Factor Models

Silvia Miranda-Agrippino
Bank of England and CFM

University of Surrey
May 8, 2017

Acknowledgment: This lecture is partly based on the lecture notes of Domenico Giannone.

[▷]

Outline

i. Introduction & Notation

ii. SFM with Spherical Idiosyncratic Variance
- ▷ Estimation: ML
- ▷ Estimation: PCA

iii. SFM with Diagonal Idiosyncratic Variance
- ▷ Estimation: ML

iv. Dynamic Factor Model
- ▷ Estimation: EM & Kalman Filter
- ▷ Approximate Factor Model: Estimation
- ▷ DFM for Large Cross-Sections

Introduction

Intuition

- Generic class of models designed to capture comovements among n (potentially very large!) series using r, ($r << n$) common (unobserved) factors.

- Factor Analysis (FA) is a way to enforce parsimony (i.e. reduce the dimensionality of problems) when variables are correlated.

Notation

- Let x_t be a $(n \times 1)$ vector of observables: $x_t = [x_1, \ldots, x_n]'$

$$\underset{[n\times 1]}{x_t} = \underbrace{\underset{[n\times r]}{\Lambda}\underset{[r\times 1]}{F_t}}_{\text{common component}} + \underbrace{\underset{[n\times 1]}{e_t}}_{\text{idiosyncratic component}}$$

- factors ➔ $F_t = [f_1, \ldots, f_r]'$
- loadings ➔ λ_{ij} $(i = 1, \ldots, n$ and $k = 1, \ldots, r)$
- idiosyncratic ➔ $e_t = [e_1, \ldots, e_n]'$

Static Factor Model with Spherical Idiosyncratic Variance

$$x_t = \Lambda F_t + e_t$$

i. F_t is i.i.d. Gaussian: $\mathbb{E}[F_t F_s'] = 0 \qquad \forall s \neq t$

ii. e_t is i.i.d. Gaussian: $\mathbb{E}[e_t e_s'] = 0 \qquad \forall s \neq t$

iii. e_t has spherical variance: $\mathbb{E}[e_t e_t'] = \sigma^2 \mathbb{I}_n$:
   - ▷ $\text{var}(e_{it}) = \text{var}(e_{jt}) = \sigma^2$
   - ▷ $\text{cov}(e_{it}, e_{jt}) = 0 \ \forall i \neq j$

iv. orthogonal common and idio components $\mathbb{E}[F_t e_t'] = 0$.

$$x_t = \Lambda F_t + e_t$$

$$\begin{pmatrix} F_t \\ e_t \end{pmatrix} \sim i.i.d.\ \mathcal{N}\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \mathbb{I}_r & 0 \\ 0 & \sigma^2 \mathbb{I}_n \end{pmatrix}\right)$$

Remark 1: variance of F_t is set to $\mathbb{I}_r$ for identification purposes: $x_t = \Lambda Q Q^{-1} F_t + e_t$ with $F_t \sim \mathcal{N}(0, QQ^{-1})$ would be observationally equivalent.

Remark 2: F_t is a random variable: we can only uniquely define its moments.

SFM with Spherical Idio Variance

$$x_t = \Lambda F_t + e_t$$

- variance of data:

$$\begin{aligned}\text{var}(x_t) &= \mathbb{E}[x_t x_t'] = \mathbb{E}[(\Lambda F_t + e_t)(\Lambda F_t + e_t)'] \\ &= \Lambda\Lambda' + \sigma^2 \mathbb{I}_n\end{aligned}$$

- covariance between data and factors:

$$\begin{aligned}\text{cov}(x_t, F_t) &= \mathbb{E}[x_t F_t'] = \mathbb{E}[(\Lambda F_t + e_t)F_t'] \\ &= \Lambda\end{aligned}$$

JOINT DISTRIBUTION OF x_t AND F_t

$$\begin{pmatrix} F_t \\ x_t \end{pmatrix} \sim i.i.d.\ \mathcal{N}\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \mathbb{I}_r & \Lambda' \\ \Lambda & \Lambda\Lambda' + \sigma^2\mathbb{I}_n \end{pmatrix}\right)$$

Distribution of $\hat{F}_t$

- $\hat{F}_t = \mathbb{E}_\theta[F_t|x^T] = \mathbb{E}_\theta[F_t x_t']\mathbb{E}_\theta[x_t x_t']^{-1} x_t$

