FACTOR MODELS

basis for APT

1 ORTHOGONAL FACTOR MODEL

- Returns $R_{1,t} \cdots R_{pt}$

  Factors $F_{1,t} \cdots F_{Kt}$

p large

K factors

$k \ll p$

$i=1,\dots,p$ $\quad R_{it} = \mu_i + \sum_{k=1}^{K} b_{ik} f_{kt} + \varepsilon_{it}$

非系统性风险

purely firm specific noise

Matrix notation : $\underset{p\times 1}{R_t} = \mu + \underset{p\times K}{B}\, \underset{K\times 1}{F_t} + \underset{p\times 1}{\varepsilon_t}$

factors are uncorrelated

Orthogonal factor model : $E(FF') = I$ Cov matirx of factor

$E(\varepsilon\varepsilon') = D$ return 间无相关性

$E(\varepsilon F') = 0$

not a contriant, 非0的

Factor return 都会跑到μ里

Also : $E(F) = 0$

$E(\varepsilon) = 0$ mean adjusted

- $E(\varepsilon\varepsilon') = D$. crucial for proving APT

  allows to form zero variance (asymptotic) portfolios

- Covariance structure $\quad R = BF + \varepsilon$

$$E(RR') = V = B\,\underbrace{E(FF')}_{\text{identity}}\,B' + E(\varepsilon\varepsilon')$$

$$\boxed{V = BB' + D}$$

$$p(p+1)/2 \qquad pk \quad + \quad p \quad = p(k+1)$$

- Factors are not unique:

$$R_t = \mu + BF_t + \varepsilon_t$$

Take any matrix P : $PP' = P'P = I$

Then : $R_t = \mu + \underbrace{BP}\,\underbrace{P'F_t} + \varepsilon_t$

$$= \mu + B^* F_t^* + \varepsilon_t$$

$$E(F_t^*) = P'E(F_t) = 0$$

$$E(F^*F^{*\prime}) = E(P'FF'P) = P'E(FF')P$$

$$= P'P = I$$

$$V = B^*B^{*\prime} + D = BB' + D$$

Factors are determined up to a rotation

Loadings B and $B^* = BP$ give the same covariance

- How to estimate a factor model?

Two main approaches
- principal component analysis
- Likelihood based analysis

2 PRINCIPAL COMPONENTS

- Random vector $R = (R_1 \cdots R_p) \sim (\mu, \boxed{V})$

Consider linear combinations of R

port 数量与 principle components 数量一样

$$\begin{cases} Y_1 = \ell_1' R \\ Y_2 = \ell_2' R \\ \vdots \\ Y_p = \ell_p' R \end{cases}$$

like port weights

$$\mathrm{var}(Y_i) = \ell_i' V \ell_i$$

$$\mathrm{cov}(Y_i, Y_j) = \ell_i' V \ell_j$$

- Principal components are ① uncorrelated combinations with ② highest possible variance

- Component 1 $\quad \ell_1 : \ \mathrm{Max}\ \ell_1' V \ell_1 \quad$ s.t. $\underbrace{\ell_1'\ell_1 = 1}_{\text{why?}}$

否则 V goes to infinity

- Component 2 $\quad \ell_2 : \ \mathrm{Max}\ \ell_2' V \ell_2 \quad$ s.t. $\ell_2'\ell_2 = 1$

$$\ell_2' V \ell_1 = 0$$

← uncorrelated with first components

Var: $l_1 > l_2 > l_3 \cdots$ 因为越往后 constraint 越多

Sum(λs) = Sum(Variance of stocks)

- **Result:** If V has eigen values $\lambda_1 \cdots \lambda_p$ * → variances of components

  and eigen vectors $\underline{e}_1 \cdots \underline{e}_p$ → ℓs vector

  $$Y_i = e_i' R$$

  $$V = E \Lambda E'$$

  $$\text{var}(Y_i) = e_i' V e_i = \lambda_i$$

  [|] trace of V is sum(Var stocks)

  $$\sigma^2_{R_1} + \cdots + \sigma^2_{R_p} = \lambda_1 + \cdots + \lambda_p$$

  trace of $E'E$ is I. (property of eigenvectors.)

