Dynamic Factor Models

Anna Simoni[2]

[2]CREST, CNRS, Ensae, Ecole Polytechnique

Outline

Introduction. I

- The large datasets available today cover a number of years which is finite thus the number of data points is limited.

- On the other hand, more and more time series are collected and made available by statistical agencies.

- T = number of points in time ; N = number of series. Typically we are in a setting where $N \geq T$.

- The case $N \geq T$ in statistics is known as high-dimensional setting and dealing with such datasets by means of standard techniques as linear regressions constitutes a hard problem to be solved due to the lack of degrees of freedom.

- For example, in macroeconomic datasets we have $N \simeq 100, 1000$ and $T \simeq 100$ (quarterly or monthly series), while in financial datasets we have $N \simeq 100, 1000$ and $T \simeq 1000$ (daily series).

Introduction. II

- Factor models are a dimension reduction technique for analysing a large panel of time series.
- Parsimonious representation of the information.
- The recent availability of large datasets made them increasingly popular in the last twenty years.
- We suppose that the observed variables can be described by a small number of latent factors and that these common factors are the source of the correlation among the observable variables.

Suppose to have quarterly observations on 207 time series (see Table 1)

- The full span of the dataset is $1959Q1 - 2014Q4$. Only 145 of the 207 series are available for this full period.
- From this dataset (after preliminary transformation of the data) one common component (factor) is extracted by using principal components analysis (see later).
- Then, the quarterly growth rate of each series (real Gross Domestic Product (GDP), total nonfarm employment, IP, and manufacturing and trade sales) is regressed on this common factor.
- None of the 4 series were used to extract the factor.
- As it can be seen from the figure, the single factor explains a large fraction of the 4-quarter variation in these four series.

Introduction. IV

- The R^2s of the four-quarter fits range from 0.73 for GDP to 0.92 for employment.
- At the same time, the estimated factor does not equal any one of these series, nor does it equal any one of the 58 series used to construct it.

