Dynamic Factor Models

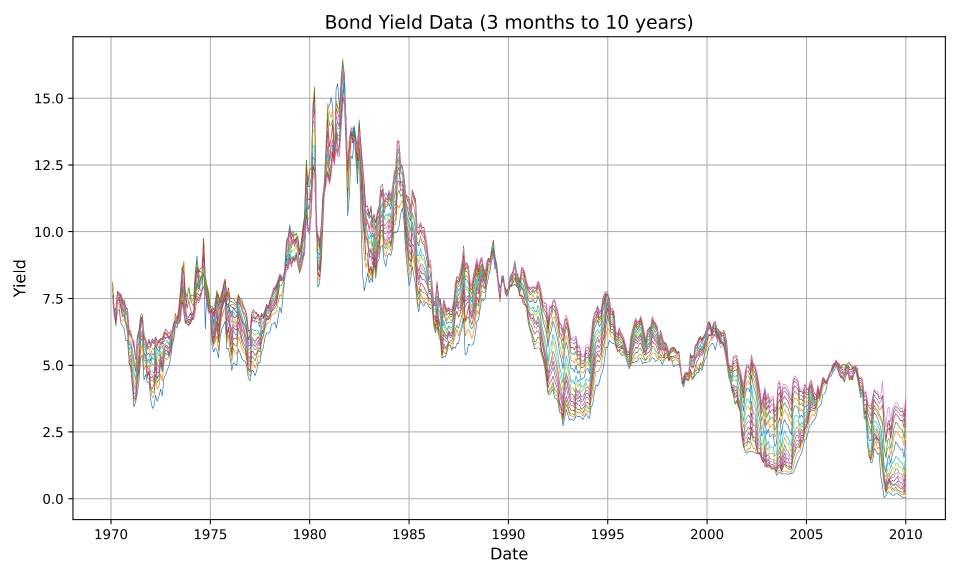
Constantino Hevia
UTDT

July 18, 2024

Factor Models

- In macroeconomics we have a large N, small T problem:
  - Large N: we have hundreds of data series (financial, macroeconomic, etc)
  - Small T: limited number of years with reliable data
  - Degrees of freedom problem: VAR with 10 variables and 2 lags has $10 + 2 \times 10^2 + 10 \times 11/2 = 365$ parameters to estimate!
- **Dynamic Factor Models:** presumption that a few common factors drive most comovements across the series.
- **This lecture**:
  - Principal components
  - Dynamic factor models (Stock and Watson, 2011)
  - FAVARs (Bernanke, Boivin and Eliasz, 2005)

Data on US treasury yields



Principal components analysis

- We observe a large number N of variables at time $t = 1, 2, ..., T$.
- **Objective:** summarize the data using few indicators.
- $\boldsymbol{y}_t = N$-vector of stationary observations at time t.
- $\hat{\mu}_i = T^{-1} \sum_{t=1}^{T} y_{i,t}$: sample mean of variable $i = 1, \ldots, N$.
- $\hat{\sigma}_{ii} = T^{-1} \sum_{t=1}^{T} (y_{i,t} - \mu_i)^2$: sample variance of variable $i = 1, \ldots, N$.
- $\tilde{y}_{i,t} = (y_{i,t} - \hat{\mu}_i) / \sqrt{\hat{\sigma}_{ii}}$: standardized value of variable $i = 1, \ldots, N$.
- $\tilde{\boldsymbol{y}}_t = (\tilde{y}_{1,t}, \tilde{y}_{2,t}, ..., \tilde{y}_{N,t})'$: N-vector of standardized variables.
- $\hat{\boldsymbol{\Omega}} = T^{-1} \sum_{t=1}^{T} \tilde{\boldsymbol{y}}_t \tilde{\boldsymbol{y}}_t'$: sample correlation matrix of $\boldsymbol{y}_t$.

First principal component: dimensionality reduction to a single factor

- Observed time series of ("large" N) data $\tilde{\boldsymbol{y}}_t \in \Re^N$.
- Approximate $\tilde{\boldsymbol{y}}_t$ using a single time-varying variable f_t:

$$\tilde{\boldsymbol{y}}_t \approx \boldsymbol{h} f_t.$$

- **Objective:** explain as much variation as possible of the vector $\tilde{\boldsymbol{y}}_t$ using the scalar f_t.
- f_t is called the first principal component of $\tilde{\boldsymbol{y}}_t$.
- The vector $\boldsymbol{h}$ contains the factor loadings.
- f_t and $\boldsymbol{h}$ determined up to an arbitrary scale factor.

First principal component: the problem

- Find a scalar f_t and a vector $\boldsymbol{h} \in \Re^N$ to minimize distance between $\tilde{\boldsymbol{y}}_t$ and $\boldsymbol{h} f_t$:

$$\min_{\boldsymbol{h}, \{f_1, f_2, \ldots, f_T\}} \sum_{t=1}^{T} (\tilde{\boldsymbol{y}}_t - \boldsymbol{h} f_t)'(\tilde{\boldsymbol{y}}_t - \boldsymbol{h} f_t)$$

- **Note**: f_t and $\boldsymbol{h}$ are not unique:
    - $\boldsymbol{h} f_t = \boldsymbol{h} q q^{-1} f_t = \boldsymbol{h}^* f_t^*$ for $q \neq 0$ ($\boldsymbol{h}^* = \boldsymbol{h} q$ and $f_t^* = q^{-1} f_t$).
    - However, the product $\boldsymbol{h} f_t$ is unique (strictly convex minimization problem).
    - Usual normalization to eliminate the indeterminacy: impose $\boldsymbol{h}' \boldsymbol{h} = 1$.

First principal component: the problem

$$\min_{\boldsymbol{h},\{f_1,f_2,\ldots,f_T\}} \sum_{t=1}^{T} (\tilde{\boldsymbol{y}}_t - \boldsymbol{h} f_t)'(\tilde{\boldsymbol{y}}_t - \boldsymbol{h} f_t)$$

subject to

$$\boldsymbol{h}'\boldsymbol{h} = 1.$$

- Solve the problem in two steps:
  1. Given $\boldsymbol{h}$, find the best $\{f_1, f_2, ..., f_T\}$. This gives $f_t(\boldsymbol{h})$.
  2. Substitute $f_t(\boldsymbol{h})$ into the objective function and minimize with respect to $\boldsymbol{h}$.

Step 1: minimize over f_t given $\boldsymbol{h}$

$$\min_{\{f_1, f_2, \ldots, f_T\}} \sum_{t=1}^{T} (\tilde{\boldsymbol{y}}_t - \boldsymbol{h} f_t)'(\tilde{\boldsymbol{y}}_t - \boldsymbol{h} f_t)$$

- Minimize period by period:

$$\min_{f_t} (\tilde{\boldsymbol{y}}_t - \boldsymbol{h} f_t)'(\tilde{\boldsymbol{y}}_t - \boldsymbol{h} f_t).$$

- OLS: $\tilde{\boldsymbol{y}}_t$ are the dependent variables, $\boldsymbol{h}$ is the vector of regressors, and f_t is the β coefficient:

$$f_t(\boldsymbol{h}) = (\boldsymbol{h}'\boldsymbol{h})^{-1} \boldsymbol{h}' \tilde{\boldsymbol{y}}_t.$$

- Optimized objective function

$$\begin{aligned}
(\tilde{\boldsymbol{y}}_t - \boldsymbol{h} f_t(\boldsymbol{h}))'(\tilde{\boldsymbol{y}}_t - \boldsymbol{h} f_t(\boldsymbol{h})) =& (\tilde{\boldsymbol{y}}_t - \boldsymbol{h}(\boldsymbol{h}'\boldsymbol{h})^{-1}\boldsymbol{h}'\tilde{\boldsymbol{y}}_t)'(\tilde{\boldsymbol{y}}_t - \boldsymbol{h}(\boldsymbol{h}'\boldsymbol{h})^{-1}\boldsymbol{h}'\tilde{\boldsymbol{y}}_t) \\
=& \tilde{\boldsymbol{y}}_t'\tilde{\boldsymbol{y}}_t - \tilde{\boldsymbol{y}}_t'\boldsymbol{h}(\boldsymbol{h}'\boldsymbol{h})^{-1}\boldsymbol{h}'\tilde{\boldsymbol{y}}_t \\
=& \tilde{\boldsymbol{y}}_t' \left(I_N - \boldsymbol{h}(\boldsymbol{h}'\boldsymbol{h})^{-1}\boldsymbol{h}' \right) \tilde{\boldsymbol{y}}_t.
\end{aligned}$$

