Lecture 14

Factor Models

Motivation

Last time, we discussed structural VARs. One of our main concerns was that shocks might not be fundamental for the system that we considered. Recall "price puzzle" example from last time – in a 2 variable VAR, due to non-fundamentalness, tightening of monetary policy might appear to induce inflation when in reality inflation caused monetary policy to tighten. Another way to state our concern is that the space of structural shocks in not spanned by the residuals in our VAR. A potential solution is to add more variables. This is not a good idea because:

- Number of parameters increases too fast with number of variables

- Number of shocks = number of variables – how do we interpret them all?

- Which variables do we use? – many variables measure similar things

Factor-augmented VARs are another approach.

Setup

Assume: there are a small number of structural shocks explaining co-movements of a large number of time series.

Our notation will follow Stock & Watson (2005).

- x_t is a stationary $n \times 1$ vector time-series. We have observations for $t = 1, ..., T$. Both n and T are large. For example in Stock and Watson, $n = 132$.

- f_t is a $q \times 1$ vector of dynamic factors

Each element of x_t is given by:

$$x_{it} = \tilde{\lambda}_i(L)f_t + u_{it} \tag{1}$$

where $\tilde{\lambda}_i(L)$ is order p and u_{it} might be autocorrelated,

$$u_{it} = \delta_i(L)u_{it-1} + v_{it} \tag{2}$$

and the dynamic factors follow a VAR process:

$$f_t = \Gamma(L)f_{t-1} + \eta_t \tag{3}$$

Assumptions:

- $E[u_{it}u_{js}] = 0 \ \forall i \neq j$. This is a stronger assumption than necessary – we could allow for weak cross-correlation, in which case we would call this an approximate dynamic factor model

- $E[f_t u_{is}] = 0$

- $E[v_{is} v_{it}] = 0 \; \forall s \neq t$

This is called an exact dynamic factor model. It is exact because of the first of the three assumptions just listed – all covariance among variables is due to the factors. It is dynamic because both current and lagged factors affect x_{it} (1).

If we plug (2) into (1), we have:

$$(I - \delta_i(L)L)x_{it} = \lambda_i(L)f_t + v_{it} \tag{4}$$

where $\lambda_i(L) = (I - \delta_i(L)L)\tilde{\lambda}_i(L)$. This equation along with (3) is another way of specifying the model. We can write (4) is matrix form as:

$$X_t = D(L)X_{t-1} + \lambda(L)f_t + v_t \tag{5}$$

where $D(L) = \begin{bmatrix} \delta_1(L) & \dots & 0 \\ 0 & \ddots & \vdots \\ 0 & \dots & \delta_n(L) \end{bmatrix}$.

We can also write the model in static form as

$$X_t = D(L)X_{t-1} + \Lambda F_t + v_t \tag{6}$$
$$F_t = \Phi(L)F_{t-1} + G\eta_t \tag{7}$$

where $F_t = [f_{t-1}, ..., f_{t-p}]$ is size $r \times 1$. We call F_t the static factors. G is an $r \times q$ matrix, where $r \leq qp$. (Recall that q is the number of dynamic factors, and p is the order of the $\lambda(L)$, $i.e.$ the number of lags of f_t that enter the equation for X_t).

Finally, we can write everything together as a VAR

$$\begin{bmatrix} F_t \\ X_t \end{bmatrix} = \begin{bmatrix} \Phi(L) & 0 \\ \Lambda\Phi(L) & D(L) \end{bmatrix} \begin{bmatrix} F_{t-1} \\ X_{t-1} \end{bmatrix} + \varepsilon_t \tag{8}$$
$$\varepsilon_t = \begin{bmatrix} I \\ \Lambda \end{bmatrix} G\eta_t + \begin{bmatrix} 0 \\ v_t \end{bmatrix}$$

Notice that the factor structure has led to many restrictions in this VAR. We have a 0 in the upper right block of the VAR. $D(L)$ is block diagonal. The variance matrix of ε_t is not full rank; it has rank q, the number of dynamic factors.

Plan of FAVAR analysis

1. Estimate reduced form model : estimate parameters & identify factors.

2. Invert to get MA representation

3. Structural analysis: label and identify structural shocks and get structural IRFs

Step 1: Estimation of Factor structure

The problem of estimating factor structure is quite complicated and non-standard. We start with much simpler model first. Consider a simplified model without dynamics(lags):

$$X_t = \lambda F_t + v_t \tag{9}$$

X_t is $n \times 1$, λ is $n \times r$, F_t is $r \times 1$, and v_t is $n \times 1$. F_t are not uniquely defined: you can take any non-singular $r \times r$ matrix A and change $\widetilde{F}_t = AF_t, \widetilde{\lambda} = \lambda A^{-1}$. We always consider F_t as defined up to linear transformation. There are two approaches to estimating factor model: MLE and least squares.

Method 1: MLE

The first approach to estimating this was by MLE. That is, one can assume that errors are normal $v \sim N(0, \text{diag}(\sigma_1^2, ..., \sigma_n^2))$, then the log likelihood is proportional to

$$l_n \sim -\frac{T}{2} \sum_{i=1}^{T} \log \sigma_i^2 - \sum_i \sum_t \frac{1}{2\sigma_i^2}(x_{it} - \lambda_i F_t)^2$$

We need to maximize log-likelihood over $\{\lambda_i, F_t, \sigma_i^2\}$ to get estimates of factors and factor structure.
This has two down-sides:

- **Numerical**: Log-likelihood is a non-linear function of many parameters($r(n+T)+n$), since we need to maximize over $\{\lambda_i, F_t, \sigma_i^2\}$. Maximizing over so many parameters will be difficult.