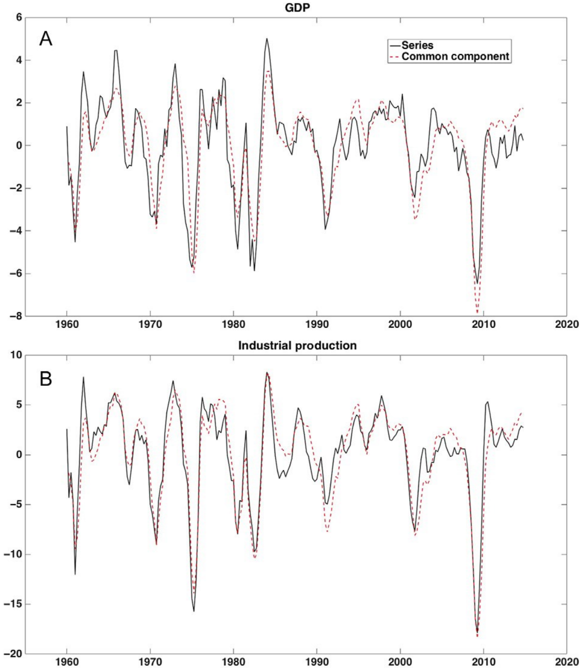
Source : Stock & Watson

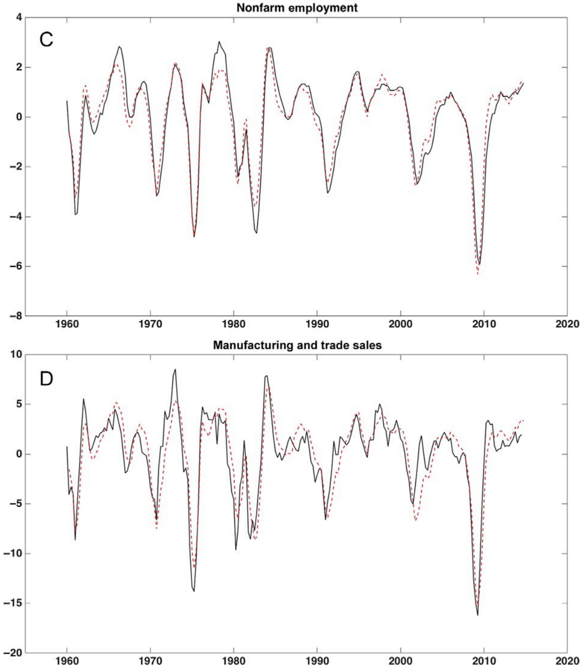
Table 1 Quarterly time series in the full dataset

	Category	Number of series	Number of series used for factor estimation
(1)	NIPA	20	12
(2)	Industrial production	11	7
(3)	Employment and unemployment	45	30
(4)	Orders, inventories, and sales	10	9
(5)	Housing starts and permits	8	6
(6)	Prices	37	24
(7)	Productivity and labor earnings	10	5
(8)	Interest rates	18	10
(9)	Money and credit	12	6
(10)	International	9	9
(11)	Asset prices, wealth, and household balance sheets	15	10
(12)	Other	2	2
(13)	Oil market variables	10	9
	Total	207	139

Notes: The real activity dataset consists of the variables in the categories 1–4.

Source : Stock & Watson




Nonfarm employment
C
4
2
0
−2
−4
−6
−8
1960
1970
1980
1990
2000
2010
2020
Manufacturing and trade sales
D
10
5
0
−5
−10
−15
−20
1960
1970
1980
1990
2000
2010
2020

Outline

1. Introduction
2. Static Factor Models
3. Dynamic Factor Model
4. Estimation
   - Estimation of Static FMs

Static Factor Models (SFM). I

The SFM expresses a N-vector X_t of observed zero-mean time series variables as depending on a reduced number $r < N$ of latent factors F_t and a mean-zero idiosyncratic component ε_t :

$$X_t \quad = \quad \Lambda F_t + \varepsilon_t, \qquad t = 1, \ldots, T \tag{1}$$

where :

- F_t is an r-vector and ε_t is an N-vector ;
- $\Lambda := (\lambda_1, \lambda_2, \ldots, \lambda_N)'$ is an $N \times r$ matrix ;
- $\mathbf{E}[\varepsilon_t] = \mathbf{E}[F_t] = 0$;
- $\mathbf{E}[F_t F_t'] = I_r$, $\mathbf{E}[F_t \varepsilon_\tau] = 0$, $\forall (t, \tau)$;
- $\mathbf{E}[F_t F_\tau] = 0$, $\forall (t, \tau)$, $t \neq \tau$;
- $\mathbf{E}[\varepsilon_t \varepsilon_\tau] = 0$, $\forall (t, \tau)$, $t \neq \tau$;
- **Exact** FM : $\mathbf{E}[\varepsilon_t \varepsilon_t'] = D = diag(d_1, \ldots, d_N)$

Static Factor Models (SFM). II

In a **static model :** the factors have only a contemporaneous effect on X_t.

The common component is $\chi_t := \Lambda F_t$.

Previous assumptions imply that $Cov(\chi_{it}, \varepsilon_{js}) = 0$, for all t, s and all $i, j \in \{1, \ldots, N\}$.

***

In an **exact** SFM : since the ε_{it} are pairwise uncorrelated, then all the correlation among the observed variables passes through the factors.

In the **Approximate SFM** : we do not suppose that $\mathbf{E}[\varepsilon_{it}\varepsilon_{jt}] = 0$ but that $\mathbf{E}[\varepsilon_{it}\varepsilon_{jt}]$ is negligible with respect to the correlation explained by the common factors.

That is, as $N \to \infty$, $\mathbf{E}[\varepsilon_t \varepsilon_t']$ remains bounded while $\Lambda\Lambda'$ is unbounded.

