ECON 616: Lecture 9: Factor Models

ED HERBST

October 30, 2018

Factor Models

- We saw in the last lecture that it's import to capture all of the observables in order to properly estimate some macroeconometric models.
- VARs were a good candidate, but more data => more parameters
- Too much even for Bayesian methods to overcome (maybe)

One way to summarize a lot of data: factor models.
In this lecture, I'll summarize factor models following Stock and Watson (2010).

Introduction

Macro data:

- Big N: lots of series, macro, financial, sectorial
- Small T: max 70 years of postwar data.
- When $N > T$, estimate via conventional methods become difficult.

Solution: Use a factor model

- Long history in statistics
- Dynamic extension: Geweke (1977)
- Macro factors: Sargent and Sims (1977)

Consistent finding: dynamic factors explain a lot of the variance in many US macro time series [Giannone, Reichlin, and Sala (2004), Watson (2004)]

Details

- N *normalized* series with T observations each, $X_t = [x_{1t}, x_{2t}, \ldots, x_{Nt}]$.
- q dynamic factors f_t
- The factor model

$$\begin{aligned} f_t &= \Phi(L) f_{t-1} + e_t \\ X_t &= \Lambda(L) f_t + \eta_n \end{aligned}$$

  - $\Phi(L)$ is $q \times q$ lag polynomial matrix, $\Lambda(L)$ is an $N \times q$ lag polynomial, e_t, η_t iid mean zero.
  - i th lag polynomial $\Lambda_i(L)$ is called the dynamic factor loading for the *ith* series, X_{it} and $\Lambda_i(L) f_t$ is called the common component of the *ith* series.
  - Assume all the series are stationary

Consideration

"Suppose one knew f_t, and(e_t, η_t) are Gaussian, then one can make efficient forecasts for an individual variable using the population regression of that variable on the lagged factors and lags of that variable."
This means that the forecaster gets the benefit of using all N variables through only q factors (a smaller number.)
Optimal one step forecast if e_t iid:

$$\begin{aligned} E[X_{it+1}|X_t, f_t, X_{it-1}, f_{t-1}, \ldots] &= E[\Lambda(L)f_{t+1} + e_{t+1}|X_t, f_t, X_{it-1}, f_{t-1}, . \\ &= E[\Lambda(L)f_{t+1}|f_t, f_{t-1}] + E[e_{t+1}|X_t, f_t, X_{i} \\ &= \lambda_i(L)\Phi(L)f_t \end{aligned}$$

Sounds great! How to estimate one of these models?

1. Kalman Filter
2. Cross Sectional Averages (nonparametric)
3. Hybrid Approach

Kalman Filter

Assume $\Lambda(L)$ and $\Phi(L)$ are 1st order polynomials (only one lag), without loss of generality.

$$\begin{aligned} X_t &= \Lambda f_t + e_t \\ f_t &= \Phi f_{t-1} + \eta_t \end{aligned}$$

people sometimes call this the "static form" (somewhat confusing)

e_t follows indiviudal AR: $d_i(L)e_t = \xi_{it}, i = 1, \ldots, N$,
$\xi_{it} \sim iidN(0, \sigma^2_{\xi_i})$

η_t is id $N(0, \sigma^2_{\eta_i})$.

This is in state space form!

We can use the Kalman filter to estimate this system!
Advantages:

- Can handle irregular data (weekly/monthly), by changing the row of Λ
- See: Aruoba, Diebold, and Scotti (2009) for a DFM with a single factor explaining weekly, monthly, and quartely variables.

Disadvantages:

- Maximizing the likelihood is difficult – normalization
- Computationally intense: need EM algorithm or Kalman Smoother.
- Number of Parameters is proportion to N!

Second Approach: Cross-Sectional Averaging

Use representation:

$$X_t = \Lambda f_t + e_t$$

Weighted averages of e_t will converge to zero by WLLN

This means that only linear combination of factors remain.