Step 2: minimize over $\boldsymbol{h}$

$$\min_{\boldsymbol{h}} \sum_{t=1}^{T} (\tilde{\boldsymbol{y}}_t - \boldsymbol{h} f_t(\boldsymbol{h}))'(\tilde{\boldsymbol{y}}_t - \boldsymbol{h} f_t(\boldsymbol{h})) = \min_{\boldsymbol{h}} \sum_{t=1}^{T} \tilde{\boldsymbol{y}}_t' \left(I_N - \boldsymbol{h}(\boldsymbol{h}'\boldsymbol{h})^{-1}\boldsymbol{h}' \right) \tilde{\boldsymbol{y}}_t$$

$$\text{subject to } \boldsymbol{h}'\boldsymbol{h} = 1.$$

- Equivalent to

$$\max_{\boldsymbol{h}} \sum_{t=1}^{T} \tilde{\boldsymbol{y}}_t' \boldsymbol{h}(\boldsymbol{h}'\boldsymbol{h})^{-1}\boldsymbol{h}'\tilde{\boldsymbol{y}}_t \quad \text{subject to} \quad \boldsymbol{h}'\boldsymbol{h} = 1.$$

- Using $\boldsymbol{h}'\boldsymbol{h} = 1$ and $\tilde{\boldsymbol{y}}_t'\boldsymbol{h} = \boldsymbol{h}'\tilde{\boldsymbol{y}}_t$, objective function can be written as

$$\begin{aligned}\sum_{t=1}^{T} \tilde{\boldsymbol{y}}_t' \boldsymbol{h}(\boldsymbol{h}'\boldsymbol{h})^{-1}\boldsymbol{h}'\tilde{\boldsymbol{y}}_t &= \sum_{t=1}^{T} \boldsymbol{h}'\tilde{\boldsymbol{y}}_t\tilde{\boldsymbol{y}}_t'\boldsymbol{h} = \boldsymbol{h}' \left(\sum_{t=1}^{T} \tilde{\boldsymbol{y}}_t\tilde{\boldsymbol{y}}_t' \right) \boldsymbol{h} \\ &= T\boldsymbol{h}'\hat{\Omega}\boldsymbol{h}.\end{aligned}$$

Step 2: minimize over $\boldsymbol{h}$

- Problem reduces to

$$\max_{\boldsymbol{h}} \boldsymbol{h}'\hat{\Omega}\boldsymbol{h} \quad \text{subject to} \quad \boldsymbol{h}'\boldsymbol{h} = 1.$$

- Consider the eigenvalues λ_i and eigenvectors $\boldsymbol{x}_i$ of $\hat{\Omega}$:

$$\hat{\Omega}\boldsymbol{x}_i = \boldsymbol{x}_i\hat{\lambda}_i \text{ for } i = 1, 2, ..., N$$

$$\begin{aligned} \hat{\Lambda} =& diag(\hat{\lambda}_1, \hat{\lambda}_2, ..., \hat{\lambda}_N) \\ X =& [\boldsymbol{x}_1, \boldsymbol{x}_2, ..., \boldsymbol{x}_N] \end{aligned}$$

where $X'X = XX' = I_N$ and $\hat{\lambda}_1 > \hat{\lambda}_2 > ... > \hat{\lambda}_N > 0$.

- Then

$$\begin{aligned} \hat{\Omega}X =& X\hat{\Lambda} \\ X'\hat{\Omega}X =& \hat{\Lambda} \end{aligned}$$

Step 2: minimize over $\boldsymbol{h}$

$$\max_{\boldsymbol{h}} \boldsymbol{h}'\hat{\Omega}\boldsymbol{h} \quad \text{subject to} \quad \boldsymbol{h}'\boldsymbol{h} = 1.$$

- Change of variable: $\tilde{\boldsymbol{h}} = X'\boldsymbol{h}$ (so that $\boldsymbol{h} = X\tilde{\boldsymbol{h}}$), where $XX' = I_N$ and $X'\hat{\Omega}X = \hat{\Lambda}$. Note also that $\tilde{\boldsymbol{h}}'\tilde{\boldsymbol{h}} = \boldsymbol{h}'XX'\boldsymbol{h} = \boldsymbol{h}'\boldsymbol{h} = 1$.
- Then

$$\boldsymbol{h}'\hat{\Omega}\boldsymbol{h} = \tilde{\boldsymbol{h}}'X'\hat{\Omega}X\tilde{\boldsymbol{h}} = \tilde{\boldsymbol{h}}'\hat{\Lambda}\tilde{\boldsymbol{h}}.$$

- Thus,

$$\max_{\boldsymbol{h}} \boldsymbol{h}'\hat{\Omega}\boldsymbol{h} \quad \text{s.t.} \quad \boldsymbol{h}'\boldsymbol{h} = 1 \iff \max_{\tilde{\boldsymbol{h}}} \tilde{\boldsymbol{h}}'\hat{\Lambda}\tilde{\boldsymbol{h}}' \quad \text{s.t.} \quad \tilde{\boldsymbol{h}}'\tilde{\boldsymbol{h}} = 1.$$

- But diagonal $\hat{\Lambda}$ implies

$$\tilde{\boldsymbol{h}}'\hat{\Lambda}\tilde{\boldsymbol{h}} = \tilde{h}_1^2\hat{\lambda}_1 + \tilde{h}_2^2\hat{\lambda}_2 + \ldots + \tilde{h}_N^2\hat{\lambda}_N.$$

Step 2: minimize over $\boldsymbol{h}$

- The problem is

$$\max_{\{\tilde{h}_1,\tilde{h}_2,\ldots,\tilde{h}_N\}} \tilde{h}_1^2\hat{\lambda}_1 + \tilde{h}_2^2\hat{\lambda}_2 + \ldots + \tilde{h}_N^2\hat{\lambda}_N \quad \text{subject to} \quad \tilde{h}_1^2 + \tilde{h}_2^2 + \ldots + \tilde{h}_N^2 = 1.$$

- Solution is (recall $\hat{\lambda}_1 > \hat{\lambda}_2 > \ldots > \hat{\lambda}_N > 0$):

$$\tilde{h}_1 = 1, \quad \tilde{h}_2 = \tilde{h}_3 = \tilde{h}_N = 0 \implies \boldsymbol{h} = X\tilde{\boldsymbol{h}} = \boldsymbol{x}_1.$$

- **Conclusion:**
  - Vector of factor loadings $\boldsymbol{h}$ is given by the eigenvector of $\hat{\Omega}$ associated with the largest eigenvalue ($\boldsymbol{x}_1$ and λ_1, respectively).
  - First principal component is the inner product of $\boldsymbol{x}_1$ with the data $\tilde{\boldsymbol{y}}_t$,

$$f_t(\boldsymbol{h}) = (\boldsymbol{h}'\boldsymbol{h})^{-1}\boldsymbol{h}'\tilde{\boldsymbol{y}}_t = \boldsymbol{x}_1'\tilde{\boldsymbol{y}}_t.$$

Two principal components

- Use 2 variables to summarize the N variables in $\tilde{\boldsymbol{y}}_t$.

- Choose a 2×1 vector $\boldsymbol{f}_t = (f_{1t}, f_{2t})'$ and an $N \times 2$ matrix $\boldsymbol{H} = [\boldsymbol{h}_1, \boldsymbol{h}_2]$ to solve

$$\min_{\{\boldsymbol{H}, \boldsymbol{f}_1, \boldsymbol{f}_2, \ldots, \boldsymbol{f}_T\}} \sum_{t=1}^{T} (\tilde{\boldsymbol{y}}_t - \boldsymbol{H}\boldsymbol{f}_t)'(\tilde{\boldsymbol{y}}_t - \boldsymbol{H}\boldsymbol{f}_t)$$

- The vector $\boldsymbol{f}_t$ are the (two) factors and the matrix $\boldsymbol{H}$ are the factor loadings.

- $\boldsymbol{f}_t$ and $\boldsymbol{H}$ not unique:
  - Let $\boldsymbol{Q}$ be a nonsingular 2×2 matrix.
  - $\boldsymbol{H}\boldsymbol{f}_t = \boldsymbol{H}\boldsymbol{Q}\boldsymbol{Q}^{-1}\boldsymbol{f}_t$, $\boldsymbol{H}^* = \boldsymbol{H}\boldsymbol{Q}$, $f_t^* = \boldsymbol{Q}^{-\boldsymbol{1}}\boldsymbol{f}_t \implies \boldsymbol{H}\boldsymbol{f}_t = \boldsymbol{H}^*\boldsymbol{f}_t^*$.
  - Normalize $\boldsymbol{H}'\boldsymbol{H} = I_2$.