CONDITIONAL DISTRIBUTION OF $\hat{F}_t$

$$\hat{F}_t = \mathbb{E}[F_t|x^T] \sim i.i.d.\ \mathcal{N}\left[\left(\frac{\Lambda'\Lambda}{\sigma^2} + \mathbb{I}_r\right)^{-1} \frac{\Lambda'}{\sigma^2} x_t\ ,\ \left(\frac{\Lambda'\Lambda}{\sigma^2} + I_r\right)^{-1}\right]$$

Remark 3: With $r = 1$ and $\Lambda = [1, \ldots, 1]'$

$$\hat{F}_t = \left(\frac{n}{\sigma^2} + 1\right)^{-1} \sum_{i=1}^{n} \frac{x_{it}}{\sigma^2} = \frac{1}{n + \sigma^2} \sum_{i=1}^{n} x_{it}$$

factors are proportional to the cross-sectional average of x_t .

Indeterminacy, the Role of Large n

▷ General case:

$$\mathbb{E}[F_t|x^T] = \left(\frac{\Lambda'\Lambda}{\sigma^2} + \mathbb{I}_r\right)^{-1} \frac{\Lambda'}{\sigma^2} x_t$$

$$= \left(\frac{\Lambda'\Lambda}{\sigma^2} + \mathbb{I}_r\right)^{-1} \frac{\Lambda'\Lambda}{\sigma^2} F_t + \left(\frac{\Lambda'\Lambda}{\sigma^2} + \mathbb{I}_r\right)^{-1} \frac{\Lambda'}{\sigma^2} e_t$$

▷ With $r = 1$ and $\Lambda = [1, \ldots, 1]'$:

▷ $$\mathbb{E}[F_t|x^T] = \left(\frac{1}{n+\sigma^2}\right) \sum_{i=1}^{n} x_{it}$$

$$= \left(\frac{n}{n+\sigma^2}\right) F_t + \left(\frac{n}{n+\sigma^2}\right) \frac{1}{n} \sum_{i=1}^{n} e_{it}$$

$\text{var}\left(\frac{1}{n}\sum_{i=1}^{n} e_{it}\right) = \frac{\sigma^2}{n}$ ➔ $\hat{F}_t$ is a consistent estimator of F_t (i.e. $\hat{F}_t$ converges to F_t) with $n \to \infty$.

$$x_t \sim i.i.d.\ \mathcal{N}\left[0,\ \Lambda\Lambda' + \sigma^2 \mathbb{I}_n\right]$$

▷ for each $x_t,\ t \in [1, T]$:

$$\Pr(x_t = \tilde{x}_t | \Lambda, \sigma^2) = \frac{n}{\sqrt{2\pi}} |\Lambda\Lambda' + \sigma^2 \mathbb{I}_n|^{-1/2} \exp\left\{-\frac{1}{2} x_t' (\Lambda\Lambda' + \sigma^2 \mathbb{I}_n)^{-1} x_t\right\}$$

Λ and σ^2 are the maximizers (minimizers) of the (minus) log likelihood function:

$$\{\Lambda, \sigma^2\} = \underset{\Lambda, \sigma^2}{\operatorname{argmin}}\ \mathcal{L}(x^T, \Lambda, \sigma^2)$$

SFM Estimation: ML

$$\{\Lambda, \sigma^2\} = \underset{\Lambda, \sigma^2}{\operatorname{argmin}} \; \mathcal{L}(x^T, \Lambda, \sigma^2)$$

$$\begin{aligned}\mathcal{L}(x^T, \Lambda, \sigma^2) &\propto \frac{T}{2} \log\left(|\Lambda\Lambda' + \sigma^2 \mathbb{I}_n|\right) - \frac{1}{2}\left(\sum_{t=1}^{T} x_t'(\Lambda\Lambda' + \sigma^2 \mathbb{I}_n)^{-1} x_t\right) \\ &= \frac{T}{2} \log\left(|\Lambda\Lambda' + \sigma^2 \mathbb{I}_n|\right) - \frac{T}{2} \operatorname{tr}(\Lambda\Lambda' + \sigma^2 I_N)^{-1} S\end{aligned}$$

where $S = T^{-1} \sum_{t=1}^{T} x_t x_t'$ is the sample covariance.