Cross Sectional Averaging Estimators

- Nonparametric (equation for factors, e_t don't matter)
- f_t is treated as a q dimensional factor using N -dimensional data.
- Chamberlain and Rothschild (1983) approximate factor model conditions:

$$N^{-1}\Lambda'\Lambda \to D_\Lambda, \text{ with } D_\Lambda \text{ full rank}$$

$$maxeig(\Sigma_e) \leq c < \infty \text{ for all } N.$$

Approximate Factor Model Assumptions

1. Factors are "pervasive" (they affect most or all of the series) and that the factors are heterogenous (so that the columns of Λ are not too similar.)
2. We also need to limit the influence of idiosyncratic component across series

Intuition: Consider using $\hat{f}_t$ to be the weighted average of X_t using weights W (*not random*), so that $W'W/N = I_N$.

if $N^{-1}W'\Lambda \to H$ as $N \to \infty$ with H full rank, and if the APF conditions hold:

$$N^{-1}W'(\Lambda f_t + e_t) = N^{-1}W'\Lambda f_t + N^{-1}W'e_t \to Hf_t \text{ as } N \to \infty$$

because $N^{-1}W'\Lambda \to H$ by assumption and $N^{-1}W'e_t \to 0$ by WLLN.

Since H is full rank, our estimator consistently estimates the space spanned by factors

Principal Components Estimation

The principal components estimator is the weighted averaging estimator (9) with $W = \hat{\Lambda}$, where $\hat{\Lambda}$ is the matrix of the eigenvetcors of the sample variance matrix X_t

$$\hat{\Sigma}_x = \frac{1}{T}\sum_{t=1}^{T} X_t X_t'$$

associated with the q largest eigenvalues.

Can be derived as solution to least squares problem:

$$\min_{f_1,\ldots,f_T} \frac{1}{NT}\sum_{t=1}^{T}(X_t - \Lambda F_t)'(X_t - \Lambda F_t), \qquad (1)$$

subject to the normalization $N^{-1}\Lambda'\Lambda = I_q$. Then

$$\hat{f}_t = N^{-1}\hat{\Lambda}' X_t$$

Some Details

Theory

- Consistency: Connor and Korajczyk (1986), Stock and Watson (2002)
- Asymptotic Distributions: Bai (2003), Bai \ Ng (2006)

Important Extension: Generalized Principal Components Estimation

- Analogue of Generalized Least Squares to Least Squares
- Idea: Σ_e is not proportional to identity matrix, then can do better by account for these correlations.
- Use Σ_e^{-1} as weighting matrix in Equation 1
- Solution: use scaled Eigenvectors of $\Sigma_e^{-1/2}\hat{\Sigma}_x\Sigma_e^{-1/2}$.
- In practice, many procedures for this: Forni et al. (2005), Stock and Watson (2005)

Hybrid Approach

Newest variants of esimates merge approaches:

- Statitistical Efficiency of State Space (+ missing data!)
- Robustness and Convienence of PCA

This is particulary helpful with noisy idiosyncratic disturbances

.

General Approach

1. estimate factors by PCA (or generalized PCA)
2. use estimated factors to estimate the unknown parameters of state space model.

This is particularly helpful if lags of the factors explain X_t, since we can achieve dimensionality reduction using the KF.

Kalman smoother: can delive time-series averaging => improved estimate.

Comparisons

PC vs GPC

- Forni et al. (2005) Monte Carlo study GPC works better.
- Boving and Ng (2005) different MC study: GPC is not much better.
- Most people just use PC.

PC vs Hybrid Approach

- Doz, Giannone, and Reichlin (2006): N is small, hybrid/State Space approach is better, but negligible once N, T reach 50.
- Reiss and Watson: even for big N, T, is there is a lot of idiosyncratic noise, hybrid/State Space approach could yield improvements.