- **Non-standard asymptotics**: MLE asymptotics treats the number of parameters as fixed, however, here the number of parameters are growing with the sample size (the number of F_t grows with T and the number of λ_i grows with the number of series, n). This situation, where the number of parameters grows with the sample size, is called the *incidental parameters problem*. As a result the usual MLE results about consistency and efficiency do not apply.

Method 2: Principal Components

No dynamics. We estimate by least squares:

$$\min_{\lambda, F} \sum_{i=1}^{N} \sum_{t=1}^{T} (x_{it} - \lambda_i F_t)^2$$

Unlike the likelihood, we can minimize this objective function analytically. However, we will need some normalization restrictions. Usually, we assume $F'F/T = I_r$. As on problem set 2, the solution to the problem is

$$F = \sqrt{T}[r \text{ eigenvectors corresponding to the largest eigenvalues of } XX']$$

Bai (*Econometrica* 2003 & other papers) gives asymptotic results for this estimator. Estimates of F_t is consistent and $\sqrt{N}$- asymptotically normal if $\frac{\sqrt{N}}{T} \rightarrow 0$. In general, the speed of convergence for F is $\min\{\sqrt{N}, T\}$.

Put dynamics back. Now, instead of considering the simplified model (9), let's return to the full model, in static form (6). The objective function is

$$\min_{D(L), \lambda, F} \sum_{i=1}^{N} \sum_{t=1}^{T} ((I - D(L)L)x_{it} - \lambda_i F_t)^2$$

How to find optimum numerically? Observe that if we knew, F_t, it would be easy to estimate λ and D. Consider the following iterative estimation procedure:

1. pick D

2. let $\tilde{X}_t = (I - D(L)L)X_t$

3. principal components of $\tilde{X}_t$ give $\hat{F}$

4. Regress X_t on its lags and $\hat{F}$ to get a new D

5. repeat until convergence

On each step the sum of squares decreases. Additional justification is needed to prove that we would converge to global maximum. The procedure above gives estimates of $\hat{F}$, $\hat{\Lambda}$, and $\hat{D}$. These estimates are consistent and asymptotically normal (see Bai). When we say that $\hat{F}$ and are consistent, we mean that the space spanned by $\hat{F}$ converge to the spaces spanned by F (factors are not uniquely defined!!! Remember, they are latent).

We still need to estimate Φ, G, f_t, and the variance of the shocks. Note that F_t is what's called "static factors" while f_t is dynamic factors. The difference that F_t is "stochastically dynamically singular", that is there are some linear combinations of lags of F_t which are exactly zero (remember that we formed F_t as p lags of f_t). Surely, for the series of estimated $\widehat{F}_t$ would not look like lags of another time series. But the dynamic residuals will be singular. Namely,

- We estimate $\hat{\Phi}$ by regressing $\hat{F}_t$ on its lags and form residuals

$$\hat{e}_t = \hat{F}_t - \hat{\Phi}(L)\hat{F}_{t-1}$$

- If e_t would be true errors from regression of "true" factors F_t on its lags then variance-covariance matrix of them is not of full rank; it's rank is equal to the number of dynamic factors q.

- One can show that sample covariance matrix of $\hat{e}_t$ $\Omega_1 = \frac{1}{T}\sum \hat{e}_t\hat{e}_t'$ has nearly factor structure. That is $\hat{e}_t = G\eta_t + error$, where η_t is $q(q < r)$ dimensional dynamic shock(shock to dynamic factor) and error nearly idiosyncratic. A consistent estimate of η_t can be obtained by extracting q principle components of $\widehat{e}_t$. (Again remember that we extracting shocks up to linear transformations).

Aside How to define the number of static and dynamic factors? Two conflicting approaches. One is information type criteria (looks like BIC and AIC)- see a series of papers by Bai and Ng (2002 & 2005). The second approach is testing suggested (separately) by Onatski and Harding. They usually suggest more factors than implied by info criteria.

Step 2: Inverting to get MA

Second step is easiest. You have to invert all lag operators (one can do it by subsequent substitution). Observe

$$F_t = (I - \Phi(L)L)^{-1}G\eta_t$$

so,

$$
\begin{aligned}
(1 - D(L)L)X_t &= \Lambda F_t + v_t \\
&= \Lambda((I - \Phi(L)L))^{-1}G\eta_t + v_t \\
X_t &= (1 - D(L)L)^{-1}\Lambda((I - \Phi(L)L))^{-1}G\eta_t + (1 - D(L)L)^{-1}v_t \\
&= B(L)\eta_t + C(L)v_t
\end{aligned}
$$

Here $B(L)$ is reduced form IRFs. To get structural IRFs we need to find a rotation A such that $\eta_t = A^{-1}\xi_t$ and ξ_t has structural interpretation. Then $B^*(L) = B(L)A$ is a set of structural IRFs. And $X_t = B^*(L)\xi_t$

Step 3: Structural Analysis.

Next time

14.384 Time Series Analysis
Fall 2013