Two principal components

- Solve the problem in two steps:
  1. Given $\boldsymbol{H}$, find $\boldsymbol{f}_t$ that minimizes the objective function $\implies \boldsymbol{f}_t(\boldsymbol{H})$
  2. Substitute $\boldsymbol{f}_t(\boldsymbol{H})$ into the objective function and minimize with respect to $\boldsymbol{H}$.
- Step 1 is OLS period-by-period

$$\begin{aligned}\min_{\{\boldsymbol{f}_t\}}(\tilde{\boldsymbol{y}}_t - \boldsymbol{H}\boldsymbol{f}_t)'(\tilde{\boldsymbol{y}}_t - \boldsymbol{H}\boldsymbol{f}_t) =& \tilde{\boldsymbol{y}}_t'(I_N - \boldsymbol{H}(\boldsymbol{H}'\boldsymbol{H})^{-1}\boldsymbol{H}')\tilde{\boldsymbol{y}}_t \\ \boldsymbol{f}_t =& (\boldsymbol{H}'\boldsymbol{H})^{-1}\boldsymbol{H}'\tilde{\boldsymbol{y}}_t.\end{aligned}$$

- Step 2:

$$\max_{\boldsymbol{H}} \sum_{t=1}^{T} = \tilde{\boldsymbol{y}}_t'\boldsymbol{H}(\boldsymbol{H}'\boldsymbol{H})^{-1}\boldsymbol{H}'\tilde{\boldsymbol{y}}_t \quad \text{subject to } \boldsymbol{H}'\boldsymbol{H} = I_2.$$

Two principal components

- Write objective function as (using $\boldsymbol{H}'\boldsymbol{H} = I_2$)

$$
\begin{aligned}
\sum_{t=1}^{T} \tilde{\boldsymbol{y}}_t' \boldsymbol{H}(\boldsymbol{H}'\boldsymbol{H})^{-1}\boldsymbol{H}'\tilde{\boldsymbol{y}}_t &= \sum_{t=1}^{T} \underbrace{\tilde{\boldsymbol{y}}_t'\boldsymbol{H}\boldsymbol{H}'\tilde{\boldsymbol{y}}_t}_{\text{scalar}} \\
&= \sum_{t=1}^{T} trace(\tilde{\boldsymbol{y}}_t'\boldsymbol{H}\boldsymbol{H}'\tilde{\boldsymbol{y}}_t) \\
&= \sum_{t=1}^{T} trace(\boldsymbol{H}'\tilde{\boldsymbol{y}}_t\tilde{\boldsymbol{y}}_t'\boldsymbol{H}) \\
&= trace(\boldsymbol{H}'\sum_{t=1}^{T}\tilde{\boldsymbol{y}}_t\tilde{\boldsymbol{y}}_t'\boldsymbol{H}) \\
&= T \times trace\left(\boldsymbol{H}'\hat{\Omega}\boldsymbol{H}\right).
\end{aligned}
$$

Two principal components

- Change of variable: $\tilde{\boldsymbol{H}} = X'\boldsymbol{H}$ so that $\boldsymbol{H} = X\tilde{\boldsymbol{H}} = X\left[\tilde{\boldsymbol{h}}_1, \tilde{\boldsymbol{h}}_2\right]$ and $\tilde{\boldsymbol{H}}'\tilde{\boldsymbol{H}} = I$. Then,

$$
\begin{aligned}
trace\left(\boldsymbol{H}'\hat{\Omega}\boldsymbol{H}\right) =& trace\left(\tilde{\boldsymbol{H}}'X'\hat{\Omega}X\tilde{\boldsymbol{H}}\right) \\
=& trace\left(\tilde{\boldsymbol{H}}'\hat{\Lambda}\tilde{\boldsymbol{H}}\right) \\
=& trace\left(\begin{bmatrix} \tilde{\boldsymbol{h}}_1' \\ \tilde{\boldsymbol{h}}_2' \end{bmatrix} \hat{\Lambda}[\tilde{\boldsymbol{h}}_1, \tilde{\boldsymbol{h}}_2]\right) \\
=& trace\left(\begin{bmatrix} \tilde{\boldsymbol{h}}_1'\hat{\Lambda}\tilde{\boldsymbol{h}}_1 & \tilde{\boldsymbol{h}}_1'\hat{\Lambda}\tilde{\boldsymbol{h}}_2 \\ \tilde{\boldsymbol{h}}_2'\hat{\Lambda}\tilde{\boldsymbol{h}}_1 & \tilde{\boldsymbol{h}}_2'\hat{\Lambda}\tilde{\boldsymbol{h}}_2 \end{bmatrix}\right) \\
=& \tilde{\boldsymbol{h}}_1'\hat{\Lambda}\tilde{\boldsymbol{h}}_1 + \tilde{\boldsymbol{h}}_2'\hat{\Lambda}\tilde{\boldsymbol{h}}_2 \\
=& \tilde{h}_{1,1}^2\hat{\lambda}_1 + \tilde{h}_{2,1}^2\hat{\lambda}_2 + \ldots + \tilde{h}_{N,1}^2\hat{\lambda}_N \\
& + \tilde{h}_{1,2}^2\hat{\lambda}_1 + \tilde{h}_{2,2}^2\hat{\lambda}_2 + \ldots + \tilde{h}_{N,2}^2\hat{\lambda}_N.
\end{aligned}
$$

Two principal components

- The problem is

$$\max_{\tilde{\boldsymbol{h}}_1,\tilde{\boldsymbol{h}}_2} \left(\tilde{h}_{1,1}^2\hat{\lambda}_1 + \tilde{h}_{2,1}^2\hat{\lambda}_2 + \ldots + \tilde{h}_{N,1}^2\hat{\lambda}_N\right) + \left(\tilde{h}_{1,2}^2\hat{\lambda}_1 + \tilde{h}_{2,2}^2\hat{\lambda}_2 + \ldots + \tilde{h}_{N,2}^2\hat{\lambda}_N\right)$$

subject to

$$\begin{aligned} \tilde{h}_{1,1}^2 + \tilde{h}_{2,1}^2 + \ldots + \tilde{h}_{N,1}^2 &= 1 \\ \tilde{h}_{1,2}^2 + \tilde{h}_{2,2}^2 + \ldots + \tilde{h}_{N,2}^2 &= 1 \\ \tilde{h}_{1,1}\tilde{h}_{1,2} + \tilde{h}_{2,1}\tilde{h}_{2,2} + \ldots + \tilde{h}_{N,1}\tilde{h}_{N,2} &= 0. \end{aligned}$$

- Recalling $\hat{\lambda}_1 > \hat{\lambda}_2 > \ldots > \hat{\lambda}_N > 0$, the solution is

$$\begin{aligned} \tilde{h}_{1,1} &= 1, \quad \tilde{h}_{2,1} = \tilde{h}_{3,1} = \ldots = \tilde{h}_{N,1} = 0 \quad \Longrightarrow \quad \boldsymbol{h}_1 = \boldsymbol{x}_1 \\ \tilde{h}_{2,2} &= 1, \quad \tilde{h}_{1,2} = \tilde{h}_{3,2} = \ldots = \tilde{h}_{N,2} = 0 \quad \Longrightarrow \quad \boldsymbol{h}_2 = \boldsymbol{x}_2. \end{aligned}$$

Two principal components

- **Conclusion:**
  - Factor loadings $\boldsymbol{h}_1$ and $\boldsymbol{h}_2$ are given by the eigenvectors of $\hat{\Omega}$ associated with the two largest eigenvalues, $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$.
  - The first two principal components f_{1t} and f_{2t} are given by the product of the eigenvectors $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$ with the data $\tilde{\boldsymbol{y}}_t$:

$$\boldsymbol{f}_t = (\boldsymbol{H}'\boldsymbol{H})^{-1}\boldsymbol{H}'\tilde{\boldsymbol{y}}_t = \boldsymbol{H}'\tilde{\boldsymbol{y}}_t = [\boldsymbol{x}_1, \, \boldsymbol{x}_2]'\tilde{\boldsymbol{y}}_t = \begin{bmatrix} \boldsymbol{x}_1'\tilde{\boldsymbol{y}}_t \\ \boldsymbol{x}_2'\tilde{\boldsymbol{y}}_t \end{bmatrix} = \begin{bmatrix} f_{1t} \\ f_{2t} \end{bmatrix}.$$

  - Second principal component is orthogonal to the first:

$$\sum_{t=1}^{T} f_{1t} f_{2t} = \sum_{t=1}^{T} \left(\boldsymbol{x}_1'\tilde{\boldsymbol{y}}_t\right)\left(\tilde{\boldsymbol{y}}_t'\boldsymbol{x}_2\right) = T\boldsymbol{x}_1'\hat{\Omega}\boldsymbol{x}_2 = 0. \tag{1}$$

In general, $S < N$ principal components

- Approximate
$$\tilde{\boldsymbol{y}}_t \approx \boldsymbol{H}\boldsymbol{f}_t$$
where $\boldsymbol{f}_t = (f_{1t}, f_{2t}, \ldots, f_{St})'$ are the S principal components, and $\boldsymbol{H} = [\boldsymbol{h}_1, \boldsymbol{h}_2, \ldots, \boldsymbol{h}_S]$ is an $N \times S$ matrix with the factor loadings.

- Factor loading $\boldsymbol{h}_1$, $\boldsymbol{h}_2$, and $\boldsymbol{h}_S$ are given by the eigenvectors of $\hat{\Omega}$ associated with the S largest eigenvalues.

- Principal component $s = 1, 2, \ldots, S$ are given by
$$f_{s,t} = \boldsymbol{h}_s' \tilde{\boldsymbol{y}}_t = \boldsymbol{x}_s' \tilde{\boldsymbol{y}}_t.$$

- Principal components are orthogonal: $\sum_{t=1}^{T} f_{i,t} f_{j,t} = 0$ for $i \neq j$.