Estimating The Number of Factors q

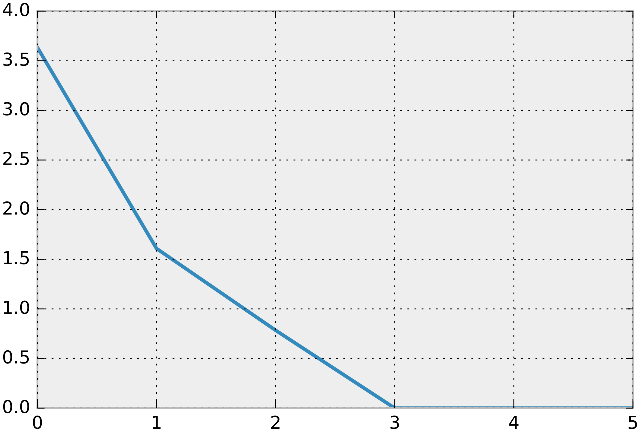
Scree plot: plot of ordered eigenvalues of $\hat{\Sigma}_x$ against the rank of that eigenvlaues.

They can help assess the marginal contribution of the ith principal component for explaining X_t.

Let's look at some Scree plots.

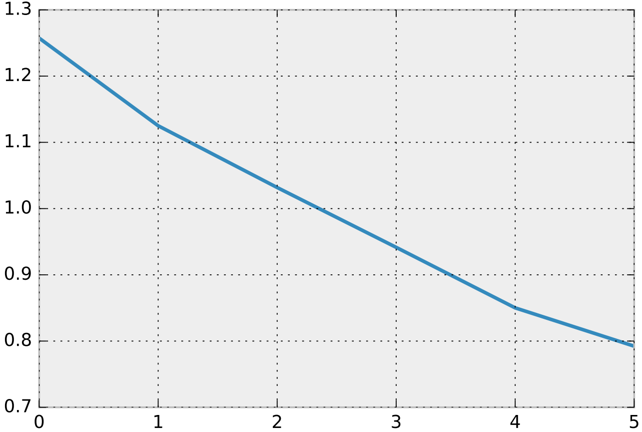
Can formalize: see Onatski (2009).

Scree Plot for Y = X A + E



X = 100x3 N(0,1),
A = 3 $\times$ 6,
$E = 200x6\ N(0, 0.01^2)$.

Scree Plot for 6 Independent (Normal) Series



Estimation of q based on information criteria

Bai and Ng (2002) developed a family of estimators for q that are motivated by the information criteria bused on model seelction:

- benefit of additional factor: fit in sample better
- cost of additional factor: higher sampling variability

Minimize a penalty likelihood or log sum of squares, where the penalty factor incrases linearly in the number of factors (parameters)

$$IC(q) = \ln V_q(\hat{\Lambda}, \hat{f}) + q \times g(N, T)$$

A good g: $g(N, T) = \frac{N+T}{NT} \ln(\min(N, T))$
With dynamic factors things are harder.

Why Estimate a Dynamic Factor Model?

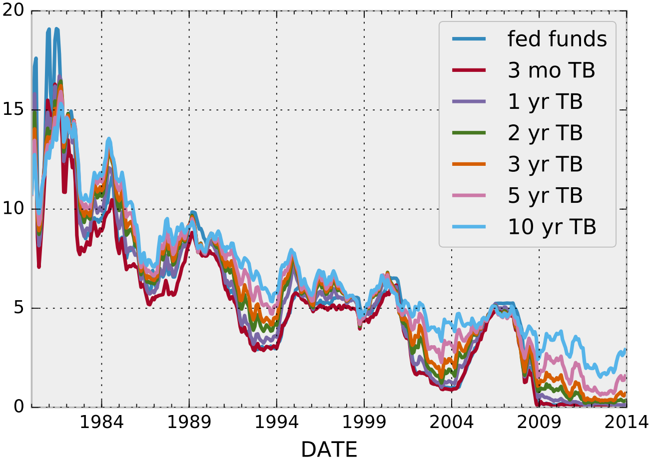
- Forecasting (basically all of the paper cited here)
- Data Reduction – Combine an estimated factor model with
  - VAR Bernanke Kuttner
  - Boivin and Giannonne

