Factor models

March 13, 2017

Factor Models

Macro economists have a peculiar data situation:

- Many data series, but usually short samples

How can we utilize all this information without running into degrees of freedom problems?

- Factor models is one approach.
- We will focus on what Stock and Watson (2010) call "second generation non-parametric" factor models

Today: Principal components and FAVARs

- Based on Stock and Watson (2010) and Bernanke, Boivin and Eliasz (2005).

Factor models

Dynamic factor model split data into a low dimensional dynamic component and a transitory series specific (idiosyncratic) component

Underlying assumption:

- A few common *factors* can explain most of the variation in many different time series

Factor models is one approach to so-called "dimension reduction" that sometimes help in estimation (and forecasting)

State space models

Factor models are a special case of state space models

$$\underset{(r\times 1)}{F_t} = AF_{t-1} + u_t$$

$$\underset{(N\times 1)}{Y_t} = \underset{(N\times r)}{W}\underset{(r\times 1)}{F_t} + v_t$$

where $N >> r$

Factor models

What are the factors?

- In most cases, there is no economic interpretation: The factors are statistical constructs that have no deeper meaning beyond their definitions
- In some context the factors may correspond to the some combination of state variables in a DSGE model (but one needs to be careful)

This has not stopped people from labeling factors:

- Real and nominal factors (Ng and Ludvigson JoF 2009)
- Level, slope and curvature (Large literature on bond yields)

When does it work in theory?

For the system

$$\underset{(r\times 1)}{F_t} = AF_{t-1} + u_t$$
$$\underset{(N\times 1)}{Y_t} = WF_t + v_t$$

we need that

1. $N^{-1}W'W \to D_W$ where D_W is an $r \times r$ matrix of rank r as $N \to \infty$
2. max $eig(E(v_t v_t')) \leq c < \infty$

Condition (1) ensures that factors affect all observables, and that the observables span the factors. Condition (2) ensures that the measurement errors in individual series cancel as the number of series is increased.

When does it work in practice?

Can we from just observing Y_t tell whether a large data set can be represented with a factor structure?

Yes:

- Scree plots (informal but useful)

Scree plots

Uses 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 of (normalized) sample covariance $E(Y_t Y_t')$

$$E\left(Y_t Y_t'\right) = W \Lambda W'$$

where W contains the eigenvectors of $E(Y_t Y_t')$ and Λ is a diagonal matrix containing the ordered eigenvalues and where the eigenvectors are orthonormal so that $WW' = I$.

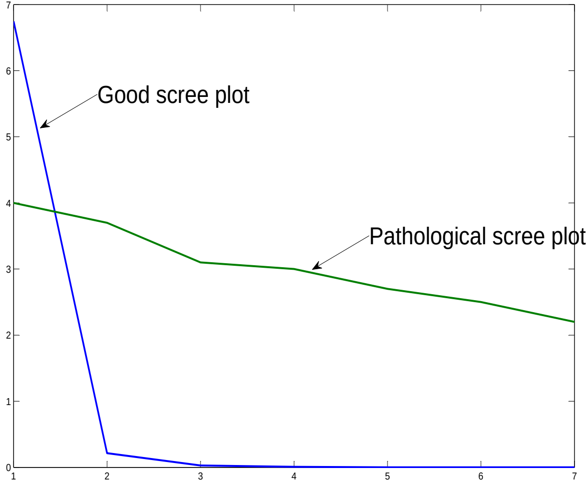


Figure: Good and bad scree plot

Choosing the number of factors

There are several approaches:

- Visual inspection of the scree plot
  - The most common and usually sufficiently accurate method
- Formal test of largest ratio of two adjacent eigenvalues
  - Identifies the sharpest "kink" in the scree plot.
- Information criteria
  - Similar to likelihood ratio tests: Cost of better fit with more factors traded against risk of over-parameterization

Visual inspection of scree plots will do it for us

An example: Bond yields

The data:

