Lecture 15: Factor Models

MIT 18.S096

Dr. Kempthorne

Fall 2013

Outline

1. Factor Models
   - Linear Factor Model
   - Macroeconomic Factor Models
   - Fundamental Factor Models
   - Statistical Factor Models: Factor Analysis
   - Principal Components Analysis
   - Statistical Factor Models: Principal Factor Method

Linear Factor Model

Data:

- m assets/instruments/indexes: $i = 1, 2, \ldots, m$
- n time periods: $t = 1, 2, \ldots, n$
- m-variate random vector for each time period:

$$\mathbf{x}_t = (x_{1,t}, x_{2,t}, \ldots, x_{m,t})'$$

E.g., returns on m stocks/futures/currencies;

interest-rate yields on m US Treasury instruments.

Factor Model

$$\begin{aligned} x_{i,t} &= \alpha_i + \beta_{1,i} f_{1,t} + \beta_{2,i} f_{2,t} + \cdots + \beta_{k,i} f_{k,t} + \epsilon_{i,t} \\ &= \alpha_i + \boldsymbol{\beta}_i' \mathbf{f}_t + \epsilon_{i,t} \quad \text{where} \end{aligned}$$

- α_i: intercept of asset i
- $\mathbf{f}_t = (f_{1,t}, f_{2,t}, \ldots, f_{K,t})'$: **common factor** variables at period t (constant over i)
- $\boldsymbol{\beta}_i = (\beta_{1,i}, \ldots, \beta_{K,i})'$: **factor loadings of asset** i (constant over t)
- $\epsilon_{i,t}$: the **specific factor of asset** i at period t.

Linear Factor Model

Linear Factor Model: Cross-Sectional Regressions

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

for each $t \in \{1, 2 \ldots, T\}$, where

$$\boldsymbol{\alpha} = \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_m \end{bmatrix} \ (m \times 1); \quad B = \begin{bmatrix} \boldsymbol{\beta}_1' \\ \boldsymbol{\beta}_2' \\ \vdots \\ \boldsymbol{\beta}_m' \end{bmatrix} = [[\beta_{i,k}]] \ (m \times K); \quad \boldsymbol{\epsilon}_t = \begin{bmatrix} \epsilon_{1,t} \\ \epsilon_{2,t} \\ \vdots \\ \epsilon_{m,t} \end{bmatrix} \ (m \times 1)$$

- $\boldsymbol{\alpha}$ and B are the same for all t.
- $\{\mathbf{f}_t\}$ is ($K-$variate) covariance stationary $I(0)$ with

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

- $\{\boldsymbol{\epsilon}_t\}$ is m-variate white noise with:

$$\begin{aligned} E[\boldsymbol{\epsilon}_t] &= \mathbf{0}_m \\ Cov[\boldsymbol{\epsilon}_t] &= E[\boldsymbol{\epsilon}_t\boldsymbol{\epsilon}_t'] = \boldsymbol{\Psi} \\ Cov[\boldsymbol{\epsilon}_t, \boldsymbol{\epsilon}_{t'}] &= E[\boldsymbol{\epsilon}_t\boldsymbol{\epsilon}_{t'}'] = \mathbf{0} \ \ \forall t \neq t' \end{aligned}$$

  $\boldsymbol{\Psi}$ is the $(m \times m)$ diagonal matrix with entries $(\sigma_1^2, \sigma_2^2, \ldots, \sigma_m^2)$ where $\sigma_i^2 = var(\epsilon_{i,t})$, the variance of the ith asset specific factor.
- The two processes $\{\mathbf{f}_t\}$ and $\{\boldsymbol{\epsilon}_t\}$ have null cross-covariances:

Linear Factor Model

Summary of Parameters

- $\boldsymbol{\alpha}$: $(m \times 1)$ intercepts for m assets
- B: $(m \times K)$ loadings on K common factors for m assets
- $\boldsymbol{\mu}_f$: $(K \times 1)$ mean vector of K common factors
- $\boldsymbol{\Omega}_f$: $(K \times K)$ covariance matrix of K common factors
- $\boldsymbol{\Psi} = diag(\sigma_1^2, \ldots, \sigma_m^2)$: m asset-specific variances

Features of Linear Factor Model

- The $m-$variate stochastic process $\{\mathbf{x}_t\}$ is a covariance-stationary multivariate time series with
  - Conditional moments:
    $$E[\mathbf{x}_t \mid \mathbf{f}_t] = \boldsymbol{\alpha} + B\mathbf{f}_t$$
    $$Cov[\mathbf{x}_t \mid \mathbf{f}_t] = \boldsymbol{\Psi}$$
  - Unconditional moments:
    $$E[\mathbf{x}_t] = \boldsymbol{\mu}_x = \boldsymbol{\alpha} + B\boldsymbol{\mu}_f$$
    $$Cov[\mathbf{x}_t] = \boldsymbol{\Sigma}_x = B\boldsymbol{\Omega}_f B' + \boldsymbol{\Psi}$$