Dynamic factor models

Three generations of factor models (Stock and Watson, 2011)

- **First generation** (parametric): write the model in state-space form and use the Kalman filter to estimate the parameters of the model and the Kalman smoother to estimate the factors: Engle and Watson (1981,1983), Stock and Watson (1989), Sargent (1989), Quah and Sargent (1993)

- **Second generation (this note)** (non-parametric): Factor estimation using cross-averaging across Y_t. Implementation using *Principal Components*. Stock and Watson (2002), Bai (2003), Bai and Ng (2006).

- **Third generation**: Hybrid principal component and state space methods: Giannone, Reichlin, and Small (2008); Doz, Giannone, and Reichlin (2006).
  - Step 1: estimate factors using principal components.
  - Step 2: estimate parameters of the state-space representation using the Kalman filter.

State space models

Factor models are a special case of linear state-space models:

$$\underset{(r\times 1)}{F_t} = \underset{(r\times r)}{C} F_{t-1} + u_t$$
$$\underset{(N\times 1)}{Y_t} = \underset{(N\times r)}{A} F_t + v_t$$

where $N >> r$.

- F_t are the “state variables” or “factors”.
- Y_t are the “observed variables”.
- u_t are shocks to the state variables.
- v_t are “measurement” errors:
  - In the factor model: everything that is not explained by the factors F_t.

What are the factors?

- In some contexts, the factors may be associated with some state variables in a DSGE model.
- In most cases, however, the factors don't have any economic interpretation.
- But people have been labeling factors:
    - Level, slope and curvature (bond yields).
    - Real and nominal factors (Ng and Ludvigson (2009)).

When does it work in theory?

Consider the system

$$\underset{(r\times 1)}{F_t} = \underset{(r\times r)}{C} F_{t-1} + u_t$$
$$\underset{(N\times 1)}{Y_t} = \underset{(N\times r)}{A} F_t + v_t.$$

We need:

1. $N^{-1}A'A \to D_A$, where D_A is a $r \times r$ matrix of rank r as $N \to \infty$.
   - This condition means that factors are pervasive: that they affect all observables and that the observables span the factors.

2. $\max\left(eig(E(v_t v_t'))\right) \leq c < \infty$ for all N.
   - This condition requires that the idiosyncratic disturbances have limited correlation across series. This, in turn, implies that the errors in the individual series cancel out as the number of series is increased.

Idea of the method: non-parametric cross-sectional averaging

- Idea is to estimate the factors by cross-sectional averaging of Y_t at each time t

$$Y_t = AF_t + v_t$$

- The average could be a weighted average (this is what principal components do).
- If we take the cross sectional average across Y_{it}, idiosyncratic disturbances will converge to zero by the weak law of large number as N increases.

Intuition: cross-averaging

- Construct the estimator of F_t as the weighted average of Y_t using a $N \times r$ matrix of weights W, where we normalize $W'W/N = I_r$,

$$\hat{F}_t = N^{-1} W' Y_t$$

- If $N^{-1} W' A \to H$ as $N \to \infty$, where H is a full rank $r \times r$ matrix and, if Conditions 1 and 2 are satisfied, $\hat{F}_t$ is consistent for the space spanned by F_t:

$$\hat{F}_t = N^{-1} W' (AF_t + v_t) = N^{-1} W' AF_t + N^{-1} W' v_t \xrightarrow{p} HF_t \text{ as } N \to \infty$$

- The key to this result is to choose W such that $N^{-1} W' A \to H$ where H is full rank. Principal components deliver this property.

Estimation with principal components

- The principal components estimator of F_t is a weighted averaging estimator with
  - $W = \hat{X}$.
  - $\hat{X}$ is the matrix of eigenvectors of the sample covariance matrix of Y_t, $\hat{\Sigma}_Y$, associated with its r largest eigenvalues.
  - Once we have an estimate of F_t, we can estimate the dynamics of the factors estimating the VAR

$$F_t = CF_{t-1} + u_t.$$

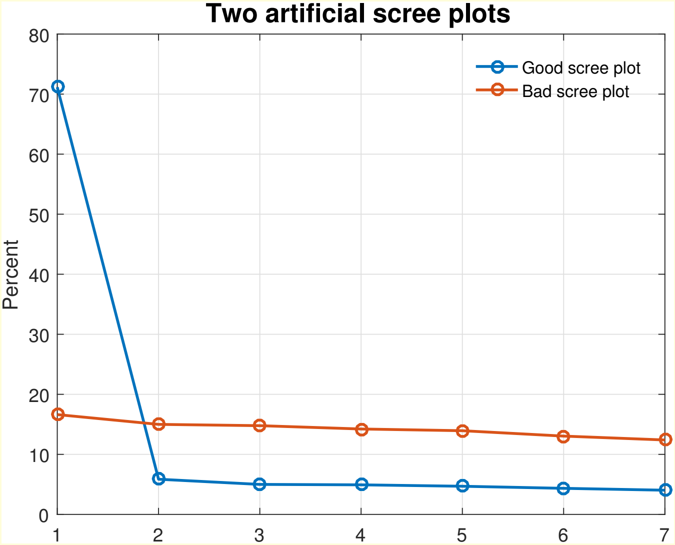
When does it work in practice?

- Just from observing the data Y_t, can we tell whether a large dataset can be represented with a factor structure?
- Answer: Yes
  - Scree plots (informal but useful).

Scree plot

- Use that $E[v_t v_t'] << E[Y_t Y_t']$ implies a particular structure of $E[Y_t Y_t']$ if $N >> r$.
- Do eigenvalue-eigenvector decomposition (i.e. principal components) of the covariance matrix $E(Y_t Y_t')$
$$E(Y_t Y_t') = X \Lambda X'$$
    - X contains the eigenvectors of $E(Y_t Y_t')$.
    - Λ is a diagonal matrix containing the ordered eigenvalues.
    - The eigenvectors are orthogonal so that $XX' = I$.
- Plot ordered eigenvalues from largest to smallest.

Scree plot



Two artificial scree plots
80
70
60
50
40
30
20
10
0
Percent
1
2
3
4
5
6
7
Good scree plot
Bad scree plot

Choosing the number of factors

Several approaches:

- Inspect the scree plot: informal but useful.
- Information criterion such as Bai and Ng, Econometrica (2002).
- Largest ratio of adjacent eigenvalues: Ahn and Horenstein, Econometrica (2013)

Bai and Ng (2002) information criterion

- Fix number of factors r and let the optimized objective be

$$V(r) = \min_{\{\boldsymbol{H}^{(r)}, \boldsymbol{f}_t^{(r)}\}} \frac{1}{NT} \sum_{t=1}^{T} (\tilde{\boldsymbol{y}}_t - \boldsymbol{H}^{(r)} \boldsymbol{f}_t^{(r)})' (\tilde{\boldsymbol{y}}_t - \boldsymbol{H}^{(r)} \boldsymbol{f}_t^{(r)})$$

  subject to

$$\boldsymbol{H}^{(r)\prime} \boldsymbol{H}^{(r)} = \boldsymbol{I}_r.$$

- Choose r to minimize

$$\log V(r) + r \frac{(N+T) \log(\min(N, T))}{NT}$$

Ahn and Horenstein (2013) criterion

- Identify the largest "kink" in the scree plot. Let

$$\hat{\Omega} = T^{-1} \sum_{t=1}^{T} \tilde{y}_t \tilde{y}_t'$$

- Order eigenvalues of $\hat{\Omega}$ from largest to smallest,

$$\hat{\lambda}_1 > \hat{\lambda}_2 > ... > \hat{\lambda}_N > 0.$$

- Choose r to be the value for which the ratio

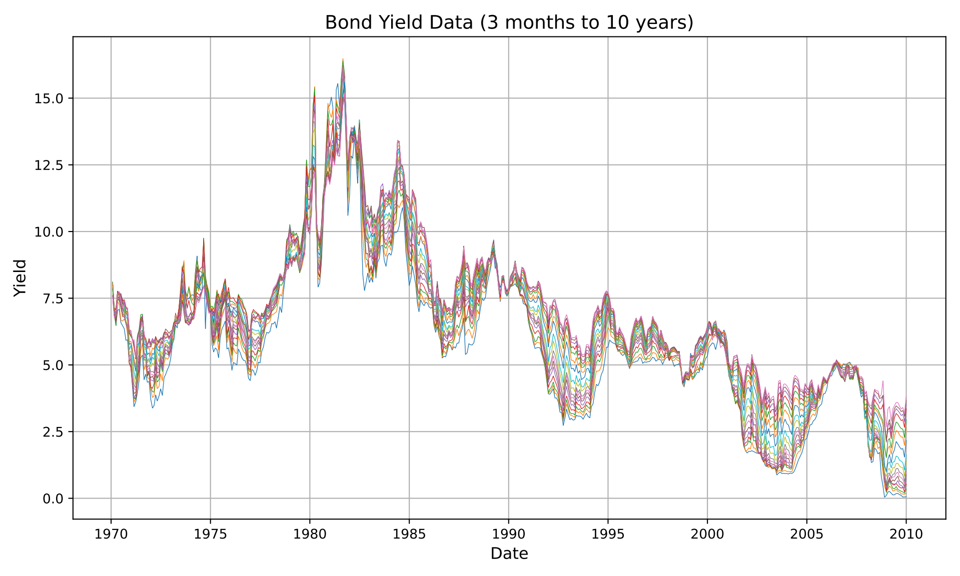
$$\frac{\hat{\lambda}_r}{\hat{\lambda}_{r+1}}$$

is largest.

