N^2 \end{pmatrix},$$

and

$$\widehat{\mathbf{\Omega}}_f = \frac{1}{T-1}\sum_{t=1}^{T}(\widehat{\mathbf{f}}_t - \overline{\mathbf{f}})(\widehat{\mathbf{f}}_t - \overline{\mathbf{f}})', \overline{\mathbf{f}} = \frac{1}{T}\sum_{t=1}^{T}\widehat{\mathbf{f}}_t$$

Factor Mimicking Portfolios

Since each sample principal component (factor), $\hat{f}_{tk}$, is a linear combinations of the returns, it is possible to construct portfolios that are perfectly correlated with the principal components by re-normalizing the weights in the $\hat{\mathbf{p}}_k^*$ vectors so that they sum to unity. Hence, the weights in the factor mimicking portfolios have the form

$$\hat{\mathbf{w}}_k = \left(\frac{1}{\mathbf{1}'_N \hat{\mathbf{p}}_k^*}\right) \cdot \hat{\mathbf{p}}_k^*, \;\; k = 1, \ldots, K \tag{15}$$

where $\mathbf{1}$ is a $(N \times 1)$ vector of ones, and the factor mimicking portfolio returns are

$$R_{k,t} = \hat{\mathbf{w}}_k' \mathbf{R}_t$$

Asymptotic Principal Components

- *Asymptotic principal component analysis* (APCA), proposed and developed in Conner and Korajczyk (1986), is similar to traditional PCA except that it relies on asymptotic results as the number of cross-sections N (assets) grows large.

- APCA is based on eigenvector analysis of the $T \times T$ matrix $\widehat{\mathbf{\Omega}}_T$. Conner and Korajczyk prove that as N grows large, eigenvector analysis of $\widehat{\mathbf{\Omega}}_T$ is asymptotically equivalent to traditional factor analysis. That is, the APCA estimates of the factors $\mathbf{f}_t$ are the first K eigenvectors of $\widehat{\mathbf{\Omega}}_T$. Specifically, let $\widehat{\mathbf{F}}$ denote the orthornormal $K \times T$ matrix consisting of the first K eigenvectors of $\widehat{\mathbf{\Omega}}_T$. Then $\widehat{\mathbf{f}}_t$ is the t^{th} column of $\widehat{\mathbf{F}}$.

The main advantages of the APCA approach are:

- It works in situations where the number of assets, N, is much greater than the number of time periods, T. Eigenvectors of the smaller $T \times T$ matrix $\widehat{\mathbf{\Omega}}_T$ only need to be computed, whereas with traditional principal component analysis eigenvalues of the larger $N \times N$ matrix $\widehat{\mathbf{\Omega}}_N$ need to be computed.

- The method allows for an approximate factor structure of returns. In an approximate factor structure, the asset specific error terms ε_{it} are allowed to be contemporaneously correlated, but this correlation is not allowed to be too large across the cross section of returns. Allowing an approximate factor structure guards against picking up local factors, e.g. industry factors, as global common factors.

Determining the Number of Factors

- In practice, the number of factors is unknown and must be determined from the data.

- If traditional factor analysis is used, then there is a likelihood ratio test for the number of factors. However, this test will not work if $N > T$.

- Connor and Korajczyk (1993) described a procedure for determining the number of factors in an approximate factor model that is valid for $N > T$. Bai and Ng (2002) proposed an information criteria that is easier and more reliable to use.

Bai and Ng Method

Bai and Ng (2002) propose some panel C_p (Mallows-type) information criteria for choosing the number of factors. Their criteria are based on the observation that eigenvector analysis on $\hat{\mathbf{\Omega}}_T$ or $\hat{\mathbf{\Omega}}_N$ solves the least squares problem

$$\min_{\boldsymbol{\beta}_i, \mathbf{f}_t} (NT)^{-1} \sum_{i=1}^{N} \sum_{t=1}^{T} (R_{it} - \alpha_i - \boldsymbol{\beta}_i' \mathbf{f}_t)^2$$

Bai and Ng's model selection or information criteria are of the form

$$\begin{aligned} IC(K) &= \hat{\sigma}^2(K) + K \cdot g(N,T) \\ \hat{\sigma}^2(K) &= \frac{1}{N} \sum_{i=1}^{N} \hat{\sigma}_i^2 \end{aligned}$$

where $\hat{\sigma}^2(K)$ is the cross-sectional average of the estimated residual variances for each asset based on a model with K factors and $g(N,T)$ is a penalty function depending only on N and T.

The preferred model is the one which minimizes the information criteria $IC(K)$ over all values of $K < K_{\text{max}}$. Bai and Ng consider several penalty functions and the preferred criteria are

$$\begin{aligned} PC_{p1}(K) &= \hat{\sigma}^2(K) + K \cdot \hat{\sigma}^2(K_{\text{max}}) \left(\frac{N+T}{NT} \right) \cdot \ln \left(\frac{NT}{N+T} \right), \\ PC_{p2}(K) &= \hat{\sigma}^2(K) + K \cdot \hat{\sigma}^2(K_{\text{max}}) \left(\frac{N+T}{NT} \right) \cdot \ln \left(C_{NT}^2 \right), \\ C_{NT} &= \min(\sqrt{N}, \sqrt{T}) \end{aligned}$$

Algorithm

First, select a number $K_{\max}$ indicating the maximum number of factors to be considered. Then for each value of $K < K_{\max}$, do the following:

1. Extract realized factors $\widehat{\mathbf{f}}_t$ using the method of APCA.

2. For each asset i, estimate the factor model

$$R_{it} = \alpha_i + \boldsymbol{\beta}_i' \widehat{\mathbf{f}}_t^K + \varepsilon_{it},$$

where the superscript K indicates that the regression has K factors, using time series regression and compute the residual variances

$$\widehat{\sigma}_i^2(K) = \frac{1}{T-K-1} \sum_{t=1}^{T} \widehat{\varepsilon}_{it}^2.$$

3. Compute the cross-sectional average of the estimated residual variances for each asset based on a model with K factors

$$\widehat{\sigma}^2(K) = \frac{1}{N} \sum_{i=1}^{N} \widehat{\sigma}_i^2(K)$$

4. Compute the cross-sectional average of the estimated residual variances for each asset based on a model with $K_{\max}$ factors, $\widehat{\sigma}^2(K_{\max})$.

5. Compute the information criteria $PC_{p1}(K)$ and $PC_{p2}(K)$.

6. Select the value of K that minimized either $PC_{p1}(K)$ or $PC_{p2}(K)$.

Bai and Ng perform an extensive simulation study and find that the selection criteria PC_{p1} and PC_{p2} yield high precision when $\min(N,T) > 40$.