  Sum of square of the components E_1 is 1.

  ⋯ E_2 is 1.

- **Special Case:** Equal correlation matrix

$$V = \underbrace{\begin{bmatrix} \sigma & & 0 \\ & \ddots & \\ 0 & & \sigma \end{bmatrix}}_{D^{\frac{1}{2}}} \underbrace{\begin{bmatrix} 1 & & \rho \\ & \ddots & \\ \rho & & 1 \end{bmatrix}}_{Cor} \underbrace{\begin{bmatrix} \sigma & & 0 \\ & \ddots & \\ 0 & & \sigma \end{bmatrix}}_{D^{\frac{1}{2}}}$$

dimension of matrix

Then: $\lambda_1 = 1 + (p-1)\rho$ ← 它最大 (like equal-weighted index)

$\lambda_2 = \cdots = \lambda_p = 1 - \rho$ ← 它们都一样

$$e_1 = \left[\frac{1}{\sqrt{p}}, \cdots, \frac{1}{\sqrt{p}}\right]$$

The first component explains a large fraction of the total variance. It is akin to an equal weighted index

* Eigen values solve $|V - \lambda I| = 0$

Eigen vector: $V\underline{e}_i = \lambda_i \underline{e}_i$

- **Estimation** Do the above computations [1] on $\hat{V}$

  Amounts to decomposing $\hat{V}$ along its eigen values and vectors

$$V = \sum_{i=1}^{p} \lambda_i \underbrace{e_i e_i'}_{p \times p} = [\underline{e}_1 \cdots \underline{e}_p] \begin{bmatrix} \lambda_1 & & 0 \\ & \ddots & \\ 0 & & \lambda_p \end{bmatrix} \begin{bmatrix} e_1' \\ \vdots \\ e_p' \end{bmatrix}$$

$$= P \Lambda P'$$

主成分 主成分的方差

.... Back to our factor model

3 Estimation of Factor Model by PC

- $V = \lambda_1 e_1 e_1' + \cdots + \lambda_p e_p e_p'$

$$= \underbrace{\left[\sqrt{\lambda_1}\, e_1 , \cdots\cdots \sqrt{\lambda_p}\, e_p\right]}_{\text{may be ignore 黄的 (to find a cutoff)}} \underbrace{\begin{bmatrix} \sqrt{\lambda_1}\, e_1' \\ \vdots \\ \sqrt{\lambda_p}\, e_p' \end{bmatrix}}$$

$$V = \underset{p \times p\ (k)}{B} \times \underset{p \times p\ (k)}{B'} + 0$$

$\underset{p \times k}{V - BB'}$

- Exact, no noise!

Useless: $p = 100$ stocks $\Rightarrow$ 100 factors!

- Idea: pick the first m principal components that explain most of the total variance

Use λ to decide what to keep .. throw

- $\underset{p \times p}{\hat{V}} \Rightarrow \underset{p \times m}{\hat{B}}$ $\quad m \ll p$

$\hat{D}$? $\left\{ \text{Diagonal elements of } \hat{V} - \underset{p \times m}{\hat{B}}\, \underset{m \times p}{\hat{B}'} \right\} = \begin{pmatrix} \hat{d}_1 & & 0 \\ & \ddots & \\ 0 & & \hat{d}_p \end{pmatrix}$

$\hat{V} \approx BB' + D$ approximation: only keeping diagonal elements.

$p \times m$ $m \times p$

- What do we have? $\underset{p \times m}{\hat{B}}$!