Linear Factor Model

Linear Factor Model: Time Series Regressions

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

for each asset $i \in \{1, 2 \ldots, m\}$, where

$$\mathbf{x}_i = \begin{bmatrix} x_{i,1} \\ \vdots \\ x_{i,t} \\ \vdots \\ x_{1,T} \end{bmatrix} \quad \boldsymbol{\epsilon}_i = \begin{bmatrix} \epsilon_{i,t} \\ \vdots \\ \epsilon_{i,t} \\ \vdots \\ \epsilon_{i,T} \end{bmatrix} \quad \mathbf{F} = \begin{bmatrix} \mathbf{f}'_1 \\ \vdots \\ \mathbf{f}'_t \\ \vdots \\ \mathbf{f}'_T \end{bmatrix} = \begin{bmatrix} f_{1,1} & f_{2,1} & \cdots & f_{K,1} \\ \vdots & \vdots & \vdots & \vdots \\ f_{1,t} & f_{2,t} & \cdots & f_{K,t} \\ \vdots & \vdots & \vdots & \vdots \\ f_{1,T} & f_{2,T} & \cdots & f_{K,T} \end{bmatrix}$$

- α_i and $\boldsymbol{\beta}_i = (\beta_{1,i}, \ldots, \beta_{K,i})$ are regression parameters.
- $\boldsymbol{\epsilon}_i$ is the T-vector of regression errors with $Cov(\boldsymbol{\epsilon}_i) = \sigma_i^2 \mathbf{I}_T$

Linear Factor Model: Multivariate Regression

$$\mathbf{X} = [\mathbf{x}_1 | \cdots | \mathbf{x}_m],\ \mathbf{E} = [\boldsymbol{\epsilon}_1 | \cdots | \boldsymbol{\epsilon}_m],\ \mathbf{B} = [\boldsymbol{\beta}_1 | \cdots | \boldsymbol{\beta}_m],$$

$$\mathbf{X} = \mathbf{1}_T \boldsymbol{\alpha}' + \mathbf{F}\mathbf{B} + \mathbf{E}$$

(note that $\mathbf{B}$ equals the transpose of cross-sectional B)

Outline

1. Factor Models
   - Linear Factor Model
   - Macroeconomic Factor Models
   - Fundamental Factor Models
   - Statistical Factor Models: Factor Analysis
   - Principal Components Analysis
   - Statistical Factor Models: Principal Factor Method

Macroeconomic Factor Models

Single Factor Model of Sharpe (1970)

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

where

- R_{Mt} is the return of the market index in excess of the risk-free rate; the **market risk factor**.
- $x_{i,t}$ is the return of asset i in excess of the risk-free rate.
- $K = 1$ and the single factor is $f_{1,t} = R_{Mt}$.
- Unconditional cross-sectional covariance matrix of the assets:

  $$Cov(\mathbf{x}_t) = \mathbf{\Sigma}_x = \sigma_M^2 \boldsymbol{\beta}\boldsymbol{\beta}' + \mathbf{\Psi} \text{ where}$$

  - $\sigma_M^2 = Var(R_{Mt})$
  - $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_m)'$
  - $\mathbf{\Psi} = diag(\sigma_1^2, \ldots, \sigma_m^2)$

Estimation of Sharpe's Single Index Model

- Single Index Model satisfies the Generalized Gauss-Markov assumptions so the least-squares estimates $(\hat{\alpha}_i, \hat{\beta}_i)$ from the time-series regression for each asset i are best linear unbiased estimates (BLUE) and the MLEs under Gaussian assumptions.
$$\mathbf{x}_i = \mathbf{1}_T \hat{\alpha}_i + \mathbf{R}_M \hat{\beta}_i + \hat{\boldsymbol{\epsilon}}_i$$
- Unbiased estimators of remaining parameters:
    - $\hat{\sigma}_i^2 = (\hat{\boldsymbol{\epsilon}}_i' \hat{\boldsymbol{\epsilon}}_i)/(T-2)$
    - $\hat{\sigma}_M^2 = [\sum_{t=1}^T (\mathbf{R}_{Mt} - \bar{\mathbf{R}}_M)^2]/(T-1)$ with $\bar{\mathbf{R}}_M = (\sum_{t=1}^T \mathbf{R}_{Mt})/T$
    - $\hat{\boldsymbol{\Psi}} = diag(\hat{\sigma}_1^2, \dots, \hat{\sigma}_m^2)$
- Estimator of unconditional covariance matrix:
$$\widehat{Cov}(\mathbf{x}_t) = \hat{\boldsymbol{\Sigma}}_x = \hat{\sigma}_M^2 \hat{\boldsymbol{\beta}} \hat{\boldsymbol{\beta}}' + \hat{\boldsymbol{\Psi}}$$

Macroeconomic Multifactor Model

The common factor variables $\{\mathbf{f}_t\}$ are realized values of macro econonomic variables, such as

- Market risk
- Price indices (CPI, PPI, commodities) / Inflation
- Industrial production (GDP)
- Money growth
- Interest rates
- Housing starts
- Unemployment

  See Chen, Ross, Roll (1986). "Economic Forces and the Stock Market"

Linear Factor Model as Time Series Regressions

$$\mathbf{x}_i = \mathbf{1}_T \alpha_i + \mathbf{F}\boldsymbol{\beta}_i + \boldsymbol{\epsilon}_i, \quad \text{where}$$