Example: U.S. bond yields

- U.S. treasury yields from zero coupon bonds of 17 different maturities.
- Maturities range from 3 months to 10 years.
- Monthly data, 1970-2010.

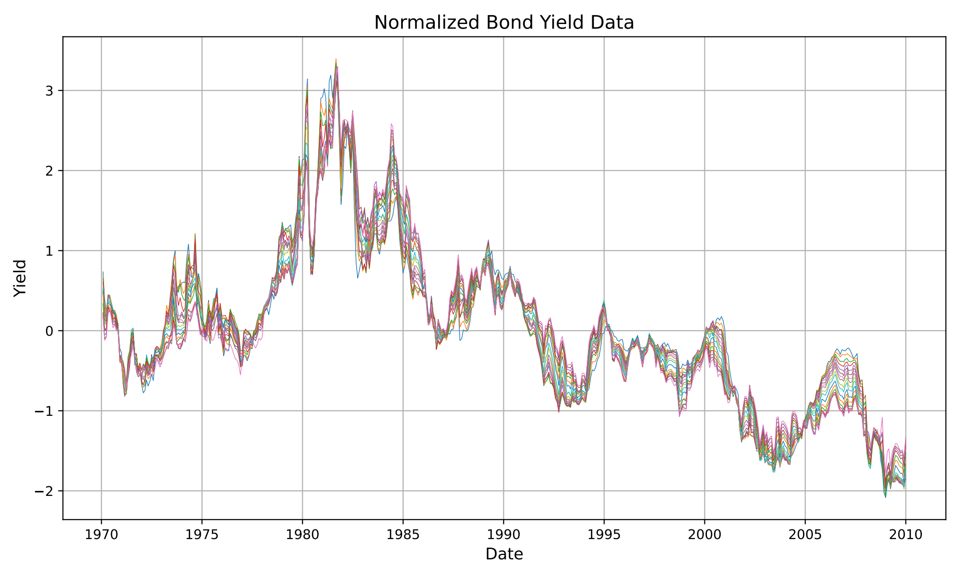
US treasury bond yields



Step 1: normalize the data

- Subtract the mean from each time series.
- Divide by the standard deviation.
- Ensures that factor structure does not depend on the unit of measurements.
    - For example, GDP measured in dollars or pesos.

Normalized data



Step 2: obtain the principal components

- Perform Eigenvalue-Eigenvector decomposition of the covariance matrix

$$T^{-1}\sum_{t=1}^{T}\tilde{\boldsymbol{y}}_t\tilde{\boldsymbol{y}}_t' = \Omega = X\Lambda X' \implies X'\Omega X = \Lambda$$

- In Matlab,

$$[X, \Lambda] = eig(\Omega)$$

  where Λ is a diagonal matrix with the eigenvalues *in ascending or descending order* depending on the properties of Ω.

  - This can be confusing. There is a better command in Matlab: pca($\cdot$).

- in Python: PCA($\cdot$) of the sklearn.decomposition library.

Step 2: in Matlab

- Use Matlab's pca (principal component analysis) command:

$$[X, F, \lambda] = pca(\text{normalized data})$$

  - X contains the N eigenvectors (factor loadings).
  - F contains the estimated factors $(T \times N)$ across columns.
  - λ is a $N \times 1$ vector with the N eigenvalues *in descending order*.

- Also, if we use the N principal components, we recover the data:

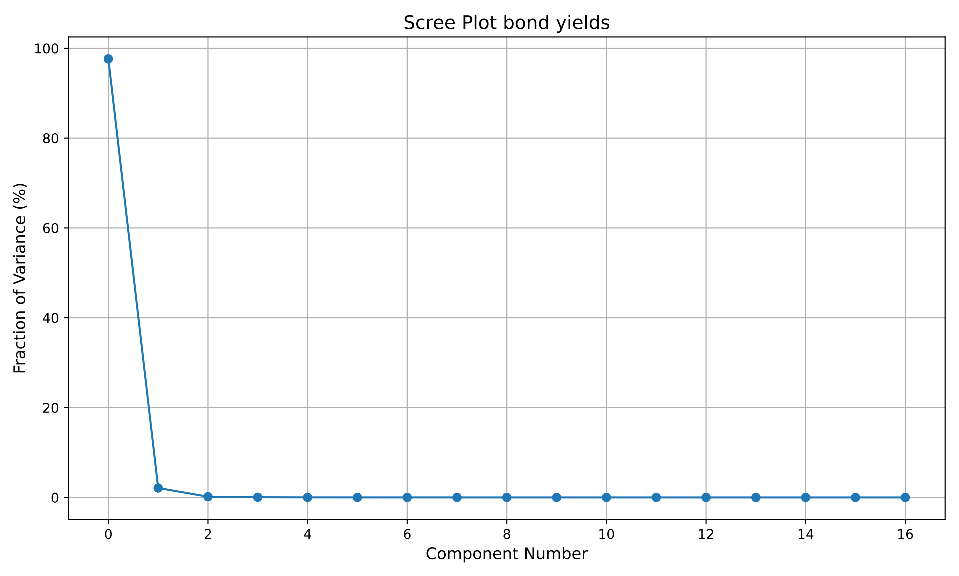
$$\tilde{\boldsymbol{y}}_t = X\boldsymbol{f}_t$$

where X is $N \times N$ and $\boldsymbol{f}_t = [f_{1t}, f_{2t}, ..., f_{Nt}]$.

Step 2: in Python

```python
# Call relevant libraries
from sklearn.decomposition import PCA
from sklearn.preprocessing import StandardScaler
...
# Standardize data for analysis
scaler = StandardScaler()
data_norm = scaler.fit_transform(data)
...
# Compute principal components of normalized data,
# factors and eigenvalues
pca = PCA()
factors = pca.fit_transform(data_norm)
eigenvalues = pca.explained_variance_
```

Scree plot for US treasury yields data

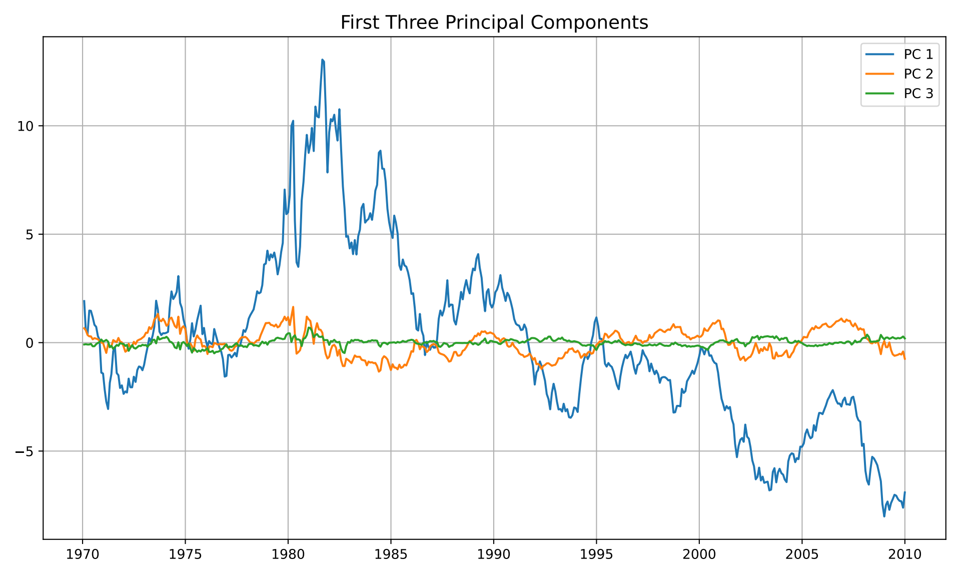


Criteria to choose number of factors

# of factors	Bai-Ng	Ahn-Horenstein
1	**0.20**	**46.40**
2	0.35	12.26
3	0.52	3.91
4	0.69	3.10
5	0.86	1.92
6	1.04	1.24
7	1.21	1.50
8	1.38	1.27
9	1.55	1.24
10	1.73	1.30

- However, must be careful with the “large” persistence in interest rates.
- Usual tests suggest interest rates have a unit root.
- Of course, this can’t be true in theory but may still screw things up in “small” samples. In finance, people typically use 3 factors.

Principal components



Fit of the three factors

- If the three factor model is a good representation the data, the three factors should fit the data well.