So, as $Var(X_t) = \Lambda\Lambda' + D$, we can consider that the part of the correlation between two variables that is not explained by the factors is negligible.

Outline

Dynamic Factor Models (DFMs). I

- The DFM represents the evolution of a vector of N observed time series, X_t, in terms of :
  - a reduced number of unobserved common factors which evolve over time,
  - plus uncorrelated disturbances (represent measurement error and/or idiosyncratic dynamics of the individual series).
- There are two ways to write the model : dynamic form vs. static form
- DFM : example of the class of state-space or hidden Markov models, in which observable variables are expressed in terms of unobserved or latent variables, which in turn evolve according to some lagged dynamics with finite dependence (*i.e.*, the law of motion of the latent variables is Markov).
- DFM important for macroeconometric applications because : the complex **comovements** of a potentially large number of observable series are summarized by a small number of common factors, which drive the **common fluctuations** of all the series.

Dynamic Factor Models (DFMs). II

Notation and Conventions :

- observable and latent variables are assumed to be second-order stationary.
- all data series are assumed to be transformed to have unit standard deviation.
- all the series in X_t are integrated of order zero ($X_t \sim I(0)$). So, remove by differencing the data stochastic trends and potential deterministic trends arising through drift if any.

State-space Models. I

- Consider a time series $\{Y_t\}_{t\geq 1}$.
- Specifying the joint finite-dimensional distributions of $(Y_1, \ldots, Y_t)$, for any $t \geq 1$, is not an easy task (independence and exchangeability not often justified in time series).
- Markovian dependence is the simplest form of dependence among the Y_t's. We say that $\{Y_t\}_{t\geq 1}$ is a *Markov chain* if, for any $t > 1$,

$$\pi(Y_t | Y_1, \ldots, Y_{t-1}) = \pi(Y_t | Y_{t-1}).$$

- Assuming a Markovian structure for the observations is, however, not appropriate in many applications.
- **State space models** build on the simple dependence structure of a Markov chain to define more complex models for $\{Y_t\}_t$.

State-space Models. II

- In a state space model we assume that there is an unobservable Markov chain θ_t, called the state process, and that Y_t is an imprecise measurement of θ_t.

- One can think of $\{\theta_t\}$ as an auxiliary time series that facilitates the task of specifying the probability distribution of the observable time series $\{Y_t\}_t$.

- Formally, a state space model consists of an $\mathbb{R}^p$valued time series $\{\theta_t; t = 0, 1, \ldots\}$ and an $\mathbb{R}^m$-valued time series $\{Y_t; t = 0, 1, \ldots\}$, satisfying the following assumptions :

  1. $\{\theta_t\}_t$ is a Markov chain,
  2. Conditionally on $\{\theta_t\}$, the Y_t's are independent and Y_t depends on θ_t only.

  As a consequence, a *state space model* is completely specified by the initial distribution $\pi(\theta_0)$ and the conditional densities $\pi(\theta_t|\theta_{t-1})$ and $\pi(y_t|\theta_t)$, $t \geq 1$. In fact, for any $t > 0$,

$$\pi(\theta_t, \ldots, \theta_0, y_1, \ldots, y_t) = \pi(\theta_0) \prod_{j=1}^{t} \pi(\theta_j|\theta_{j-1})\pi(y_j|\theta_j).$$

Dynamic form of the DFM. I

The DFM expresses a N-vector X_t of observed time series variables as depending on a reduced number q of latent factors f_t and a mean-zero idiosyncratic component ε_t :

$$\begin{aligned} X_t &= \lambda(L)f_t + \varepsilon_t &(2)\\ f_t &= \Psi(L)f_{t-1} + \eta_t, &(3) \end{aligned}$$

where :

- f_t and ε_t are in general *serially correlated* (to capture dynamics) ;
- $\mathbf{E}[f_t] = 0$ and $\mathbf{E}[\varepsilon_t] = 0$;
- $\lambda(L)$ and $\Psi(L)$ are $N \times q$ and $q \times q$ lag polynomial matrices ;
- $\lambda(L)f_t = \lambda_0 f_t + \ldots + \lambda_p f_{t-p}$
- η_t is *serially uncorrelated* and has mean zero ;
- $\mathbf{E}[\varepsilon_t \eta'_{t-k}] = 0, \forall k$.