  Factor model: $\underset{p \times 1}{R_t} = \mu + \underset{p \times m}{\hat{B}} \underset{m \times 1}{F_t} + \varepsilon_t$ [2]

  y X

  ↳ number of factors

  We don't have F_t ! "score"

- F_t : For each time t, regress R_t on B

  a cross sectional regression with p observations $(r_{1t} \cdots r_{pt})$ and noise variance $(d_1, \ldots d_p)$

  Heteroskedastic : $\varepsilon_t \sim N(0, \hat{D})$

  requires GLS

  Scores $\underset{m \times 1}{\hat{F}_t} = (\hat{B}' \hat{D}^{-1} \hat{B})^{-1} (\hat{B}' \hat{D}^{-1} R_t)$

Step 2

- If $\hat{B}$ is noisy .... because $\hat{V}$ is noisy

  .... maybe T is not large relative to p

  the GLS regression of [2] suffers from errors-in-the-variables.

- Stocks : $p = 1000$, $T = ?$] $\underset{1000 \times 1000}{\hat{V}} \rightarrow \hat{B} \rightarrow$ EIV in 2nd step

④ Estimation of FACTOR MODEL BY M. L.

T observations of p returns

- $R_t = \mu + B F_t + \varepsilon_t$ — F_t: m number of factors ≥ 3 factor, 4 ..

$F_t \sim N(0, I_{m\times m})$
$\varepsilon_t \sim N(0, D)$ $\Longrightarrow$ write the likelihood. $\ell(\mu, V \mid R)$

Maximize the likelihood s.t $V = \underbrace{B}_{p\times m}B' + \underbrace{D}_{p}$ *

$\Longrightarrow \hat{B}, \hat{D}$

① • $\hat{B}_m$'s will differ from $\hat{B}_{m+1}$'s first m components
contrast with principal components

---

② • Then same step as above to estimate
$\hat{F}_t$ by cross-sectional GLS $R_t = \mu + \hat{B} F_t + \varepsilon_t$

用 likelihood ratio test 去看几个 factor 更好

* Need further restriction for B, typically $B'D^{-1}B$ diagonal.

⑤ Asymptotic Principal Components

- Connor & Korajczyk (86, 88) test APT and flip this around:

Fixed income 用不了, 因为没那么多 instruments.

$p = 1500$ stocks
$T = 60$ $\Bigg\}$ $\hat{V} \rightarrow p \times p$ makes no sense, it is not full rank.

- Consider the returns moment matrix:

$$\underset{T\times T}{\Omega} = \frac{R'R}{p} \qquad \underset{p\times T}{R} = ||||| \quad \underset{T\times p}{R'}\ \underset{p\times T}{R}$$

They show that, as $p \to \infty$, the eigen vectors of Ω estimate the factors with arbitrary precision

need #stock to be 大, noise 没了 $\Rightarrow \hat{F}_t$

- Second pass

   - $R_t = \mu + B\hat{F}_t + \varepsilon_t$ → precise if p large

   Run p, time series OLS regressions to get B

- $p\nearrow \Rightarrow \hat{F}_t$ precise $\Rightarrow$ No EIV in 2nd Pass

6 BAYESIAN APPROACH (ignoring identification)

$$R_t = \alpha + BF + \varepsilon \qquad \varepsilon \sim N(0, D)$$

- Prior $\quad p(\alpha, B, D) \propto |D|^{-1/2} = \dfrac{1}{\prod \sigma_{\varepsilon_i}}$
- Posterior $\ p(\alpha, B, D \mid R) \propto |D|^{-1/2}\ p(R \mid \alpha, B, D)$
- Data augmentation - condition on F

1 $\left.\begin{array}{l} p(\alpha, B \mid F, R, D) \\ p(D \mid F, R) \end{array}\right\}$ simple regression

2 $p(F \mid \alpha, B, D, R)$?

F, R are jointly normal $\quad \begin{pmatrix} F_t \\ R_t \end{pmatrix} \sim N\left(\begin{bmatrix} 0 \\ \alpha \end{bmatrix}, \begin{bmatrix} I & B' \\ B & BB'+D \end{bmatrix} \right)$

Conditional distribution

$$E[F_t \mid \alpha, B, D, R_t] = B'(BB'+D)^{-1}(R_t - \alpha)$$

$$V[F_t \mid \alpha, B, D, R_t] = I - B'(BB'+D)^{-1}B$$

(1) and (2) form a Gibbs Sampling Algorithm