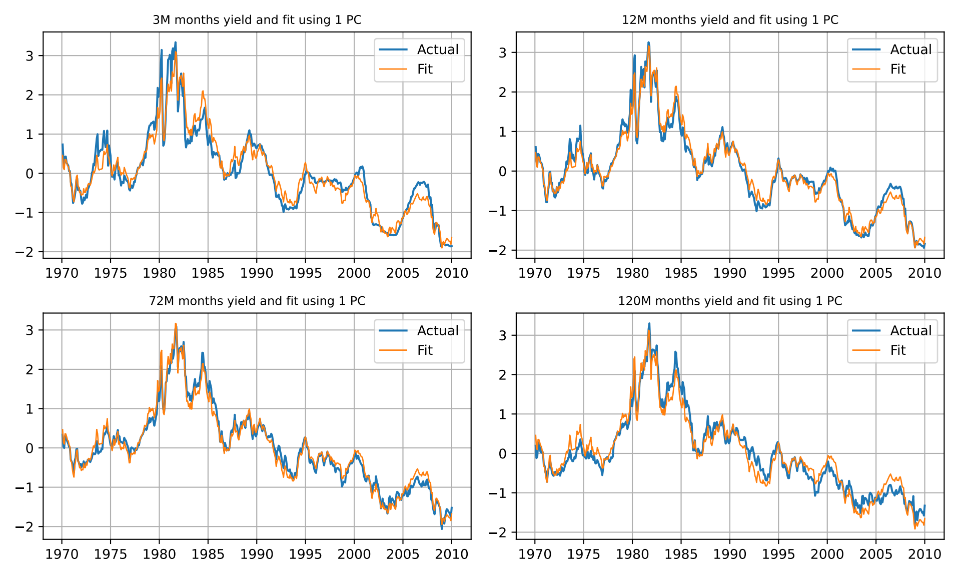
- Recall that we constructed the factors that solve

$$
\begin{aligned}
\tilde{\boldsymbol{y}}_t &= X\boldsymbol{f}_t \\
&= \begin{bmatrix} \boldsymbol{x}_1, & \cdots & , \boldsymbol{x}_N \end{bmatrix} \begin{bmatrix} f_{1,t} \\ \vdots \\ f_{N,t} \end{bmatrix}
\end{aligned}
$$

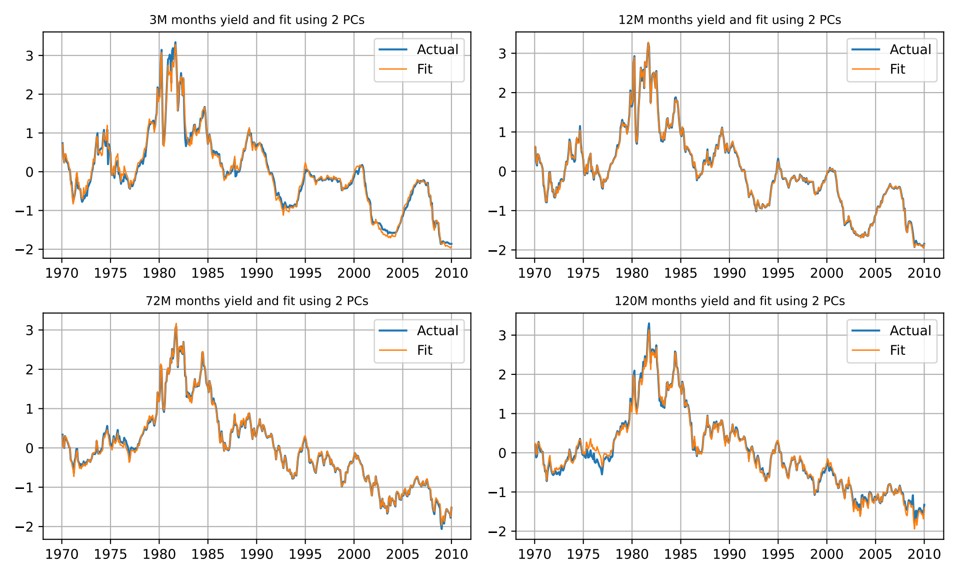
- We now plot the fitted values of yields with maturity 3-month, 12 months, 6 years, and 10 years using the first three principal components, in order

$$
\begin{aligned}
\tilde{\boldsymbol{y}}_t^{fit,1} &= \boldsymbol{x}_1 f_{1,t} \\
\tilde{\boldsymbol{y}}_t^{fit,2} &= \boldsymbol{x}_1 f_{1,t} + \boldsymbol{x}_2 f_{2,t} \\
\tilde{\boldsymbol{y}}_t^{fit,3} &= \boldsymbol{x}_1 f_{1,t} + \boldsymbol{x}_2 f_{2,t} + \boldsymbol{x}_3 f_{3,t}.
\end{aligned}
$$

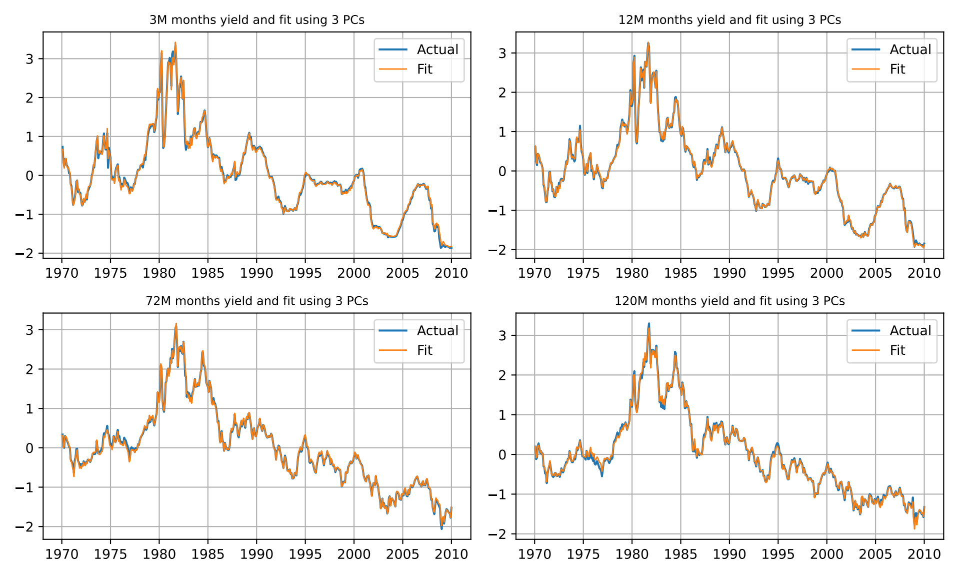
Fit using one principal component



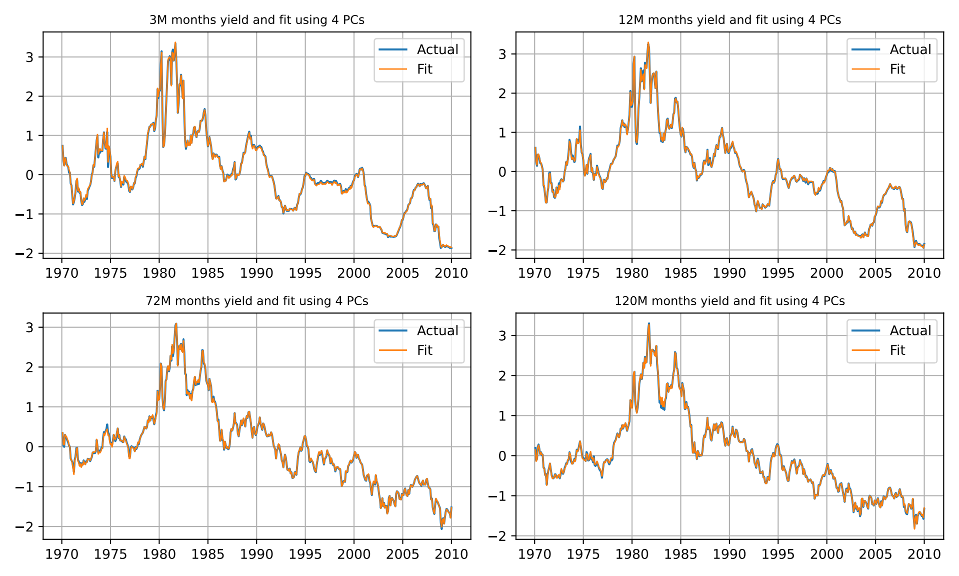
Fit using two principal components



Fit using three principal components



Fit using four principal components



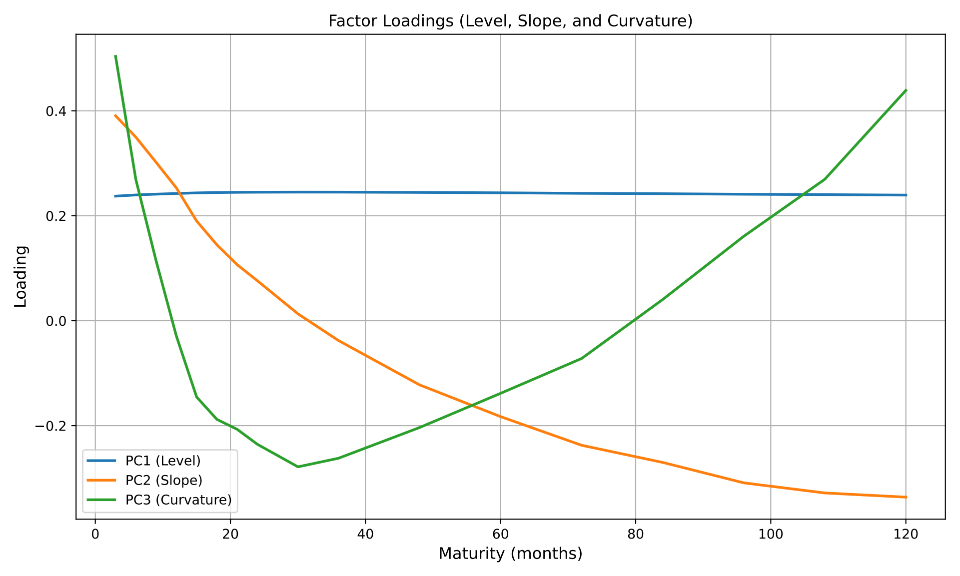
Interpretation of the factors

- In principle, they don't have a structural intepretation. The factors are statistical constructs.