Dynamic form of the DFM. II

($*$) The i-th row of $\lambda(L)$, $\lambda_i(L)$, is called the dynamic factor loading for the i-th series X_{it}.

($*$) If $p = 0$: the **relation** between X_t and f_t is **static** (but X_t and f_t are dynamic processes).

($*$) The DFM capture the dynamics in the data in two ways :

- factors are autocorrelated and their dynamic is modeled as a VAR (or VARMA),
- and the observed variables can be affected by contemporaneous as well as past values of the factors.

Particular case for the measurement equation :

$$X_t = \lambda_0 f_t + \varepsilon_t,$$

the factors enter the equation only through their contemporaneous value.

($*$) The system of equations (2)-(3) is in the form of a state-space model. (2) is called *measurement equation* while (3) is called *state equation*.

Dynamic form of the DFM. III

$(*)$ The idiosyncratic error ε_t can be serially correlated. A simple and tractable model is the following :

$$\varepsilon_{it} = \delta_i(L)\varepsilon_{i,t-1} + \nu_{it}, \tag{4}$$

where ν_{it} is serially uncorrelated.

Dynamic form of the DFM. IV

1. Exact DFM

When $\mathbf{E}[\varepsilon_{it}\varepsilon_{js}] = 0$, $\forall t, s, i \neq j$.

So, the correlation of one series with another occurs only through the latent factors f_t.

Forecast : under Gaussianity of (ν_t, η_t) and (4),

$$\mathbf{E}[X_{i,t+1}|X_t, f_t, X_{t-1}, f_{t-1}, \ldots] = \alpha_i^f(L)f_t + \delta_i(L)X_{it}, \tag{5}$$

where $\alpha_i^f(L) = \lambda_{i0}\Psi(L) - \delta_i(L)\lambda_i(L) + L^{-1}(\lambda_i(L) - \lambda_0)$.

Estimation procedure :

- the dynamics of f_t and ε_t are specified ;
- the model is cast in a state-space form ;
- the likelihood can be computed through the Kalman filter under a Gaussian assumption, for any given value of the parameters of the model ;
- the MLE of the parameters is obtained through a numerical procedure.

2. Approximate DFM

- The assumption that ε_t is uncorrelated across series is unrealistic in many applications.
- For example, data derived from the same survey might have correlated measurement error and multiple series for a given sector might have unmodeled sector-specific dynamics.
- Chamberlain and Rothschild's (1983) approximate factor model allows for such correlation.
- Under the approximate DFM, (5) would contain additional observable variables relevant for forecasting series X_{it}.
- Used when $N \uparrow \infty$. The components of ε_t might be correlated among them but we assume that the part of the correlation between observables due to the idiosyncratic components is negligible with respect to the part due to the common factors.

Static form of the DFM. I

The dynamic representation in (2)-(3) captures the dependence of the observed variables on the lags of the factors explicitly, while the static representation of the DFM embeds those dynamics implicitly.

The 2 forms lead to different estimation methods.

Two types of Static factor models :

- exact factor models : the factors explain all the correlation among the variables ;
- approximate factor models : adapted to the case where the number of observed variables $\uparrow \infty$, the factors explain most of the correlation among variables (the remaining part of the correlation is negligible).

Static form of the DFM. I

We rewrite the dynamic form above to depend on r static factors F_t instead of the q dynamic factors f_t, where $r > q$.