- $\mathbf{F} = [\mathbf{f}_1, \mathbf{f}_2, \ldots \mathbf{f}_T]'$ is the $(T \times K)$ matrix of realized values of $(K > 0)$ macroeconomic factors.
- Unconditional cross-sectional covariance matrix of the assets:

  $$Cov(\mathbf{x}_t) = B\boldsymbol{\Omega}_f B' + \boldsymbol{\Psi}$$

  where $B = (\boldsymbol{\beta}_1, \ldots, \boldsymbol{\beta}_m)'$ is $(m \times K)$

Estimation of Multifactor Model

- Multifactor model satisfies the Generalized Gauss-Markov assumptions so the least-squares estimates $\hat{\alpha}_i$ and $\hat{\boldsymbol{\beta}}_i$ $_{(K \times 1)}$ from the time-series regression for each asset i are best linear unbiased estimates (BLUE) and the MLEs under Gaussian assumptions.

$$\mathbf{x}_i = \mathbf{1}_T \hat{\alpha}_i + \mathbf{F}\hat{\boldsymbol{\beta}}_i + \hat{\boldsymbol{\epsilon}}_i$$

- Unbiased estimators of remaining parameters:
  - $\hat{\sigma}_i^2 = (\hat{\boldsymbol{\epsilon}}_i' \hat{\boldsymbol{\epsilon}}_i)/[T - (k+1)]$
  - $\hat{\boldsymbol{\Psi}} = diag(\hat{\sigma}_1^2, \ldots, \hat{\sigma}_m^2)$
  - $\hat{\boldsymbol{\Omega}}_f = [\sum_{t=1}^{T}(\mathbf{f}_t - \bar{\mathbf{f}})(\mathbf{f}_t - \bar{\mathbf{f}})']/(T-1)$
    with $\bar{\mathbf{f}} = (\sum_{t=1}^{T} \mathbf{f}_t)/T$

- Estimator of unconditional covariance matrix:

$$\widehat{Cov(\mathbf{x}_t)} = \hat{\boldsymbol{\Sigma}}_x = \hat{\sigma}_M^2 \hat{B} \hat{\boldsymbol{\Omega}}_f \hat{B}' + \hat{\boldsymbol{\Psi}}$$

Outline

1 Factor Models
- Linear Factor Model
- Macroeconomic Factor Models
- Fundamental Factor Models
- Statistical Factor Models: Factor Analysis
- Principal Components Analysis
- Statistical Factor Models: Principal Factor Method

Fundamental Factor Models

The common-factor variables $\{\mathbf{f}_t\}$ are determined using fundamental, asset-specific attributes such as

- Sector/industry membership.
- Firm size (market capitalization)
- Dividend yield
- Style (growth/value as measured by price-to-book, earnings-to-price, ...)
- Etc.

BARRA Approach (Barr Rosenberg)

- Treat observable asset-specific attributes as **factor betas**
- Factor realizations $\{\mathbf{f}_t\}$ are unobservable, but are estimated.

Fama-French Approach (Eugene Fama and Kenneth French)

- For every time period t, apply cross-sectional sorts to define factor realizations
  - For a given asset attribute, sort the assets at period t by that attribute and define quintile portfolios based on splitting the assets into 5 equal-weighted portfolios.
  - Form the hedge portfolio which is long the top quintile assets and short the bottom quintile assets.
- Define the common factor realizations for period t as the period-t returns for the K hedge portfolios corresponding to the K fundamental asset attributes.
- Estimate the factor loadings on assets using time series regressions, separately for each asset i.

Barra Industry Factor Model

- Suppose the m assets $(i = 1, 2, \ldots, m)$ separate into K industry groups $(k = 1, \ldots, K)$
- For each asset i, define the factor loadings $(k = 1, \ldots K)$
$$\beta_{i,k} = \begin{cases} 1 & \text{if asset } i \text{ is in industry group } k \\ 0 & \text{otherwise} \end{cases}$$
These loadings are time invariant.
- For time period t, denote the realization of the K factors as
$$\mathbf{f}_t = (f_{1t}, \ldots, f_{Kt})'$$
These $K-$ vector realizations are unobserved.

The Industry Factor Model is
$$X_{i,t} = \beta_{i,1} f_{1t} + \cdots + \beta_{i,K} f_{Kt} + \epsilon_{it}, \quad \forall i, t$$
where
$$\begin{array}{lcll} var(\epsilon_{it}) & = & \sigma_i^2, & \forall i \\ cov(\epsilon_{it}, f_{kt}) & = & 0, & \forall i, k, t \\ cov(f_{k't}, f_{kt}) & = & [\mathbf{\Omega}_f]_{k',k}, & \forall k', k, t \end{array}$$

Barra Industry Factor Model

Estimation of the Factor Realizations

For each time period t consider the cross-sectional regression for the factor model:

$$\mathbf{x}_t = B\mathbf{f}_t + \boldsymbol{\epsilon}_t \qquad (\boldsymbol{\alpha} = 0 \text{ so it does not appear})$$

with

$$\mathbf{x}_t = \begin{bmatrix} x_{1,t} \\ x_{2,t} \\ \vdots \\ x_{m,t} \end{bmatrix} \ (m \times 1); \quad B = \begin{bmatrix} \boldsymbol{\beta}_1' \\ \boldsymbol{\beta}_2' \\ \vdots \\ \boldsymbol{\beta}_m' \end{bmatrix} = [[\beta_{i,k}]] \ (m \times K); \quad \boldsymbol{\epsilon}_t = \begin{bmatrix} \epsilon_{1,t} \\ \epsilon_{2,t} \\ \vdots \\ \epsilon_{m,t} \end{bmatrix} \ (m \times 1)$$

where $E[\epsilon_t] = \mathbf{0}_m$, $E[\epsilon_t \epsilon_t'] = \boldsymbol{\Psi}$, and $Cov(\mathbf{f}_t) = \boldsymbol{\Omega}_f$.