ML SOLUTION

$$\hat{\Lambda} = V\left(D - \hat{\sigma}^2 \mathbb{I}_r\right)^{1/2}$$

$$\hat{\sigma}^2 = \frac{1}{n}\mathrm{tr}\left(S - \hat{\Lambda}\hat{\Lambda}'\right)$$

Where:

- $SD = VD$
- $\underset{[r\times 1]}{D} \equiv \begin{pmatrix} d_1 & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & d_r \end{pmatrix}$ ➔ r largest eigenvalues of S
- $\underset{[n\times r]}{V} \equiv [v_1, \ldots, v_r]'$ ➔ associated eigenvectors

PCA

Use the eigenvectors of S as a (orthonormal) basis to transform the correlated series in x_t into a set of orthogonal components z_t.

↓

z_t are the normalized <u>principal components</u> of x_t

$$z_{jt} = \frac{1}{\sqrt{d_j}} v_j' x_t = \frac{1}{\sqrt{d_j}} \sum_{i=1}^{n} v_{ij}' x_{jt}$$

- eigenvalues in decreasing order: $d_1 \geq d_2 \geq \ldots \geq d_n$

$$v_i' v_j = \begin{cases} 1 & i = j \\ 0 & i \neq j \end{cases} \qquad v_i' S v_j = \begin{cases} d_i & i = j \\ 0 & i \neq j \end{cases}$$

SFM Estimation: PC

- Projecting x_t onto its first principal component reduces the sample variance S by d_1

$$
\begin{aligned}
\hat{x}_t^{(j)} &= \left(\frac{1}{T}\sum_{t=1}^{T} z_{jt}^2\right)^{-1}\left(\frac{1}{T}\sum_{t=1}^{T} z_{jt}x_t\right) z_{jt} \\
&= \left(\frac{1}{T}\sum_{t=1}^{T} \frac{1}{d_j} v_j' x_t x_t' v_j'\right)^{-1}\left(\frac{1}{T}\sum_{t=1}^{T} \frac{1}{d_j} v_j' x_t x_t'\right) \frac{v_j' x_t}{\sqrt{d_j}} \\
&= v_j v_j' x_t
\end{aligned}
$$

$$
\begin{aligned}
&\operatorname{tr}\left(\frac{1}{T}\sum_{t=1}^{T}\left(x_t - x_t^{(j)}\right)\left(x_t - x_t^{(j)}\right)'\right) = \\
&= \operatorname{tr}(S - d_j) = \sum_{i=1}^{n} d_i - d_j
\end{aligned}
$$

- Generalizing to projection onto first r principal components:

$$\mathrm{Proj}(x_t|z_{1t},\ldots,z_{rt}) = \sum_{j=1}^{r} v_j v_j' x_t$$

$$\Downarrow$$

$$\begin{aligned} S &= \sum_{j=1}^{r} v_j v_j' d_j + \sum_{j=r+1}^{n} v_j v_j' d_j \\ &= V_{(r)} V_{(r)}' D_{(r)} + V_{(N-r)} V_{(N-r)}' D_{(N-r)} \end{aligned}$$

Let $z_t^{(r)}$ be the first r principal components of x_t:

i. $z_t^{(r)} = (z_{1t}, \dots, z_{rt})' = D_{(r)}^{-1/2} V_{(r)}' x_t$

ii. $\text{var}(z_t^{(r)}) = \frac{1}{T}\sum_{t=1}^{T} z_t^{(r)} z_t^{(r)'} = \mathbb{I}_r$

iii. $\text{Proj}(x_t | z_t^{(r)}) = V_{(r)} V_{(r)}' x_t$

iv. projection coefficients

$$\Theta = \left(\frac{1}{T} \sum_{t=1}^{T} z_t^{(r)} z_t^{(r)'} \right)^{-1} \left(\frac{1}{T} \sum_{t=1}^{T} z_t^{(r)} x_t \right) = D_{(r)}^{1/2} V_{(r)}$$

- ML:

$$\hat{\Lambda} = V_{(r)} \left(D_{(r)} - \hat{\sigma}^2 \mathbb{I}_r\right)^{1/2}$$
$$\hat{F}_t = \left(D_{(r)} - \hat{\sigma}^2 \mathbb{I}_r\right)^{1/2} D_{(r)}^{-1} V'_{(r)} x_t$$

- PC:

$$\hat{\Theta} = D_{(r)}^{1/2} V_{(r)}$$
$$\hat{z}_t = D_{(r)}^{-1/2} V'_{(r)} x_t$$

If $\text{var}(e_t) = \sigma^2 \mathbb{I}_n$, PC are proportional to the estimated common factors, and converge to the true factors for $\sigma^2 \to 0$.