This rewriting makes the model amenable to principal components analysis (PCA).

- Let p =degree of the lag polynomial matrix $\lambda(L)$;
- let $F_t := (f_t', \ldots, f_{t-p}')'$ de an r-vector of *static* factors ($r = q(p+1)$);
- let $\Lambda := (\lambda_0, \lambda_1, \ldots, \lambda_p)$, where λ_h is a $N \times q$ matrix of coefficients on the h-th lag in $\lambda(L)$;
- let $\Phi(L)$ be the matrix consisting of 0s, 1s, and the elements of $\Psi(L)$ such that the VAR in (3) can be written as

$$\begin{aligned} X_t &= \Lambda F_t + \varepsilon_t && (6) \\ F_t &= \Phi(L) F_{t-1} + G\eta_t, && (7) \end{aligned}$$

where $G = [I_q \; 0_{q\times(r-q)}]'$.

Static form of the DFM. II

- If $\varepsilon_t \sim AR(1)$ as in (4) and if $(\nu_t, \eta_t) \sim \mathcal{N}$, the one step ahead forecast of the i-th variable in the static factor model is

$$\mathbf{E}[X_{i,t+1}|X_t, F_t x_{t-1}, F_{t-1}, \ldots] = \alpha_i^F(L)F_t + \delta_i(L)X_{it}, \tag{8}$$

where $\alpha_i^F(L) = \Lambda_i \Phi(L) - \delta_i(L)\Lambda_i$.

Identification. I

Because the factors are unobserved, they are identified only up to arbitrary normalizations :

- In the static DFM, the space spanned by F_t is identified, but F_t itself is not identified :

$$\Lambda F_t = \Lambda Q^{-1} Q F_t,$$

  where Q is any invertible $r \times r$ matrix.

- Indeed, we will see that the method of PCs estimates the space spanned by F_t instead of F_t :

  if $\widehat{F}_t$ denotes the r-vector of factor estimates, there exists an $r \times r$ invertible matrix Q such that $\widehat{F}_t$ estimates $Q' F_t$.

- For many applications (*e.g.* macro monitoring and forecasting), the object of interest is the conditional mean and so it is necessary only to identify the space spanned by the factors (*i.e.* their linear combinations), not the factors themselves. So, Q is irrelevant.

- For such applications, the lack of identification is resolved by imposing a convenient normalization : the *principal components* normalization.

Principal Components Normalization :
the columns of Λ are orthogonal and are scaled to have unit norm :

$$\frac{1}{N}\Lambda'\Lambda = I_r \qquad \text{and} \qquad F'F \text{ diagonal.}$$

Alternatively,

$$\frac{1}{T}F'F = I_r \qquad \text{and} \qquad \Lambda'\Lambda \text{ diagonal,}$$

where $F := (F_1, \ldots, F_T)$ is the $(T \times r)$ matrix of factors.

While these restrictions identify the space spanned by the columns of F and the space spanned by the columns of Λ, they do not necessarily identify the individual columns of F or of Λ.

IDENTIFICATION OF STATIC FACTORS :

- We are interested in conditions under which we can identity the columns of F and the columns of Λ from the product $F\Lambda'$.
- Since $F\Lambda' = FQQ^{-1}\Lambda$ for any $(r \times r)$ invertible matrix Q, and Q has r^2 free parameters, we need at least r^2 restrictions to identify F and Λ.
- Bai & Ng (2013, JoE) propose three sets of restrictions that lead to exact identification :