- Compute $\hat{\mathbf{f}}_t$ by least-squares regression of $\mathbf{x}_t$ on B with regression parameter $\mathbf{f}_t$.
- B is $(m \times K)$ matrix of indicator variables (same for all t)

  $$B'B = diag(m_1, \ldots m_K),$$

  where m_k is the count of assets i in industry k, and $\sum_{k=1}^{K} m_k = m$.
- $\hat{\mathbf{f}}_t = (B'B)^{-1}B'\mathbf{x}_t$ (vector of industry averages!)
- $\hat{\epsilon}_t = \mathbf{x}_t - B\hat{\mathbf{f}}_t$

Barra Industry Factor Model

Estimation of Factor Covariance Matrix

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

Estimation of Residual Covariance Matrix $\hat{\mathbf{\Psi}}$

$$\hat{\mathbf{\Psi}} = diag(\hat{\sigma}_1^2, \ldots, \hat{\sigma}_m^2)$$

where

$$\hat{\sigma}_i^2 = \frac{1}{T-1}\sum_{t=1}^{T}[\hat{\epsilon}_{i,t} - \bar{\hat{\epsilon}}_i]^2$$
$$\bar{\hat{\epsilon}}_i = \frac{1}{T}\sum_{t=1}^{T}\hat{\epsilon}_{i,t}$$

Estimation of Industry Factor Model Covariance Matrix

$$\hat{\mathbf{\Sigma}} = B'\hat{\mathbf{\Omega}}_f B + \hat{\mathbf{\Psi}}$$

Barra Industry Factor Model

Further Details

- Inefficiency of least squares estimates due to heteroscedasticity in $\boldsymbol{\Psi}$.
  Resolution: apply Generalized Least Squares (GLS) estimating $\boldsymbol{\Psi}$ in the cross-sectional regressions.
- The factor realizations can be rescaled to represent **factor mimicking portfolios**
- The Barra Industry Factor Model can be expressed as a **seemingly unrelated regression (SUR) model**

Outline

1. Factor Models
   - Linear Factor Model
   - Macroeconomic Factor Models
   - Fundamental Factor Models
   - Statistical Factor Models: Factor Analysis
   - Principal Components Analysis
   - Statistical Factor Models: Principal Factor Method

Statistical Factor Models

The common-factor variables $\{\mathbf{f}_t\}$ are hidden (*latent*) and their structure is deduced from analysis of the observed returns/data $\{\mathbf{x}_t\}$. The primary methods for extraction of factor structure are:

- Factor Analysis
- Principal Components Analysis

Both methods model the $\mathbf{\Sigma}$, the covariance matrix of $\{\mathbf{x}_t, t = 1, \ldots, T\}$ by focusing on the sample covariance matrix $\hat{\mathbf{\Sigma}}$, computed as follows:

$$\mathbf{X} = [\mathbf{x}_1 : \cdots \; \mathbf{x}_T]_{\ (m \times T)}$$

$$\mathbf{X}^* = \mathbf{X} \cdot (\mathbf{I}_T - \tfrac{1}{T}\mathbf{1}_T\mathbf{1}_T') \text{ ('de-meaned' by row)}$$

$$\hat{\mathbf{\Sigma}}_x = \tfrac{1}{T}\mathbf{X}^*(\mathbf{X}^*)'$$

Factor Analysis Model

Linear Factor Model as Cross-Sectional Regression

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

for each $t \in \{1, 2 \dots, T\}$ (m equations expressed in vector/matrix form) where

- $\boldsymbol{\alpha}$ and B are the same for all t.
- $\{\mathbf{f}_t\}$ is ($K-$variate) covariance stationary $I(0)$ with $E[\mathbf{f}_t] = \boldsymbol{\mu}_f$, $Cov[\mathbf{f}_t] = \boldsymbol{\Omega}_f$
- $\{\epsilon_t\}$ is m-variate white noise with $E[\epsilon_t] = \mathbf{0}_m$ and $Cov[\epsilon_t] = \boldsymbol{\Psi} = diag(\sigma_i^2)$