Some Data

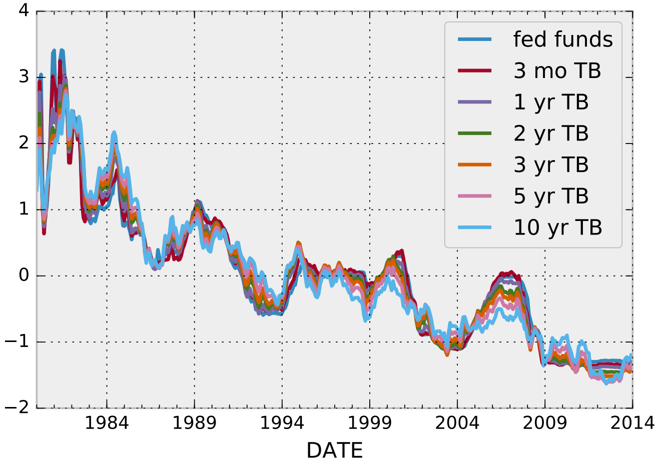
Let's look at the yield curve:

- Fed Funds Rate
- 3 month TB,
- 1, 2, 3, 5, 10 TB Treasuries
- Sample period: 1980 - 2014

Yields

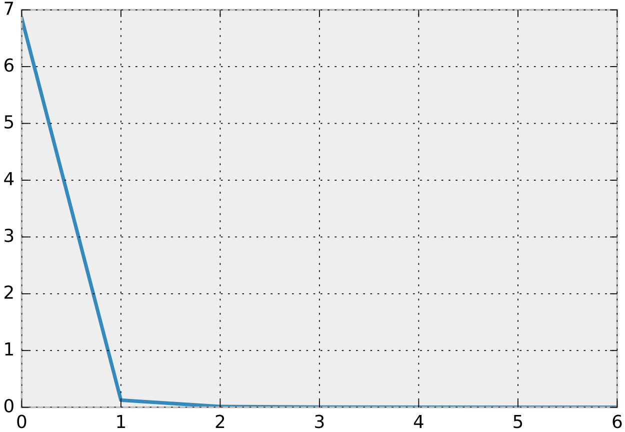


20
15
10
5
0
1984
1989
1994
1999
2004
2009
2014
DATE
fed funds
3 mo TB
1 yr TB
2 yr TB
3 yr TB
5 yr TB
10 yr TB

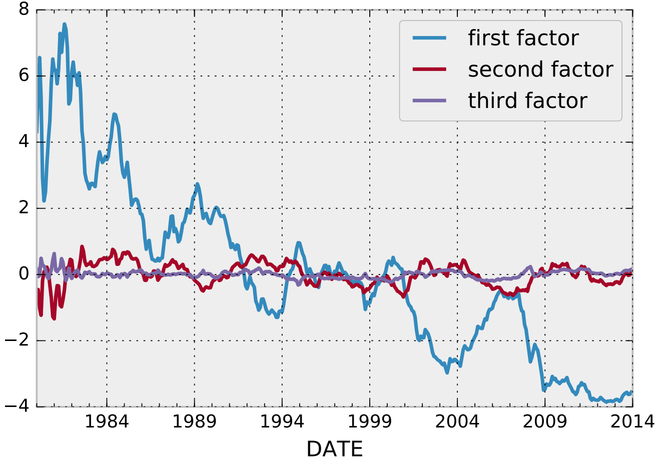


4
3
2
1
0
−1
−2
1984
1989
1994
1999
2004
2009
2014
DATE
fed funds
3 mo TB
1 yr TB
2 yr TB
3 yr TB
5 yr TB
10 yr TB