Identifying restrictions:		
	Restrictions on F	Restrictions on Λ
(2.1): PC1	$\frac{1}{T}F'F = I_r$	$\Lambda'\Lambda$ is a diagonal matrix with distinct entries
(2.2): PC2	$\frac{1}{T}F'F = I_r$	$\Lambda = \begin{pmatrix} \Lambda_1 \\ \Lambda_2 \end{pmatrix}$, $\Lambda_1 = \begin{pmatrix} \lambda_{11} & 0 & \cdots & 0 \\ \lambda_{21} & \lambda_{22} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ \lambda_{r1} & \lambda_{r2} & \cdots & \lambda_{rr} \end{pmatrix}$, $\lambda_{ii} \neq 0, i = 1, \ldots, r$
(2.3): PC3	Unrestricted	$\Lambda = \begin{pmatrix} I_r \\ \Lambda_2 \end{pmatrix}$

Normalization of the factors. IV

If the true F and true Λ satisfy these restrictions, then the corresponding rotation matrix is asymptotically I_r :

PC1. $H = I_r + \mathcal{O}_p(\delta_{NT}^{-2})$, with $\delta_{NT} := \min\{\sqrt{N}, \sqrt{T}\}$.

PC2. If $r = 1$, $H = I_r + \mathcal{O}_p(\delta_{NT}^{-2})$, with $\delta_{NT} := \min\{\sqrt{N}, \sqrt{T}\}$.
If $r > 1$, $H = I_r + \mathcal{O}_p(T^{-1/2})$.

PC3. H converges in probability to I_r.

Both PC1 and PC2 identify F and Λ up to a column sign change.

Outline

Estimation of the Factors and Parameters. I

Estimation of : the factors, the loadings, the number of factors, other parameters.

- Exact Static factor model :

  $\Rightarrow$ Estimation : MLE.

- Approximate Static factor model :

  $\Rightarrow$ Estimation : Principal Component Analysis (PCA)

  Chamberlain-Rotschild (1983) : large N case + strong assumption. Stock & Watson (2002)

- Exact Dynamic factor model :

  $\Rightarrow$ Estimation : MLE using Kalman filter

  Sargent & Sims (1977), Geweke (1977), Engle & Watson (1980, 1981, 1983), Stock & Watson (1989, 1991, 1993) : small N case.

Estimation of the Factors and Parameters. II

- Approximate Dynamic factor model :

  $\Rightarrow$ Estimation : PCA in the time domain , PCA in the frequency domain, Kalman Filter, QMLE

  Stock & Watson (1989, 1991, 1993), Forni and Reichlin (1996, 1998), Forni, Hallin, Lippi & Reichlin (1999), Bai & Ng : large N case.

Outline

ML estimation of small Exact SFMs. I

We consider an **Exact Static Factor Model** written in the matrix notation :

$$X = F\Lambda' + \varepsilon,$$

where

$$X := \begin{pmatrix} X_1' \\ \vdots \\ X_T' \end{pmatrix}, \qquad F := \begin{pmatrix} F_1' \\ \vdots \\ F_T' \end{pmatrix}, \qquad \varepsilon := \begin{pmatrix} \varepsilon_1' \\ \vdots \\ \varepsilon_T' \end{pmatrix}.$$

- Assume $\varepsilon_t | F \sim i.i.d.\mathcal{N}(0, \Omega)$, Ω p.d. and not related to F.
- The conditional log-likelihood function is :

$$-2\ell(\Lambda, F, \Omega) = TN\log(2\pi) + T\log(|\Omega|) + tr\left\{\Omega^{-1}(X - F\Lambda')'(X - F\Lambda')\right\}. \tag{9}$$

ML estimation of small Exact SFMs. II

- Assume for the moment that Ω is known. The conditional MLE (CMLE) of Λ is

$$\widehat{\Lambda} = X'F(F'F)^{-1}. \tag{10}$$

Replace this in (9) and use the standardization $F'F = TI_r$ to obtain the conditional MLE (CMLE) for F :

$$\widehat{F} = \sqrt{T} \times \Big(\text{matrix of eigenvectors corresponding to the } r \text{ largest eigenvalues of } X\Omega^{-1}X'\Big).$$

Hence, $\widehat{\Lambda} = \frac{1}{T}X'\widehat{F}$.