Use Fed Funds Rate, 3, 12, 24, 36, 48, 60 month bond yields

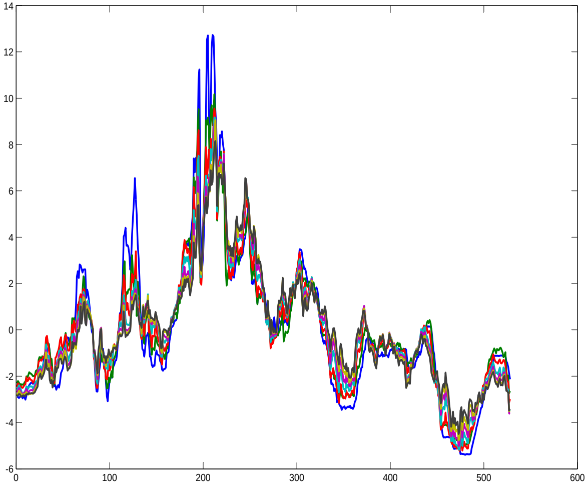


Figure: Bond yield data

Step 1: Normalize the time series

Normalize variances

$$\widehat{y}_{n,t} = \frac{y_{n,t}}{\sqrt{E y_{n,t} y'_{n,t}}}$$

This is done to ensure that factors structure is not sensitive to unit of measurement

- Example: GDP measured in Euros and Pesetas

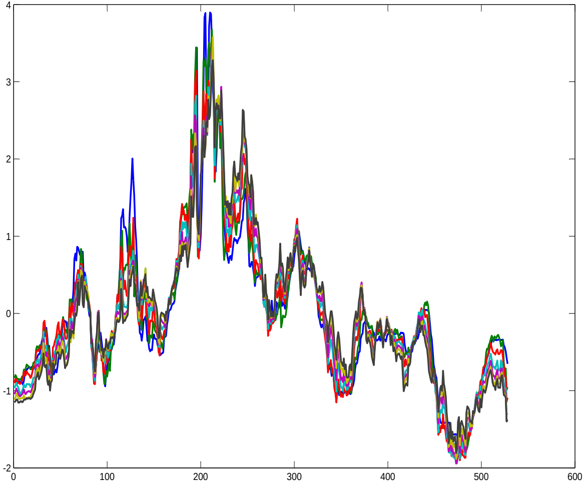


Figure: Normalized Bond yield data

Step 2: Eigenvalue decomposition of covariance matrix

$$EY_tY_t' = \Sigma_y = W\Lambda W'$$

Eigenvalue decomposition of covariance matrix in MatLab
$[W, \Lambda] = \text{eig}(\Sigma_y)$
gives Λ with eigenvalues in *ascending* or *descending* order depending on the properties of Σ_y.

- This can be confusing, since in the literature the *first* eigenvalue usually means the largest eigenvalue

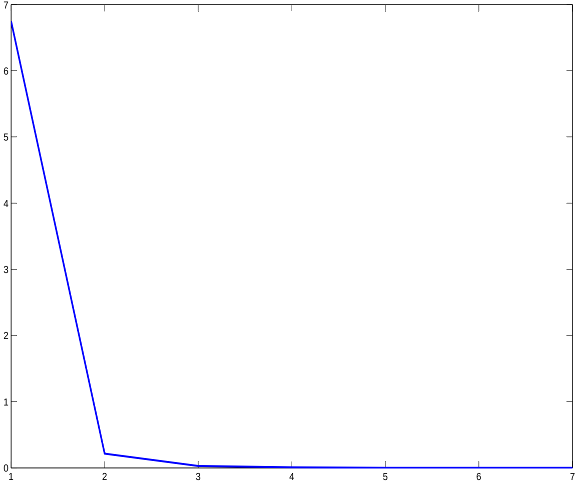


Figure: Scree plot

Step 3: Get the factors

Since

$$\widehat{Y}_t = WF_t$$

and $W^{-1} = W'$ we can get the factors from

$$F_t = W'\widehat{Y}_t$$

The three top rows of F_t contains the principal components (i.e. the factors) associated with the three largest eigenvalues.

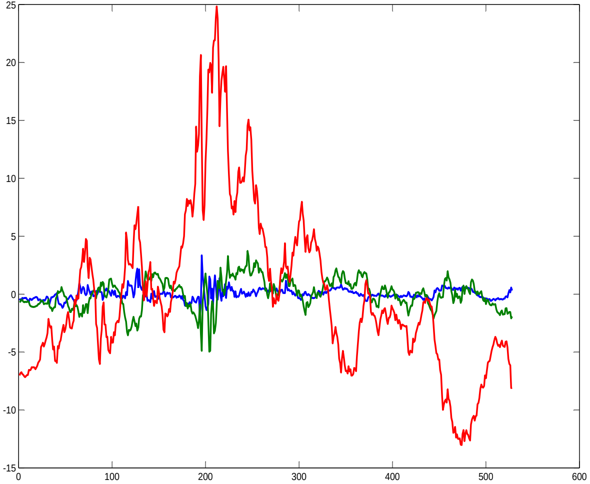


Figure: Time series of (first) three factors

Let's check how we are doing

If three factor model is a good representation of the data, three factors should be able to fit the data well

$$\begin{aligned}\widehat{Y}_t &= WF_t \\ &= \begin{bmatrix} w_1 & \cdots & w_r \end{bmatrix} \begin{bmatrix} f_{1,t} \\ \vdots \\ f_{r,t} \end{bmatrix}\end{aligned}$$