Static Factor Model with Diagonal Idiosyncratic Variance

SFM with Diagonal Idio Variance

$$x_t = \Lambda F_t + e_t$$

$$\begin{pmatrix} F_t \\ e_t \end{pmatrix} \sim i.i.d.\ \mathcal{N}\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \mathbb{I}_r & 0 \\ 0 & \Psi \end{pmatrix} \right)$$

- The idiosyncratic term is heteroskedastic, but still i.i.d. ➔ i.e. no cross- or time- dependence

JOINT DISTRIBUTION OF x_t AND F_t

$$\begin{pmatrix} F_t \\ x_t \end{pmatrix} \sim i.i.d.\ \mathcal{N}\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \mathbb{I}_r & \Lambda' \\ \Lambda & \Lambda\Lambda' + \Psi \end{pmatrix} \right)$$

Distribution of $\hat{F}_t$

- $\hat{F}_t = \mathbb{E}_\theta[F_t | x^T] = \mathbb{E}_\theta[F_t x_t']\mathbb{E}_\theta[x_t x_t']^{-1} x_t$

CONDITIONAL DISTRIBUTION OF $\hat{F}_t$

$$\hat{F}_t \sim i.i.d.\ \mathcal{N}\left[\left(\Lambda'\Psi^{-1}\Lambda + \mathbb{I}_r\right)^{-1}\Lambda'\Psi^{-1}x_t\ ,\ \left(\Lambda'\Psi^{-1}\Lambda + I_r\right)^{-1}\right]$$

Estimation via ML:

- Iteratively Weighted PC
- EM Algorithm [Dempster et al. 1977, Rubin and Thayer, 1982]

Intuition: Use estimate of Ψ to transform the model into spherical idio variance, iterate until (log) likelihood convergence.

At each iteration j:

i. Given $\Psi_{[j-1]}$:

- $x_t = \Lambda F_t + e_t, \qquad \text{var}(e_t) = \Psi_{[j-1]}$
- $\underbrace{\Psi_{[j-1]}^{-1/2} x_t}_{\tilde{x}_t} = \underbrace{\Psi_{[j-1]}^{-1/2} \Lambda}_{\tilde{\Lambda}} F_t + \underbrace{\Psi_{[j-1]}^{-1/2} e_t}_{\tilde{e}_t}$
- $\text{var}(\tilde{e}_t) = \mathbb{I}_n$

ii. PC to the transformed model to estimate $\Lambda_{[j]}$, $\hat{F}_t$, and $\Psi_{[j]}$:

- $\tilde{S}_{[j]} = \frac{1}{T}\sum_{t=1}^{T} \tilde{x}_t \tilde{x}_t' = \frac{1}{T}\sum_{t=1}^{T} \Psi_{[j-1]}^{-1/2} x_t x_t' \Psi_{[j-1]}^{-1/2}$
- $\left\{\tilde{V}, \tilde{D}: \quad \Psi_{[j-1]}^{-1/2}\left(\tilde{S}_{[j]} - \Psi_{[j-1]}\right)\Psi_{[j-1]}^{-1/2}\tilde{V} = \tilde{V}\tilde{D}\right\}$
- $\Lambda_{[j]} = \Psi_{[j-1]}^{1/2}\tilde{V}_{(r)}(\tilde{D}_{(r)} - \mathbb{I}_r)^{-1/2}$
- $\Psi_{[j]} = \text{diag}(S - \Lambda_{[j]}\Lambda_{[j]}')$
- $\mathcal{L}_{[j]}(x^T, \Lambda_{[j]}, \Psi_{[j]}) \propto \log|\Lambda_{[j]}\Lambda_{[j]}' + \Psi_{[j]}| + \text{tr}(S[\Lambda_{[j]}\Lambda_{[j]}' + \Psi_{[j]}]^{-1})$

The algorithm is initialized with a value for Ψ (`Psi_init`)
<u>reference MatLab code:</u> `SFM_ZZ.ms`.