ML estimation of small Exact SFMs. III

Alternatively :

- The log-likelihood (9) can be maximized with respect to F first. In this case (by using the standardization $\widetilde{\Lambda}'\Omega^{-1}\widetilde{\Lambda} = NI_r$) we get :

$$\widetilde{\Lambda} = \sqrt{N} \times \Big(\text{matrix of eigenvectors corresponding to the } r \text{ largest eigenvalues} \\ \text{of } \Omega^{-1}X'X\Omega^{-1}\Big)$$

$$\widetilde{F} = \frac{1}{N} X\Omega^{-1}\widetilde{\Lambda}.$$

- It holds : $\widehat{F}\widehat{\Lambda}' = \widetilde{F}\widetilde{\Lambda}'$.

ML estimation of small Exact SFMs. IV

- The conditional MLE of Ω is

$$\widehat{\Omega} = \frac{1}{N}(X - \widehat{F}\widehat{\Lambda})'(X - \widehat{F}\widehat{\Lambda}).$$

- In practice, an initial estimator of Ω is required (for instance use PCA).
- Notice that $\widehat{F}$ and $\widehat{\Lambda}$ are not MLEs since the MLEs require knowing the probabilistic structure of $\{F_t\}$ and use iterative computation methods.
- Without the normality assumption, the CMLEs $\widehat{\Lambda}$ and $\widehat{F}$ are called *generalized principal component estimators* (GPCEs) since they take into account the nonspherical structure of the error's var-cov matrix (as do the GLS estimator), see Choi 2012 and Stock & Watson 2006.

Principal Components (PCs) estimation of large Approximate SFMs. I

Nonparametric methods estimate the static factors directly without specifying a model for the factors or assuming specific distributions for the disturbances.

- PCs solve the least-squares problem in which Λ and F_t in (6) are treated as unknown parameters :

$$\min_{F,\Lambda} V_r(\Lambda, F), \quad \text{where } V_r(\Lambda, F) := \frac{1}{NT} \sum_{t=1}^{T} (X_t - \Lambda F_t)'(X_t - \Lambda F_t),$$
$$\text{s.t. } \frac{1}{N}\Lambda'\Lambda = I_r \quad \text{or} \quad \frac{1}{T}F'F = I_r. \qquad (11)$$

- The data have to be centered and standardized.

Principal Components (PCs) estimation of large Approximate SFMs. II

- If we concentrate out Λ and use the normalization that $\frac{1}{T}F'F = I_r$, the optimization problem is identical to

$$\max tr(F'(XX')F).$$

The estimated factor matrix is

$$\check{F} = \sqrt{T} \times \text{eigenvectors corresponding to the } r \text{ largest eigenvalues of } \mathbf{XX'}.$$

Given $\check{F}$, $\check{\Lambda}' = (\check{F}'\check{F})^{-1}\check{F}'X = \check{F}'X/T$.

- Another solution to the least-squares problem (11) is given by

$$\widehat{\Lambda} = \sqrt{N} \times \text{eigenvectors corresponding to the } r \text{ largest eigenvalues of } \mathbf{X'X}.$$

Given $\widehat{\Lambda}$, $\widehat{F} = X\widehat{\Lambda}/N$.

- The first solution is less computationally costly when $N > T$, while the second one when $T > N$.

Principal component analysis. I

We now look at the justification of the previous procedure.

- N fixed and finite.
- Assume to have a panel of i.i.d. data, *i.e.* with no serial correlation : $Cov(X_{it}, X_{j,t-k}) = 0$ for any k and any $i, j = 1, \ldots, n$.
- Assume $\mathbf{E}[X_t] = 0$ and define $\Sigma_X = \mathbf{E}[X_t X_t']$.