Let's plot the fitted values using only the first factors

$$\widehat{Y}_t^{fit3} = \begin{bmatrix} w_1 & w_2 & w_3 \end{bmatrix} \begin{bmatrix} f_{1,t} \\ f_{2,t} \\ f_{3,t} \end{bmatrix}$$

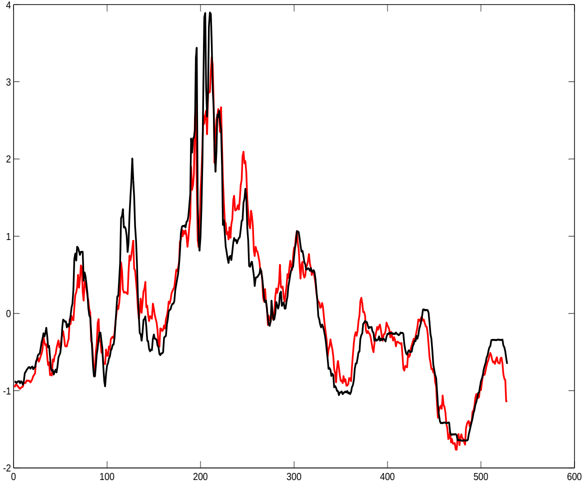


Figure: Fit of Federal Funds Rate using only 1^{st} PC

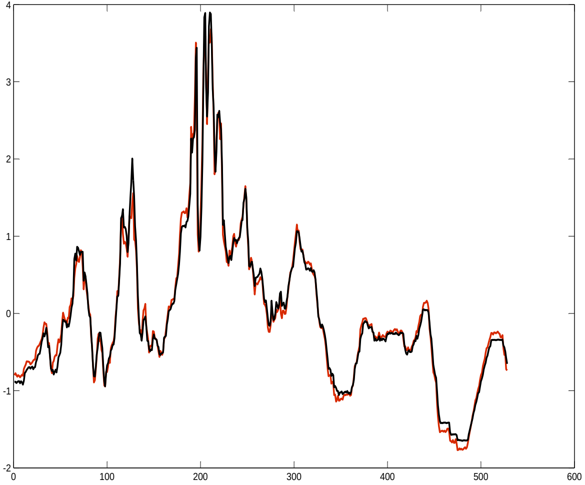


Figure: Fit of Federal Funds Rate using 1st and 2nd PC

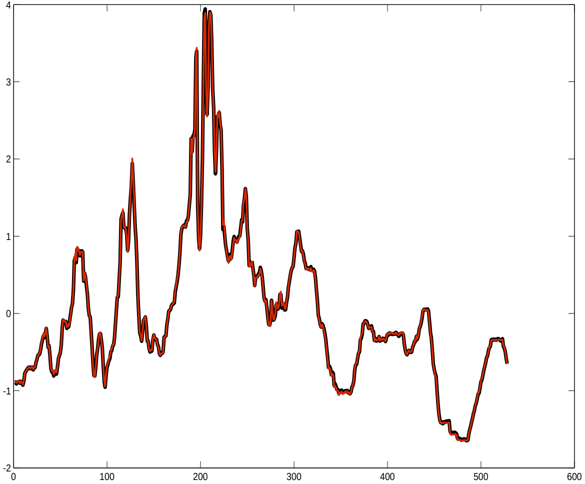


Figure: Fit of Federal Funds Rate using 1^{st}, 2^{nd} and 3^{rd} PC

Step 4

Estimate the dynamic evolution of the factors

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

This can be done OLS so that

$$A = \sum_{t=2}^{T} F_{3,t} F_{3,t-1} \left[\sum_{t=2}^{T} F_{3,t-1} F_{3,t-1} \right]^{-1}$$

What have we achieved?