Scree Plot for Bond Yields

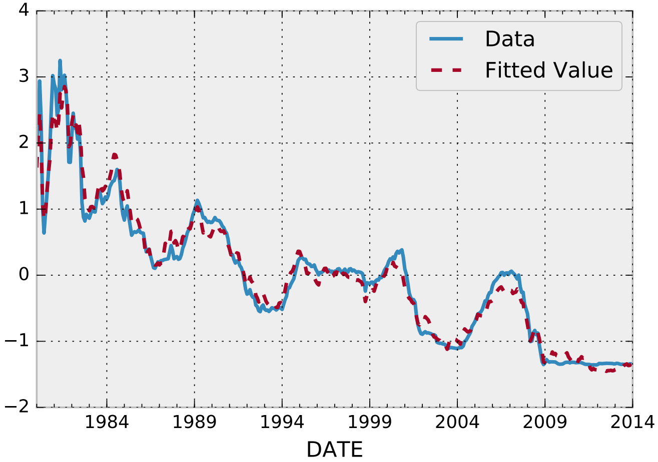


The Factors

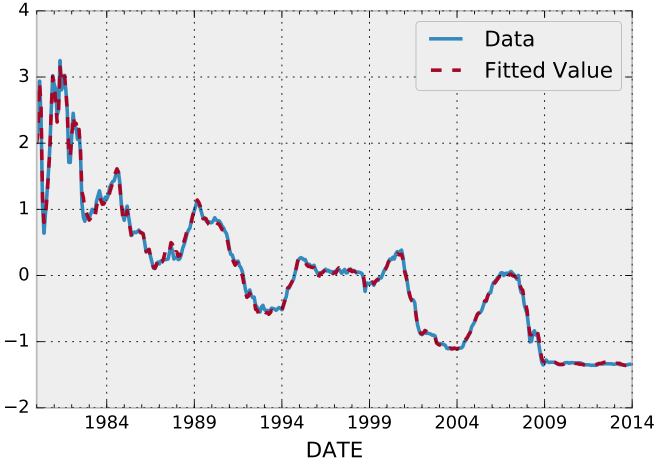


first factor
second factor
third factor
8
6
4
2
0
−2
−4
1984
1989
1994
1999
2004
2009
2014
DATE

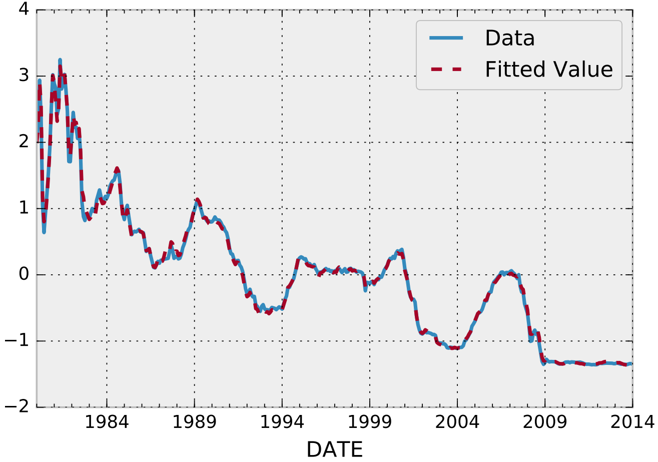
Predicting the 3 Month TB, Using 1 Factor



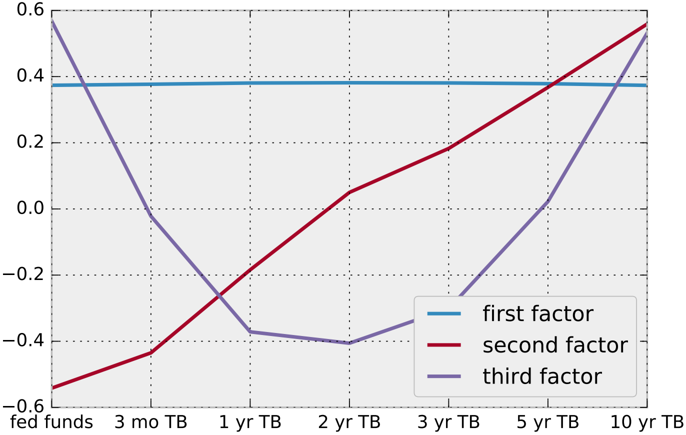
Predicting the 3 Month TB, Using 2 Factors



Predicting the 3 Month TB, Using 3 Factors



The Loadings



Level, Slope, Curvature?

Forecasting With Factor Models