- Yet, in the case of the yield curve they have been given names that are suggestive for their effect on the shape of the yield curve.

- Let's plot the *factor loadings* $[\boldsymbol{x}_1 \; \boldsymbol{x}_2 \; \boldsymbol{x}_3]$ as a function of the 17 maturities (first three columns of X).

Factor loadings



Step 3: estimate the evolution of the factors F_t

- Estimate the dynamic evolution of the factors as a VAR.
- For example, in the VAR(1) case,

$$
\begin{aligned}
F_{3,t} &= CF_{3,t-1} + u_t \\
F_{3,t} &\equiv [f_{1,t}\ f_{2,t}\ f_{3,t}]'
\end{aligned}
$$

- Run the VAR using the estimated factors to obtain an estimate of
  - C: 9 parameters.
  - $E(u_t u_t')$: 6 parameters.
  - Total of 15 parameters.

Dimensionality reduction of the procedure

Our estimated model for the dynamics of $\hat{Y}_t$ is now

$$\begin{aligned} \hat{Y}_t &= [\boldsymbol{x}_1 \; \boldsymbol{x}_2 \; \boldsymbol{x}_3] \, F_{3,t} \\ F_{3,t} &= C F_{3,t-1} + u_t \end{aligned}$$

- We estimated $15 + 17 \times 3 = 66$ parameters $(C$, $E(u_t u_t')$, and $[\boldsymbol{x}_1 \; \boldsymbol{x}_2 \; \boldsymbol{x}_3])$
- Compare this number with the $17^2 + 17 \times 18/2 = 442$ parameters of a 17 variable VAR(1) for bond yields.
- We estimate only 15% of the large VAR parameters (66/442).

What can we do with a factor model?

- Forecasting.
- FAVARs (factor augmented VARs).
- Factors as instruments in IV regressions or GMM (not today).
- Nowcasting (not today).

Forecasting with factor models

- Consider the factor model

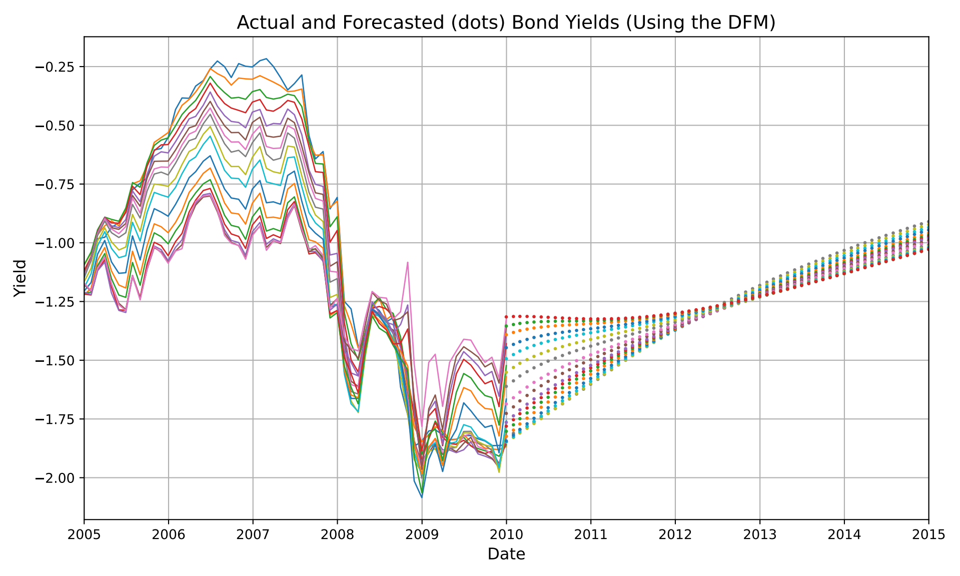
$$
\begin{aligned}
F_t &= CF_{t-1} + u_t \\
Y_t &= WF_t + v_t
\end{aligned}
$$

- Can produce s period-ahead forecasts for all N variables by using

$$E_t[Y_{t+s}|F_t] = WC^s F_t.$$

- Advantages:
  - It often improves out-of-sample fit because it reduces the overfitting problem.
  - No degrees of freedom problem.
- If factors follow a VAR(p), it is convenient to write the factor model in companion form.

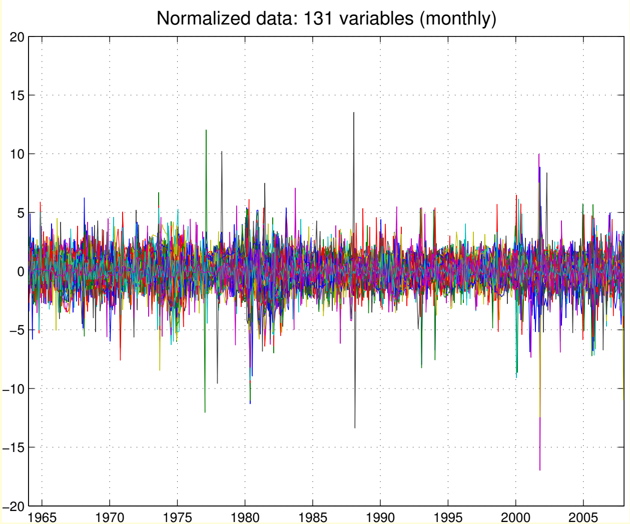
Forecasting interest rates



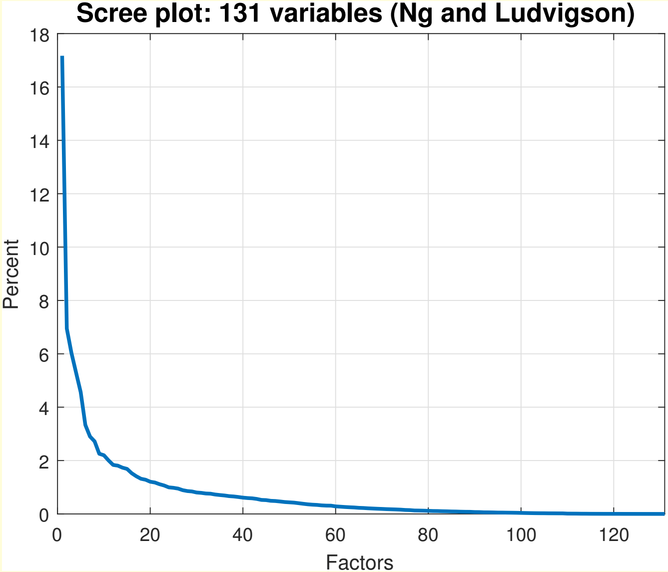
But sometimes the factor structure is not so obvious

- In bond prices the factor structure is obvious: yields move more or less together.
- But this is not always the case...
- Ludvigson and Ng (2009). Macro Factors in Bond Risk Premia.
- Investigate linkage between variation in excess bond returns and macroeconomic fundamentals.
- 131 monthly macro and financial variables series, 1964:1 - 2007:12.
- How does normalized data look like?

Ludvigson and Ng data

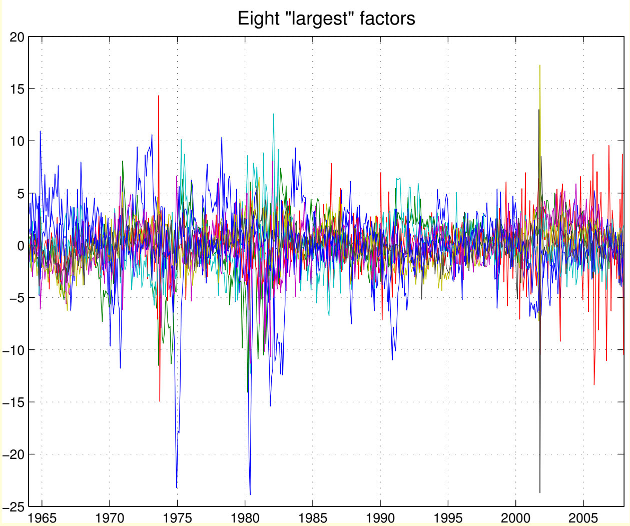


Scree plot



Scree plot: 131 variables (Ng and Ludvigson)
Percent
Factors
0
2
4
6
8
10
12
14
16
18
0
20
40
60
80
100
120

Factors



Eight "largest" factors
20
15
10
5
0
−5
−10
−15
−20
−25
1965
1970
1975
1980
1985
1990
1995
2000
2005

FAVARs: Identifying structural shocks using large data sets

Bernanke, Boivin and Eliasz (QJE, 2005): Measuring the effects of monetary policy: A Factor Augmented Vector Autoregressive (FAVAR) approach.