We now have an estimated model for the dynamics of $\widehat{Y}_t$

$$\begin{aligned}\widehat{Y}_t &= \begin{bmatrix} w_1 & w_2 & w_3 \end{bmatrix} F_{3,t} \\ F_{3,t} &= AF_{3,t-1} + u_t\end{aligned}$$

where we only estimated 15 + 7x3 = 36 parameters (W, A and $Eu_t u_t'$)

- Compare this with the 49 + 28 = 77 parameters of a 7 variable VAR.

Can we interpret the factors?

Not really, but they have been given names that are suggestive for their effect on the yield curve.

- Plot the columns of $\begin{bmatrix} w_1 & w_2 & w_3 \end{bmatrix}$

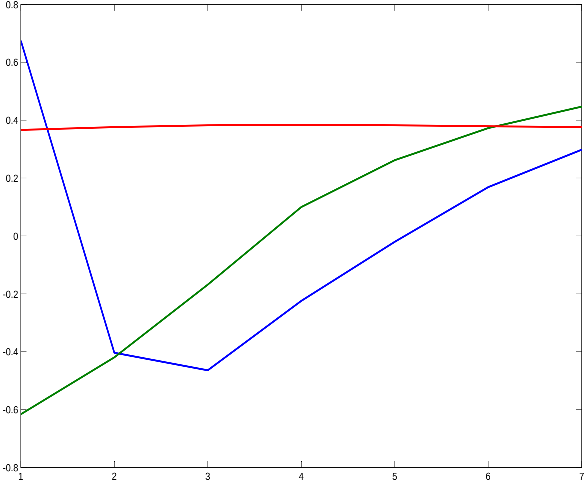


Figure: Factor loadings (Level, slope and curvature)

What can we do with a factor model?

- Forecasting
- FAVARS

Forecasting with factor models

We can produce s period forecasts for all N variables by using

$$E\left[Y_{t+s} \mid F_t\right] = WA^s F_t$$

Advantages:

- Dimension reduction often improves out-of-sample forecasting
- No degrees of freedom problems

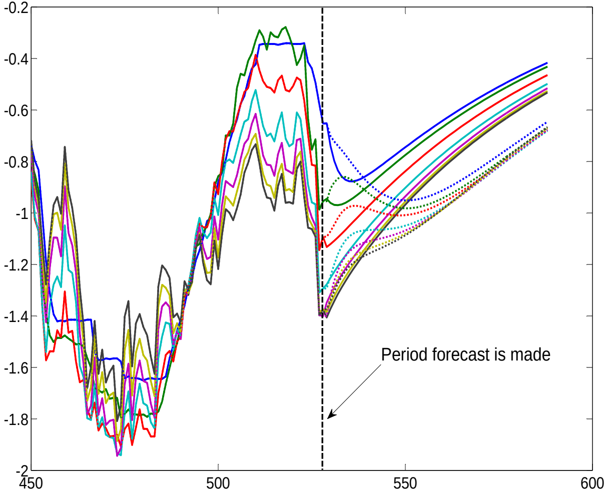
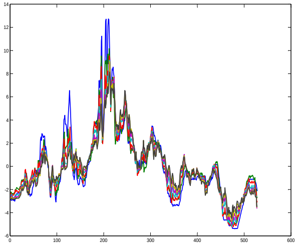


Figure: Actual and forecasted bond yields from factor model (solid) and 7 dimensional VAR(1) (dottted)

Warning: Eye-balling data is not always informative....

Factor structure in bond yields is obvious from looking at bond yield data



- This is not always the case, especially for macro data

Example: Australian Macro data (15 series)

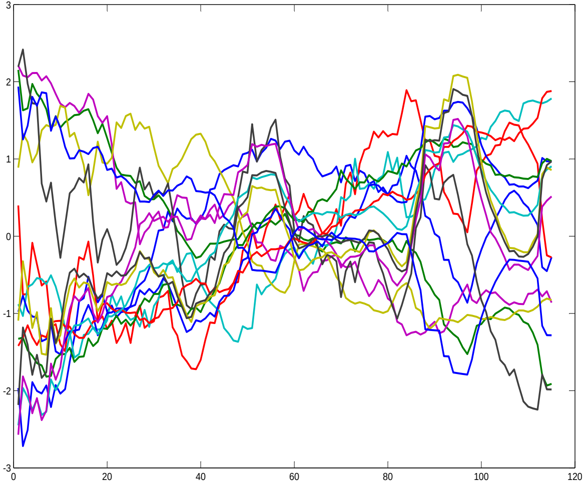


Figure: Australian Macro data (15 series)