Invariance to Linear Tranforms of $\mathbf{f}_t$

- For any $(K \times K)$ invertible matrix H define

$$\mathbf{f}_t^* = Hf_t \text{ and } B^* = BH^{-1}$$

- Then the linear factor model holds replacing $\mathbf{f}_t$ and B

$$\begin{aligned} \mathbf{x}_t &= \boldsymbol{\alpha} + B^*\mathbf{f}_t^* + \boldsymbol{\epsilon}_t = \boldsymbol{\alpha} + BH^{-1}H\mathbf{f}_t + \boldsymbol{\epsilon}_t \\ &= \boldsymbol{\alpha} + B\mathbf{f}_t + \boldsymbol{\epsilon}_t \end{aligned}$$

and replacing $\boldsymbol{\mu}_f$ and $\boldsymbol{\Omega}_f$ with

$$\boldsymbol{\Omega}_f^* = Cov(\mathbf{f}_t^*) = Cov(H\mathbf{f}_t) = HCov(\mathbf{f}_t)H' = H\boldsymbol{\Omega}_f H'$$

$$\boldsymbol{\mu}_f^* = H\boldsymbol{\mu}_f$$

Factor Analysis Model

Standard Formulation of Factor Analysis Model

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

Under these assumptions the unconditional covariance matrix is
$$Cov(\mathbf{x}_t) = \mathbf{\Sigma}_x = BB' + \mathbf{\Psi}$$

Factor Analysis Model

Maximum Likelihood Estimation

For the model

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

- $\boldsymbol{\alpha}$ and B are vector/matrix constants.
- All random variables are Normal/Gaussian:
  - $\mathbf{x}_t$ i.i.d. $N_m(\boldsymbol{\alpha}, \boldsymbol{\Sigma}_x)$
  - $\mathbf{f}_t$ i.i.d. $N_K(\mathbf{0}_K \mathbf{I}_K)$
  - $\boldsymbol{\epsilon}_t$ i.i.d. $N_m(\mathbf{0}_m, \boldsymbol{\Psi})$
- $Cov(\mathbf{x}_t) = \boldsymbol{\Sigma}_x = BB' + \boldsymbol{\Psi}$

Model Likelihood

$$\begin{aligned} L(\boldsymbol{\alpha}, \boldsymbol{\Sigma}_x) &= p(\mathbf{x}_1, \ldots, \mathbf{x}_T \mid \boldsymbol{\alpha}, \boldsymbol{\Sigma}) \\ &= \prod_{t=1}^{T}[p(\mathbf{x}_t \mid \boldsymbol{\alpha}, \boldsymbol{\Sigma})] \\ &= \prod_{t=1}^{T}[(2\pi)^{-m/2}|\boldsymbol{\Sigma}|^{-\frac{1}{2}} exp\left(-\tfrac{1}{2}(\mathbf{x}_t - \boldsymbol{\alpha})'\boldsymbol{\Sigma}_x^{-1}(\mathbf{x}_t - \boldsymbol{\alpha})\right)] \\ &= (2\pi)^{-Tm/2}|\boldsymbol{\Sigma}|^{-\frac{T}{2}} exp\left[-\tfrac{1}{2}\sum_{t=1}^{T}(\mathbf{x}_t - \boldsymbol{\alpha})'\boldsymbol{\Sigma}_x^{-1}(\mathbf{x}_t - \boldsymbol{\alpha})\right] \end{aligned}$$

Factor Analysis Model

Log Likelihood of the Factor Model

$$\begin{aligned} l(\boldsymbol{\alpha}, \boldsymbol{\Sigma}_x) &= \log\ L(\boldsymbol{\alpha}, \boldsymbol{\Sigma}_x) \\ &= -\frac{TK}{2} log(2\pi) - \frac{K}{2} log(|\boldsymbol{\Sigma}|) \\ &\quad -\frac{1}{2}\sum_{t=1}^{T}(\mathbf{x}_t - \boldsymbol{\alpha})'\boldsymbol{\Sigma}_x^{-1}(\mathbf{x}_t - \boldsymbol{\alpha}) \end{aligned}$$

Maximum Likelihood Estimates (MLEs)

- The MLEs of $\boldsymbol{\alpha}$, B, $\boldsymbol{\Psi}$ are the values which
  Maximize $l(\boldsymbol{\alpha}, \boldsymbol{\Sigma}_x)$
  Subject to: $\boldsymbol{\Sigma}_x = BB' + \boldsymbol{\Psi}$
- The MLEs are computed numerically applying the Expectation-Maximization (EM) algorithm*

* Optional Reading: Dempster, Laird, and Rubin (1977), Rubin and Thayer (1983).

Factor Analysis Model

ML Specification of the Factor Model

- Apply EM algorithm to compute $\hat{\alpha}$ and $\hat{B}$ and $\hat{\mathbf{\Psi}}$.
- Estimate factor realizations $\{\mathbf{f}_t\}$
  - Apply the cross-sectional regression models for each time period t:
    $$\mathbf{x}_t - \hat{\alpha} = \hat{B}\mathbf{f}_t + \hat{\epsilon}_t$$
    Solving for $\hat{\mathbf{f}}$ as the regression parameter estimates of the regression of observed $\mathbf{x}_t$ on the estimated factor loadings matrix. Taking account of the heteroscedasticity in ϵ, apply GLS estimates:
    $$\hat{\mathbf{f}}_t = [\hat{B}'\hat{\mathbf{\Psi}}^{-1}\hat{B}]^{-1}[\hat{B}'\hat{\mathbf{\Psi}}^{-1}(\mathbf{x}_t - \hat{\alpha})]$$
- (Optional) Consider coordinate rotations of orthonormal factors as alternate interpretations of model.

