CFRM 546
Estimating Statistical Factor Models for Asset Returns

Eric Zivot

June 5, 2015

Statistical Factor Models for Returns

- In statistical factor models, the factor realizations $\mathbf{f}_t$ in (**??**) 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.

Statistical Factor Models: Advantages and Disadvantages

- Extracted factors often explain a large portion of the variability of returns - typically more than fundamental factors

- Interpretation of statistical factors can be difficult
  - For equity returns, typically only a few factors can be easily interpreted (e.g. market and industry factors)
  - For fixed income returns, extracted factors can be interpreted as yield curve descriptors (e.g. level, slope, curvature)
  - For hedge funds, extracted factors can be interpreted as strategy descriptors

Statistical Factor Models: Estimation

$$\begin{aligned} R_{it} &= \alpha_i + \boldsymbol{\beta}_i' \mathbf{f}_t + \varepsilon_{it} \\ \underset{N \times N}{\boldsymbol{\Omega}} &= \text{var}(\mathbf{R}_t) \end{aligned}$$

Econometric problems:

- Both factor loadings $\boldsymbol{\beta}_i$ and factor returns $\mathbf{f}_t$ are not observed - only asset returns R_{it} are observed
- First, extract $\mathbf{f}_t$ from covariance/correlation matrix of returns
- Second estimate $\boldsymbol{\beta}_i$ by regressing asset returns on extracted factors

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

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

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 $\mathbf{\Omega} = E[(\mathbf{R}_t - E[\mathbf{R}_t])(\mathbf{R}_t - E[\mathbf{R}_t])']$. 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:

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

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

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

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

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

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

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

Spectral (Eigenvalue) Decomposition of Ω

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

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

- $\boldsymbol{\Lambda}$ 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

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

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

$$R_{it} = \alpha_i + \boldsymbol{\beta}_i' \hat{\mathbf{f}}_t + \varepsilon_{it},\ t = 1, \ldots, T \tag{2}$$

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

$$\widehat{\boldsymbol{\Omega}}_{FM} = \widehat{\mathbf{B}} \widehat{\boldsymbol{\Omega}}_f \widehat{\mathbf{B}}' + \widehat{\mathbf{D}} \tag{3}$$

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

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

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

$$\hat{\mathbf{w}}_k = \left(\frac{1}{\mathbf{1}'_N \hat{\mathbf{p}}_k^*} \right) \cdot \hat{\mathbf{p}}_k^*, \ k = 1, \ldots, K \tag{4}$$

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 $\widehat{\mathbf{\Omega}}_T$ or $\widehat{\mathbf{\Omega}}_N$ 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

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

where $\widehat{\sigma}^2(K)$ 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 $K < K_{\max}$. Bai and Ng consider several penalty functions and the preferred criteria are

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

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 $\min(N, T) > 40$.