EM Algorithm

- With unknown factors, the exact form of the likelihood is unknown
- Maximize the expected likelihood given initial estimates of models' parameters
- Substitute sample statistics with expected sufficient statistics ➔ equivalent to maximizing $\mathbb{E}[\mathcal{L}(x^T|F^T,\Lambda,\Psi)]$.

JOINT LOG-LIKELIHOOD

$$\mathcal{L}(x^T,F^T,\Lambda,\Psi) = \mathcal{L}(x^T|F^T,\Lambda,\Psi) + \mathcal{L}(F^T,\Lambda,\Psi)$$

↓

$$\{\Lambda_{[j]},\Psi_{[j]}\} = \underbrace{\operatorname*{argmax}_{\Lambda,\Psi}}_{\text{M-step}} \underbrace{\mathbb{E}_{\Lambda_{[j]},\Psi_{[j]}}\left[\mathcal{L}(x^T,F_t,\Lambda_{[j-1]},\Psi_{[j-1]})\right]}_{\text{E-step}}$$

At each iteration j:

- $\underset{[r\times T]}{F_{[j]}} = \underset{[n\times T]}{x} \Psi_{[j-1]}^{-1} \left(\Lambda'_{[j-1]} \Psi_{[j-1]}^{-1} \Lambda_{[j-1]} + \mathbb{I}_r \right)^{-1}$

- $\Lambda_{[j]} = x' F_{[j]} \left(F'_{[j]} F_{[j]} + \mathrm{var}(F_{[j]}) \right)^{-1}$

- $\Psi_{[j]} = \mathrm{diag} \left(\frac{x'x}{T} - \frac{x' F_{[j]}}{T} \left(F'_{[j]} F_{[j]} + \mathrm{var}(F_{[j]}) \right)^{-1} \frac{F'_{[j]} X}{T} \right)$

- $\mathcal{L}_{[j]}(X^T, F^T, \Lambda, \Psi) \propto$
  $\log |\Psi_{[j]}^{-1}| + \mathrm{tr} \left(\Psi_{[j]}^{-1} \frac{1}{T} \sum_{t=1}^{T} (X - \Lambda_{[j]} F_{[j]})'(X - \Lambda_{[j]} F_{t[j]}) \right)$

Expected Sufficient Statistics

- With known factors:

$$\hat{\Lambda} = \left(\frac{1}{T} \sum_{t=1}^{T} F_t F_t' \right)^{-1} \left(\frac{1}{T} \sum_{t=1}^{T} F_t x_t' \right)$$

- With unknown factors:

$$\hat{\Lambda} = \mathbb{E}_{\Lambda,\Psi} \left(\frac{1}{T} \sum_{t=1}^{T} F_t F_t' \right)^{-1} \mathbb{E}_{\Lambda,\Psi} \left(\frac{1}{T} \sum_{t=1}^{T} F_t x_t' \right)$$

  - $\mathbb{E}_{\Lambda,\Psi} \left(\frac{1}{T} \sum_{t=1}^{T} F_t F_t' \right) = \frac{1}{T} \sum_{t=1}^{T} \mathbb{E}_{\Lambda,\Psi} \left(F_t F_t' | x^T \right)$
  - $\mathbb{E}_{\Lambda,\Psi} \left(F_t F_t' | x^T \right) = \mathbb{E}[\hat{F}_t \hat{F}_t'] + \text{var}(\hat{F})$

reference MatLab code: `SFM_EM.m`

Dynamic Factor Model

Two possible alternatives:

- **Pseudo-ML:** treat the data as if they were i.i.d. ➔ estimate SFM
    i. If the data are stationary, the covariance matrix is consistently estimated
    ii. Factor loadings and idiosyncratic variances are consistently estimated

- **Dynamic Factor Model:** parameterize the serial dependence

$$x_t = \Lambda F_t + e_t$$

- $F_t = A_1 F_{t-1} + \ldots + A_p F_{t-p} + u_t$
- $u_t \sim i.i.d.\ \mathcal{N}(0, \Sigma_u)$
- $e_t \sim i.i.d.\ \mathcal{N}(0, \Psi)$