Factor Analysis Model

Further Details of ML Specification

- Estimated factor realizations can be rescaled to represent **factor mimicking portfolios**
- Likelihood Ratio test can be applied to test for the number of factors.

  Test Statistic: $LR(K) = 2[l(\tilde{\alpha}, \tilde{\Sigma}) - l(\hat{\alpha}, \hat{B}, \hat{\Psi})]$

  where H_0: K factors are sufficient to model $\mathbf{\Sigma}$ and $\tilde{\alpha}$ and $\tilde{\Sigma}$ are the MLEs with no factor-model restrictions.

Outline

1. Factor Models
   - Linear Factor Model
   - Macroeconomic Factor Models
   - Fundamental Factor Models
   - Statistical Factor Models: Factor Analysis
   - Principal Components Analysis
   - Statistical Factor Models: Principal Factor Method

Principal Components Analysis (PCA)

- An $m-$variate random variable:

$$\mathbf{x} = \begin{bmatrix} x_1 \\ \vdots \\ x_m \end{bmatrix}, \text{ with } E[\mathbf{x}] = \boldsymbol{\alpha} \in \Re^m, \text{ and } Cov[\mathbf{x}] = \mathbf{\Sigma}_{(m \times m)}$$

- Eigenvalues/eigenvectors of $\mathbf{\Sigma}$:
  - $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_m \geq 0$: m eigenvalues.
  - $\gamma_1, \gamma_2, \dots, \gamma_m$: m orthonormal eigenvectors:

$$\begin{aligned} \mathbf{\Sigma}\gamma_i &= \lambda_i \gamma_i, \quad i = 1, \dots, m \\ \gamma_i' \gamma_i &= 1, \quad \forall i \\ \gamma_i' \gamma_{i'} &= 0, \quad \forall i \neq i' \end{aligned}$$

  - $\mathbf{\Sigma} = \sum_{i=1}^m \lambda_i \gamma_i \gamma_i'$
- Principal Component Variables:

$$p_i = \gamma_i'(\mathbf{x} - \boldsymbol{\alpha}), \quad i = 1, \dots, m$$

Principal Components Analysis

Principal Components in Vector/Matrix Form

- $m-$Variate $\mathbf{x}$: $\quad E[\mathbf{x}] = \boldsymbol{\alpha},\ Cov[\mathbf{x}] = \boldsymbol{\Sigma}$
- $\boldsymbol{\Sigma} = \boldsymbol{\Gamma}\boldsymbol{\Lambda}\boldsymbol{\Gamma}'$, where

$$\begin{aligned} \boldsymbol{\Lambda} &= diag(\lambda_1, \lambda_2, \ldots, \lambda_m) \\ \boldsymbol{\Gamma} &= [\gamma_1 : \gamma_2 : \ \cdots \ : \gamma_m] \\ \boldsymbol{\Gamma}'\boldsymbol{\Gamma} &= \mathbf{I}_m \end{aligned}$$

- $\mathbf{p} = \begin{bmatrix} p_1 \\ \vdots \\ p_m \end{bmatrix} = \boldsymbol{\Gamma}'(\mathbf{x} - \boldsymbol{\alpha})$, $m-$Variate PC variables

$$\begin{aligned} E[\mathbf{p}] &= E[\boldsymbol{\Gamma}'(\mathbf{x} - \boldsymbol{\alpha})] = \boldsymbol{\Gamma}' E[(\mathbf{x} - E[\mathbf{x}])] = \mathbf{0}_m \\ Cov[\mathbf{p}] &= Cov[\boldsymbol{\Gamma}'(\mathbf{x} - \boldsymbol{\alpha})] = \boldsymbol{\Gamma}' Cov[\mathbf{x}]\boldsymbol{\Gamma} \\ &= \boldsymbol{\Gamma}'\boldsymbol{\Sigma}\boldsymbol{\Gamma} = \boldsymbol{\Gamma}'(\boldsymbol{\Gamma}\lambda\boldsymbol{\Gamma}')\boldsymbol{\Gamma} = \boldsymbol{\Lambda} \end{aligned}$$

- $\mathbf{p}$ is a vector of zero-mean, uncorrelated random variables that provides an *orthogonal basis* for $\mathbf{x}$.