- Investigate the effect of monetary policy shocks.
- Using much more information than what is feasible in SVAR.
  - Large N/T generates degrees of freedom problem.
- Avoid the price puzzle using standard identifying assumption (short run).
- Impulse responses can be computed for many variables.

FAVAR: Implementation as in Bernanke, Boivin and Eliasz

Consider the model

$$\begin{bmatrix} F_t \\ r_t \end{bmatrix} = \Phi \begin{bmatrix} F_{t-1} \\ r_{t-1} \end{bmatrix} + v_t$$

- F_t are factors constructed from some macroeconomic variables Y_t ($N \times 1$)
    - The basic assumption is that the factors F_t help describe the dynamic of the variables of interest Y_t
- r_t is the nominal short interest rate (Federal Fund rate).

Strategy:

1. Find the factors using PCA.
2. Estimate reduced form FAVAR.
3. Identify the effects of a monetary policy shock using short run restrictions.

Step 1: Find the factors

- We want the factors to capture information that is orthogonal to the interest rate r_t.
- Start regressing the N variables Y_t on r_t

$$Y_t = \beta r_t + u_t$$

  Note that β and u_t are both $N \times 1$ vectors.
- Construct the factors F_t as the principal components of the residuals u_t.

Step 2: Estimate the reduced form FAVAR

The reduced form FAVAR is obtained by estimating the following VAR

$$\begin{bmatrix} F_t \\ r_t \end{bmatrix} = \Phi \begin{bmatrix} F_{t-1} \\ r_{t-1} \end{bmatrix} + v_t$$

by OLS.

Step 3: Identify a shock to the interest rate r_t

- Interest rate r_t is ordered last.

- Economic content:
  - Policy makers observe all the variables and factors when choosing the interest rate.

  - The factors do not respond simultaneously to changes in the interest rate.

- Identification using Cholesky

$$\hat{\Omega} = E v_t v_t'$$
$$CC' = \hat{\Omega}$$

- Structural FAVAR:

$$A_0 \begin{bmatrix} F_t \\ r_t \end{bmatrix} = A_1 \begin{bmatrix} F_{t-1} \\ r_{t-1} \end{bmatrix} + \epsilon_t;\ \epsilon_t \sim N(0, I)$$

where $A_0 = C^{-1}$ and $A_1 = C^{-1}\hat{\Phi}$.

Step 3: Identify a shock to the interest rate r_t

- We now discuss how to compute the impulse response function in this model.
- Recall that
$$Y_{t+s} = WF_{t+s} + \beta r_{t+s}$$
and that factors evolve as
$$\begin{bmatrix} F_t \\ r_t \end{bmatrix} = \Phi \begin{bmatrix} F_{t-1} \\ r_{t-1} \end{bmatrix} + C\epsilon_t$$
- Then
$$Y_t = \begin{bmatrix} W & \beta \end{bmatrix} \begin{bmatrix} F_t \\ r_t \end{bmatrix} = \begin{bmatrix} W & \beta \end{bmatrix} \Phi \begin{bmatrix} F_{t-1} \\ r_{t-1} \end{bmatrix} + \begin{bmatrix} W & \beta \end{bmatrix} C\epsilon_t .$$

Step 3: Identify a shock to the interest rate r_t

- For Y_{t+1}

$$\begin{aligned} Y_{t+1} &= [W \quad \beta] \begin{bmatrix} F_{t+1} \\ r_{t+1} \end{bmatrix} \\ &= [W \quad \beta]\Phi \begin{bmatrix} F_t \\ r_t \end{bmatrix} + [W \quad \beta] C\epsilon_{t+1} \\ &= [W \quad \beta]\Phi^2 \begin{bmatrix} F_{t-1} \\ r_{t-1} \end{bmatrix} + [W \quad \beta]\Phi C\epsilon_t + [W \quad \beta] C\epsilon_{t+1}. \end{aligned}$$

- For Y_{t+2} :

$$\begin{aligned} Y_{t+2} &= [W \quad \beta] \begin{bmatrix} F_{t+2} \\ r_{t+2} \end{bmatrix} \\ &= [W \quad \beta]\Phi^3 \begin{bmatrix} F_{t-1} \\ r_{t-1} \end{bmatrix} + [W \quad \beta]\Phi^2 C\epsilon_t + [W \quad \beta]\Phi C\epsilon_{t+1} + [W \quad \beta] C\epsilon_{t+2}. \end{aligned}$$

Step 3: Identify a shock to the interest rate r_t

- In general,

$$Y_{t+s} = [W\,\beta]\Phi^{s+1}\begin{bmatrix} F_{t-1} \\ r_{t-1} \end{bmatrix} + [W\,\beta]\Phi^{s}C\epsilon_t + [W\,\beta]\Phi^{s-1}C\epsilon_{t+1} + \ldots + [W\,\beta]C\epsilon_{t+s}$$

- The impulse response for the general vector of shocks at t is thus

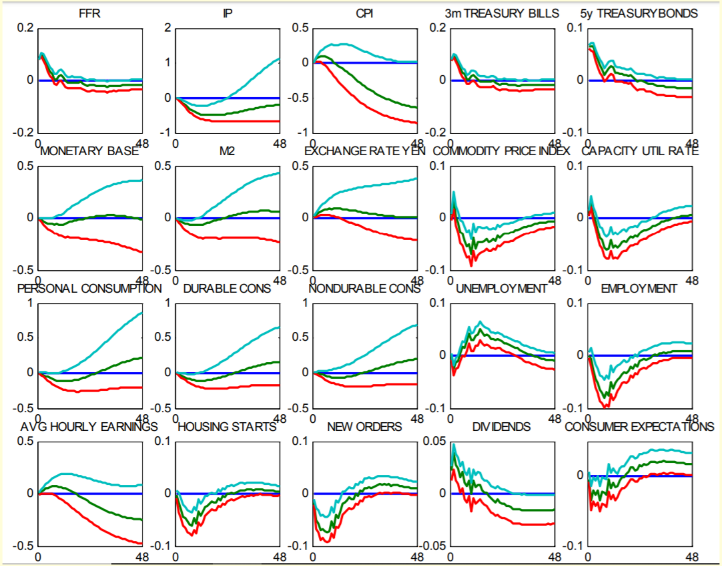
$$E_t(Y_{t+s}) - E_{t-1}(Y_{t+s}) = [W\,\beta]\Phi^{s}C\epsilon_t.$$

- Therefore, to compute the impulse response function to the identified monetary policy shock, use

$$IR(s, \epsilon_t^r) = [W\,\beta]\Phi^{s}C\iota_r,$$

  where ι_r is a vector of zeros with a 1 in the last element (corresponding to the interest rate shock).

Impulse responses



References

- Bai, J. (2003), "Inferential Theory for Factor Models of Large Dimensions," Econometrica, 71, 135-172.
- Bai, J., and S. Ng (2006a), "Confidence Intervals for Diffusion Index Forecasts and Inference for Factor-Augmented Regressions," Econometrica, 74,1133-1150.
- Bernanke, Boivin and Eliasz (2005) Measuring the effects of monetary policy: A Factor Augmented Vector Autoregressive (FAVAR) approach. *Quarterly Journal of Economics*
- Engle, R.F., and M.W. Watson (1981), "A One-Factor Multivariate Time Series Model of Metropolitan Wage Rates," Journal of the American Statistical Association, 76, 774-781.
- Engle, R.F., and M.W. Watson (1983), "Alternative Algorithms for Estimation of Dynamic MIMIC, Factor, and Time Varying Coefficient Regression Models," Journal of Econometrics, Vol. 23, pp. 385-400.
- Stock, J.H., and M.W. Watson (1989), "New Indexes of Coincident and Leading Economic Indicators," NBER Macroeconomics Annual 1989, 351-393.
- Stock, J.H., and M.W. Watson (2002), "Forecasting Using Principal Components from a Large Number of Predictors," Journal of the American Statistical Association, 97, 1167-1179.
- Stock, James and Mark Watson (2011). Dynamic Factor Models.
- Sargent, T.J. (1989), "Two Models of Measurements and the Investment Accelerator," Journal of Political Economy 97:251–287.
- Quah, D., and T.J. Sargent (1993), "A Dynamic Index Model for Large Cross Sections", in Business Cycles, Indicators, and Forecasting, ed. by J.H. Stock and M.W. Watson, Chicago: University of Chicago Press, 285-310.
- Sargent and Sims (1977). Business cycle modeling without pretending to have too much a-priori economic theory.