Remark 4: Ψ is diagonal ➔ impose cross-sectional orthogonality (i.e. exact factor structure) but allows heteroskedasticity.
[Sargent and Sims (1977), Watson and Engle (1981)]

▷ STEP 1: <u>initialization</u>

i. Compute the first r sample principal components ➔ $F_{pc,t}$

ii. Estimate the model's parameters $\hat{\Lambda}$, $\hat{A}(L)$, $\hat{\Psi}$ using OLS ➔ treat $F_{pc,t}$ as the true common factors

Note: Because of the Pseudo-ML result, initializations is typically quite good

- STEP 2: factors extraction

  i. Condition on initial parameters' estimates $\theta \equiv \left\{\hat{\Lambda},\ \hat{A}(L),\ \hat{\Psi}\right\}$

  ii. Use Kalman Smoother to obtain a new estimate of the factors

$$\hat{F}_{\theta,t} = \mathbb{E}_{\hat{\theta}}(F_t|x_1, \ldots, x_T).$$

Note: This is the two-step estimator of Doz, Giannone and Reichlin (2011).
[Giannone, Reichlin and Sala (2004); Giannone, Reichlin and Small (2005)]

▷ STEP 3: parameters re-estimation

i. Treat $\hat{F}_{\theta,t}$ as if they were the true factors

ii. Estimate model's parameters using OLS ➔ $\hat{\Lambda}$, $\hat{A}(L)$, $\hat{\Psi}$

Remark 5: If (a) replace sample statistics with expected sufficient statistics and (b) iterate ➔ EM algorithm.

Remark 6: Kalman Smoother produces all relevant elements to account for uncertainty in factors' estimates.

State Space

$$
\begin{aligned}
x_t &= C s_t + e_t, & e_t &\sim \mathcal{N}(0, R) \\
s_t &= A s_{t-1} + u_t, & u_t &\sim \mathcal{N}(0, Q)
\end{aligned}
$$

Objects of interest:

- $\mathbb{E}[s_t | x^t]$ ➔ factors
- $\mathbb{V}ar[s_t | x^t]$ ➔ factors' variance

- Given initial conditions $s_{0|0}$ and $P_{0|0}$
- Kalman Filter is an algorithm that delivers $\forall t \in [1, T]$

$$s_{t|t} \equiv \mathbb{E}\left[s_t | x^t\right]$$

$$P_{t|t} \equiv \mathbb{V}ar\left[s_t | x^t\right]$$

- Alternates between 2 steps
  i. Prediction ➔ $s_{t|t-1},\ P_{t|t-1}$
  ii. Update ➔ $s_{t|t},\ P_{t|t}$

JOINT CONDITIONAL DISTRIBUTION OF s_t AND x_t

$$\left(\begin{array}{c} s_t \\ x_t \end{array} \middle| x^{t-1} \right) \sim \mathcal{N}\left(\left(\begin{array}{c} s_{t|t-1} \\ x_{t|t-1} \end{array} \right), \left(\begin{array}{cc} P_{t|t-1} & B'_{t|t-1} \\ B_{t|t-1} & H_{t|t-1} \end{array} \right) \right)$$

Prediction

- $s_{t|t-1} = A s_{t-1|t-1}$
- $x_{t|t-1} = C s_{t|t-1}$
- $P_{t|t-1} = A P_{t-1|t-1} A' + Q$
- $B_{t|t-1} = C P_{t|t-1}$ and $H_{t|t-1} = C P_{t|t-1} C' + R$

Update

- $s_{t|t} = s_{t|t-1} + \underbrace{B'_{t|t-1} H^{-1}_{t|t-1}}_{\text{gain}} \left(x_t - x_{t|t-1} \right)$
- $P_{t|t} = P_{t|t-1} - B'_{t|t-1} H^{-1}_{t|t-1} B_{t|t-1}$

- Ultimately, we are interested in an estimate for the states (s_t) that is conditional on the entire sample:

$$s_{t|T} \equiv \mathbb{E}\left(s_t | x^T\right)$$
$$P_{t|T} \equiv \mathbb{V}ar\left(s_t | X^T\right)$$

where $X^T = (X_T, X_{T-1}, ..., X_1)$

- Works with backward recursion similar to KF
  i. Start with $s_{T|T}$ and $P_{T|T}$
  ii. Iterate backward for $s_{T-j|T}$ and $P_{T-j|T},\ j = 1, \ldots, T-1$

- If $x_{i,t}$ contains missing values, solve by attributing infinite variance in the observation equation $(R_t)_{i,i} = \infty$ ➔ the model becomes time-varying in presence of missing data!
- Alternative is to use selection matrices for data that are 'visible' at time t
- Low-frequency data as high-frequency ones but with periodically missing data ➔ e.g. quarterly variables function of latent monthly series with systematically missing data.