Principal Components Analysis

$m-$Variate $\mathbf{x}$ in Principal Components Form

- $\mathbf{x} = \begin{bmatrix} x_1 \\ \vdots \\ x_m \end{bmatrix} = \boldsymbol{\alpha} + \mathbf{\Gamma p}$, where $E[\mathbf{p}] = \mathbf{0}_m$, $Cov[\mathbf{p}] = \mathbf{\Lambda}$
- Partition $\mathbf{\Gamma} = [\mathbf{\Gamma}_1 \mathbf{\Gamma}_2]$ where $\mathbf{\Gamma}_1$ corresponds to the K $(< m)$ largest eigenvalues of $\mathbf{\Sigma}$.
- Partition $\mathbf{p} = \begin{bmatrix} \mathbf{p}_1 \\ \mathbf{p}_2 \end{bmatrix}$ where $\mathbf{p}_1$ contains the first K elements.
- $\mathbf{x} = \boldsymbol{\alpha} + \mathbf{\Gamma}_1 \mathbf{p}_1 + \mathbf{\Gamma}_2 \mathbf{p}_2 = \boldsymbol{\alpha} + B\mathbf{f} + \boldsymbol{\epsilon}$
  where
  $$\begin{array}{lcll} B & = & \mathbf{\Gamma}_1 & (m \times K) \\ \mathbf{f} & = & \mathbf{p}_1 & (K \times 1) \\ \boldsymbol{\epsilon} & = & \mathbf{\Gamma}_2 \mathbf{p}_2 & (m \times 1) \end{array}$$
  Like factor model except $Cov[\boldsymbol{\epsilon}] = \mathbf{\Gamma}_2 \mathbf{\Lambda}_2 \mathbf{\Gamma}_2'$, where $\mathbf{\Lambda}_2$ is diagonal matrix of last $(m - K)$ eigenvalues.

Empirical Principal Components Analysis

The principal components analysis of

$$\mathbf{X} = [\mathbf{x}_1 : \cdots \mathbf{x}_T] \ (m \times T)$$

consists of the following computational steps:

- Component/row means : $\bar{\mathbf{x}} = (\frac{1}{T})\mathbf{X}\mathbf{1}_T$
- 'De-meaned' matrix: $\mathbf{X}^* = \mathbf{X} - \bar{\mathbf{x}}\mathbf{1}'_T$
- Sample covariance matrix: $\hat{\mathbf{\Sigma}}_x = \frac{1}{T}\mathbf{X}^*(\mathbf{X}^*)'$
- Eigenvalue/vector decomposition: $\hat{\mathbf{\Sigma}}_x = \hat{\mathbf{\Gamma}}\hat{\mathbf{\Lambda}}\hat{\mathbf{\Gamma}}'$ yielding estimates of $\mathbf{\Gamma}$ and $\mathbf{\Lambda}$.
- Sample Principal Components:

$$\mathbf{P} = [\mathbf{p}_1 : \cdots : \mathbf{p}_T] = \hat{\mathbf{\Gamma}}'\mathbf{X}^*. \ (m \times T)$$

Empirical Principal Components Analysis

PCA Using Singular Value Decomposition

Consider the Singular Value Decomposition (SVD) of the de-meaned matrix:

$$\mathbf{X}^* = \mathbf{V}\mathbf{D}\mathbf{U}'$$

where

- $\mathbf{V}$: $(m \times m)$ orthogonal matrix, $\mathbf{V}\mathbf{V}' = \mathbf{I}_m$.
- $\mathbf{U}$: $(m \times T)$ row-orthonormal matrix, $\mathbf{U}\mathbf{V}' = \mathbf{I}_m$.
- $\mathbf{D}$: $(m \times m)$ diagonal matrix, $\mathbf{D} = diag(d_1, \ldots, d_m)$ with $d_1 \geq d_2 \geq \cdots \geq 0$.

Exercise: Show that

- $\hat{\mathbf{\Lambda}} = \frac{1}{T}\mathbf{D}^2$
- $\hat{\mathbf{\Gamma}} = \mathbf{V}$
- $\mathbf{P} = \hat{\mathbf{\Gamma}}'\mathbf{X}^* = \mathbf{D}\mathbf{U}'$

Alternate Definition of PC Variables

Given the $m-$variate $\mathbf{x}: E[\mathbf{x}] = \alpha$ and $Cov[\mathbf{x}] = \mathbf{\Sigma}$

- Define the **First Principal Component Variable** as
  $$p_1 = \mathbf{w}'\mathbf{x} = (w_1 x_1 + w_2 x_2 + \cdots + w_m x_m)$$
  where the coefficients $\mathbf{w} = (w_1, w_2, \ldots, w_m)'$ are chosen to
  maximize: $Var(p_1) = \mathbf{w}'\mathbf{\Sigma}_x\mathbf{w}$
  subject to: $|\mathbf{w}|^2 = \sum_{i=1}^m w_i^2 = 1$.
- Define the **Second Principal Component Variable** as
  $$p_2 = \mathbf{v}'\mathbf{x} = (v_1 x_1 + v_2 x_2 + \cdots + v_m x_m)$$
  where the coefficients $\mathbf{v} = (v_1, v_2, \ldots, v_m)'$ are chosen to
  maximize: $Var(p_2) = \mathbf{v}'\mathbf{\Sigma}_x\mathbf{v}$
  subject to: $|\mathbf{v}|^2 = \sum_{i=1}^m v_i^2 = 1$, **and** $\mathbf{v}'\mathbf{w} = 0$.
- Etc., defining up to p_m, The coefficient vectors are given by
  $$[\mathbf{w} : \mathbf{v} : \cdots] = [\gamma_1 : \gamma_2 : \cdots] = \mathbf{\Gamma}$$

Principal Components Analysis

Further Details

- PCA provides a decomposition of the **Total Variance**:

$$\begin{aligned}
\text{Total Variance }(\mathbf{x}) &= \sum_{i=1}^{m} Var(\mathbf{x}_i) = trace(\mathbf{\Sigma}_x) \\
&= trace(\mathbf{\Gamma\Lambda\Gamma}') = trace(\mathbf{\Lambda\Gamma}'\mathbf{\Gamma}) = trace(\mathbf{\Lambda}) \\
&= \sum_{k=1}^{m} \lambda_k \\
&= \sum_{k=1}^{m} Var(p_k) \\
&= \text{Total Variance }(\mathbf{p})
\end{aligned}$$

- The transformation from $\mathbf{x}$ to $\mathbf{p}$ is a change in coordinate system which shifts the origin to the mean/expectation $E[\mathbf{x}] = \alpha$ and rotates the coordinate axes to align with the Principal Component Variables. Distance in the space is preserved (due to orthogonality of the rotation).