The covariance matrix Σ_X is a symmetric and positive definite matrix and can always be factorised as

$$\Sigma_X = PDP',$$

where $P = (p_1, \ldots, p_N)$ is an $N \times N$ orthogonal matrix of eigenvectors and $D = diag(\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_N > 0)$ contains the ordered eigenvalues. So $PP' = P'P = I_N$,

Then,

$$\Sigma_X p_j = \lambda_j p_j$$

Principal component analysis. II

and

$$\Sigma_X = \sum_{j=1}^{N} \lambda_j p_j p_j'.$$

Recall that the eigenvalues solve : $|\Sigma_X - \lambda_j I_N| = 0$, then once we have found the eigenvalues we can find the eigenvectors by solving :

$$(\Sigma_X - \lambda_j I_N) p_j = 0.$$

Equivalently, eigenvalues and eigenvectors are the solutions of the problem :

$$\begin{aligned} p_1 &= \arg\max_a a'\Sigma_X a \qquad \text{such that } a'a = 1 \\ p_j &= \arg\max_a a'\Sigma_X a \qquad \text{such that } a'a = 1 \quad \text{and } a'p_1 = \ldots = a'p_{j-1} = 0,\ j > 1. \end{aligned}$$

We can also write :

$$p_1 = \arg\max_a Var\Big(\sum_{i=1}^{N} a_i X_{it}\Big) \qquad \text{such that } \sum_{i=1}^{N} a_i^2 = 1.$$

Principal component analysis. III

- $p_j' X_t$ is called j-th principal component of X_t and has variance λ_j.
- The projection of X_t onto the j-th principal component is the vector

  $$y_j := (p_j' X_t) p_j$$

  which has length λ_j and direction p_j.

Assume that we want to reduce the dimension of the data from N to $r < N$, consider the decomposition dictated by principal components :

$$X_t = \sum_{j=1}^{N} (p_j' X_t) p_j = \underbrace{\sum_{j=1}^{r} (p_j' X_t) p_j}_{X_{t,[r]}} + \underbrace{\sum_{j=r+1}^{N} (p_j' X_t) p_j}_{X_t - X_{t,[r]}}. \tag{12}$$

This dimension reduction technique is called **Principal Component Analysis (PCA)**. PCA is usually used to estimate static factor models.

Principal component analysis. IV

Denote $\xi_t := X_t - X_{t,[r]}$ with covariance Σ_ξ. Then, consider any other projection $\widetilde{X}_{t,[r]}$ with residuals $\widetilde{\xi}_t := \widetilde{X}_t - \widetilde{X}_{t,[r]}$ and covariance matrix $\Sigma_{\widetilde{\xi}}$. We can show that $X_{t,[r]}$ is the best r-dimensional representation of X_t in the sense that

$$tr\Sigma_\xi \leq tr\Sigma_{\widetilde{\xi}}.$$

So, this decomposition minimises the sum of the residual variances (which is equivalent to minimising the sum of all eigenvalues of the residual covariance Σ_ξ.

The covariance matrix of $X_{t,[r]}$ has just r eigenvalues different from zero and therefore has rank r, while the covariance matrix of ξ_t has rank $N - r$.

Principal component analysis. V

Now assume the static exact factor model for x_t :

$$X_t = \Lambda F_t + \varepsilon_t,$$

where F_t is a r-vector of static factors, Λ is a $N \times r$ matrix of loadings, and ε_t is an idiosyncratic error component as opposed to the common component ΛF_t.

- Assume orthogonality between the common and idiosyncratic components : $\mathbf{E}[F_t \varepsilon_t'] = 0$,
- Assume exact factor structure : $\mathbf{E}[\varepsilon_t \varepsilon_t'] = \Gamma_\varepsilon$ to be a positive definite and diagonal matrix.
- Without loss of generality, we can impose $\mathbf{E}[F_t F_t'] = I_r$. Therefore,

$$\Sigma_X = \Lambda \Lambda' + \Gamma_\varepsilon$$

where $\Lambda\Lambda'$ has rank r and Γ_ε has rank N.