DFM for Large Cross-Sections

- The Factor Model is a good representation for correlated data.
- What happens when n is large?
  i. Computational feasibility?
  ii. Asymptotic properties?
- For identification and parsimony, the factor structure is assumed exact (i.e. diagonal idio variance) ➔ incredible restrictions?
  i. n large likely to to introduce correlation among idiosyncratic components
  ii. The larger n, the more incredible the restrictions...

[Stock and Watson (2002), Forni, Hallin, Lippi and Reichlin (2002), Bai and Ng (2002), Onatsky (2008), Doz, Giannone and Reichlin (2013), Bernanke and Boivin (2002)]

- If we increase the cross-section and common factors are pervasive

$$\lim_{n\to\infty} \lambda_{min}(\Lambda'\Lambda) \to \infty$$

$$\Downarrow$$

$$\lim_{n\to\infty} \mathbb{E}_\theta(\hat{F}_{\theta,t} - F_t)(\hat{F}_{\theta,t} - F_t)' = \lim_{n\to\infty} \left(\mathbb{I}_r + \frac{\Lambda'\Lambda}{\sigma^2}\right)^{-1} = 0$$

- If the idiosyncratic components is (weakly) serially and cross-sectionally correlated, we still have

$$\hat{F}_\theta \to F_t \quad \text{as} \quad n \to \infty$$

Approximate factor model

- Recall: $\mathbb{E}_\theta\left[F_t|x_1,\ldots,x_T\right] = \left(\Lambda'\Lambda + \sigma^2\mathbb{I}_r\right)^{-1}\Lambda'\Lambda F_t + \left(\Lambda'\Lambda + \sigma^2\mathbb{I}_r\right)^{-1}\Lambda' e_t$

  CR1 : Factors are pervasive: $\liminf_{n\to\infty}\frac{1}{n}\lambda_{min}(\Lambda'\Lambda) > 0$

  CR2 : Idio weakly cross-correlated: $\lim_{n\to\infty}\frac{1}{n}\lambda_{max}(\mathbb{E}\left[e_t e_t'\right]) = 0$

- then as $n \to \infty$

  i. $\left(\Lambda'\Lambda + \sigma^2\mathbb{I}_r\right)^{-1}\Lambda'\Lambda \to \mathbb{I}_r$

  ii. $\mathbb{V}ar\left[\left(\Lambda'\Lambda + \sigma^2\mathbb{I}_r\right)^{-1}\Lambda' e_t\right] \to 0$

↓

$$plim_{n\to\infty}\mathbb{E}_\theta\left[F_t|x_1,\ldots,x_T\right] = F_t$$

[Chamberlain and Rotshild, (1983), Two-Step and ML Estimator in Doz, Giannone and Reichlin (2011,2012)]

Maximum Likelihood for DFM

- $F = (F_1, \dots, F_T)'$ ➔ true common factors
- $F_{\hat{\theta}_{ML}} = (F_{\hat{\theta}_{ML},1}, \dots, f_{\hat{\theta}_{ML},T})'$ ➔ expected common factors estimated under $\hat{\theta}_{ML}$

Under CR1 and CR2

$$\frac{1}{T}\mathrm{tr}(F - \hat{K}\hat{F}_{\hat{\theta}_{ML}})'(F - \hat{K}\hat{F}_{\hat{\theta}_{ML}}) = O_p\left(\frac{1}{\Delta(n,T)}\right) \qquad \mathrm{as} n, T \to \infty$$

- $\Delta_{nT} = \min\left\{\sqrt{T}, \frac{n}{\log(n)}\right\}$
- $\hat{K} = \left(\hat{F}'_{\hat{\theta}_{ML}}\hat{F}_{\hat{\theta}_{ML}}\right)^{-1}\hat{F}'_{\hat{\theta}_{ML}}F$