Outline

1. Factor Models
   - Linear Factor Model
   - Macroeconomic Factor Models
   - Fundamental Factor Models
   - Statistical Factor Models: Factor Analysis
   - Principal Components Analysis
   - Statistical Factor Models: Principal Factor Method

Factor Analysis Model

For $\{\mathbf{x}_t, t = 1, \ldots, T\}$, the factor model is:

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

- $\boldsymbol{\alpha}$ and B are vector/matrix constants.
- All random variables are Normal/Gaussian:
  - $\mathbf{x}_t$ i.i.d. $N_m(\boldsymbol{\alpha}, \boldsymbol{\Sigma}_x)$
  - $\mathbf{f}_t$ i.i.d. $N_K(\mathbf{0}_K \mathbf{I}_K)$
  - ϵ_t i.i.d. $N_m(\mathbf{0}_m, \boldsymbol{\Psi})$
- $Cov(\mathbf{x}_t) = \boldsymbol{\Sigma}_x = BB' + \boldsymbol{\Psi}$

Principal Factor Method of Estimation

To fit a $K-$factor model with fixed $K < m$, define

$$\mathbf{X} = [\mathbf{x}_1 : \cdots \mathbf{x}_T] \ {\scriptstyle (m \times T)}$$

Principal Factor Method of Estimation

Step 1: Conduct the computational steps of principal components analysis:

- Component/row means : $\bar{\mathbf{x}} = (\frac{1}{T})\mathbf{X}\mathbf{1}_T$
- 'De-meaned' matrix: $\mathbf{X}^* = \mathbf{X} - \bar{\mathbf{x}}\mathbf{1}'_T$
- Sample covariance matrix: $\hat{\boldsymbol{\Sigma}}_x = \frac{1}{T}\mathbf{X}^*(\mathbf{X}^*)'$
- Eigenvalue/vector decomposition: $\hat{\boldsymbol{\Sigma}}_x = \hat{\boldsymbol{\Gamma}}\hat{\boldsymbol{\Lambda}}\hat{\boldsymbol{\Gamma}}'$ yielding estimates of $\boldsymbol{\Gamma}$ and $\boldsymbol{\Lambda}$.

Step 2: Specify initial estimates (index $s = 0$)

- $\tilde{\alpha}_0 = \bar{\mathbf{x}}$
- $\tilde{B}_0 = \hat{\Gamma}^{(K)}(\hat{\Lambda}^{(K)})^{\frac{1}{2}}$, where
  $\hat{\Gamma}^{(K)}$ is submatrix of $\hat{\boldsymbol{\Gamma}}$ (first K columns)
  $\hat{\Lambda}^{(K)}$ is submatrix of $\hat{\boldsymbol{\Lambda}}$ (first K columns)
- $\tilde{\boldsymbol{\Psi}}_0 = diag(\hat{\boldsymbol{\Sigma}}_x) - diag(\tilde{B}_0\tilde{B}_0')$
- $\tilde{\boldsymbol{\Sigma}}_0 = \tilde{B}_0\tilde{B}_0' + \tilde{\boldsymbol{\Psi}}_0$

Principal Factor Method of Estimation

Step 3: Adjust the sample covariance matrix to

$$\hat{\mathbf{\Sigma}}_x^* = \hat{\mathbf{\Sigma}}_x - \tilde{\mathbf{\Psi}}_0$$

- Compute the eigenvalue/vector decomposition:

$$\hat{\mathbf{\Sigma}}_x^* = \tilde{\mathbf{\Gamma}}\tilde{\mathbf{\Lambda}}\tilde{\mathbf{\Gamma}}'$$

  yielding updated estimates of $\mathbf{\Gamma}$ and $\mathbf{\Lambda}$
- Repeat Step 2 with these new estimates obtaining $\tilde{B}_1$, $\tilde{\mathbf{\Psi}}_1$, $\tilde{\mathbf{\Sigma}}_1 = \tilde{B}_1\tilde{B}_1' + \tilde{\mathbf{\Psi}}_1$

Step 4: Repeat Step 3 generating a sequence of estimates $(\tilde{B}_s, \tilde{\mathbf{\Psi}}_s, \tilde{\mathbf{\Sigma}}_s)$ $s = 1, 2, \ldots,$ until successive changes in $\tilde{\mathbf{\Psi}}_s$ are sufficiently negligible.

Step 5: Use the estimates from the last iteration in Step 4.

18.S096 Topics in Mathematics with Applications in Finance
Fall 2013