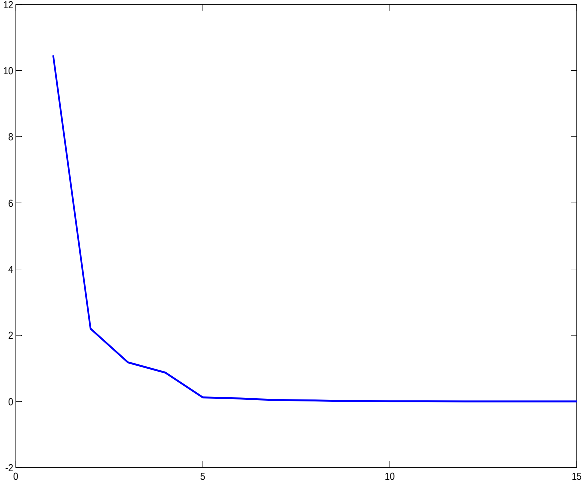


Figure: Scree plot for Australian Macro data (15 series)

FAVARS: Combining large data sets with identifying assumptions

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

- Investigate the effects of a monetary policy shock
- Use more information than what is possible in SVAR
  - Degrees of freedom issues for large N/T
- Avoid price puzzle
- IRFs can be computed to more variables

FAVAR implementation

Consider the model

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

where F_t are factors constructed from some macroeconomic variables of interest Y_t.

- The basic premise is that the factors F_t help describe the dynamics of the variables of interest Y_t.

Strategy:

1. Find the factors
2. Estimate reduced form FAVAR
3. Identify the the effects of a policy shock using standard techniques (contemporaneous restrictions)

Step 1: Finding the factors

We want the factors to capture information that is orthogonal to r_t
Start by regressing Y_t on r_t

$$\widehat{Y}_t = \beta r_t + \upsilon_t$$

The factors F_t can then be constructed as the principal components of the residuals υ_t

Step 2: Estimate the reduced form of the FAVAR

The reduced form of the FAVAR can be found by estimating

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

by OLS.

Step 3: Identify shock to r_t

With r_t ordered last, this can be done using Cholesky decomposition of errors covariance

$$\widehat{\Omega} = (T-1)^{-1} \sum \left(\begin{bmatrix} F_t \\ r_t \end{bmatrix} - \widehat{\Phi} \begin{bmatrix} F_{t-1} \\ r_{t-1} \end{bmatrix} \right) \times \left(\begin{bmatrix} F_t \\ r_t \end{bmatrix} - \widehat{\Phi} \begin{bmatrix} F_{t-1} \\ r_{t-1} \end{bmatrix} \right)'$$

so that we can recover A_0 and C in

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

Step 3: Identify shock to r_t

To compute the impulse response of the variables in Y_t to identified shock, use

$$\frac{\partial Y_{t+s}}{\partial \varepsilon_t^r} = \left[\begin{array}{cc} W & \beta \end{array} \right] \Phi^s C_r$$

since

$$Y_{t+s} = WF_{t+s} + \beta r_{t+s}$$

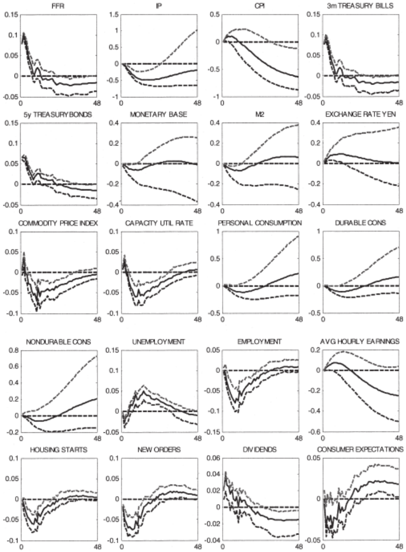


FIGURE II
Impulse Responses Generated from FAVAR with Three Factors and FFR Estimated by Principal Components with Two-Step Bootstrap

Summing up factor models

Dynamic Factor Models

- Dimension reduction: Suitable when large number of time series are available that share a lot of common variation
- Feasible when data sets are "tall", i.e. with many time series but relatively short samples
- Good record in forecasting
- Can be combined with identifying assumptions to estimate effects of structural shocks on all the variables in the sample

Implementation

- Scree plots can be used to check if factor structure is appropriate
- Easy to estimate: Eigenvalue decomposition of covariance matrix + OLS

That's it for today.