Factor Models Part 2

Summary of FAVAR

Take the same model as last time:

$$x_{it} = \lambda_i(L)f_t + \delta_i(L)x_{it-1} + v_{it} \tag{1}$$
$$f_t = \Gamma(L)f_{t-1} + \eta_t$$

Step 1: estimation

The space spanned by the factors f_t is consistently estimable, *i.e.* there exists an invertible H such that

$$||\hat{f}_t - Hf_t||_2 \xrightarrow{p} 0$$

We estimate $\hat{f}_t$ in two steps:

1. *Static factors*: Estimate static factors F_t, which is size $r \times 1$, with $r \geq q$ (f_t is $q \times 1$) by iterations:

    (a) pick $\delta_i(L)$

    (b) let $\tilde{x}_{it} = (I - \delta_i(L)L)x_{it}$

    (c) principal components (eigenvectors corresponding to largest eigenvalues) of $\tilde{x}_{it}$ give $\hat{F}$

    (d) Regress x_{it} on its lags and $\hat{F}$ to get a new $\delta_i(L)$

    (e) repeat until convergence

2. *Dynamic factors*: the static factors evolve as:

    $$F_t = \Phi(L)F_{t-1} + \varepsilon_t$$

    where $\varepsilon_t = G\eta_t$ and ε_t is $r \times 1$, G is $r \times q$, and η_t is $q \times 1$, so the variance-covariance matrix of ε_t is not full rank. Two ways to estimate space spanned by η_t:

    - Observe

      $$x_{it} = \lambda_i\Phi(L)F_t + \delta_i(L)x_{it-1} + \epsilon_{xit}$$
      $$\epsilon_{xit} = \lambda_i G\eta_t + v_{it}$$

      so regress x_{it} on its lags F_t and lags to get residuals, $\hat{\epsilon}_{xit}$. Take the q principal components of $\hat{\epsilon}_x\hat{\epsilon}'_x$ to estimate the space spanned by η

    - Observe : $F_t = \Phi(L)F_{t-1} + \varepsilon_t$ and $\varepsilon_t = G\eta_t$. Take the q principal components of $\hat{\varepsilon}_F\hat{\varepsilon}'_F$

Step 2: reduced form MA

We can write the model in MA form as:

$$x_t = B(L)\eta_t + J(L)v_t$$

We can get this representation in two ways: either formally inverting model (1) or by regressing x_{it} on lags of η. Note that B is identified up to linear transformation (as well as shocks): we could replace $B(L)\eta_t$ with $B(L)AA^{-1}\eta_t$ for any invertible A without changing the observed x_t. Just like SVARs, we need to choose an H for identification.

Step 3: Structural Analysis.

Partial (block) identification: Structural analysis is mainly done by timing restrictions. For example, assume that monetary shock is identified as follows. All variables divided into 3 groups: slow moving(react to monetary shock with a lag), fast moving (react to monetary shock immediately) and interest rate (transferring variable), arrange them from slow to fast. Often implausible to specify a full ordering, but suppose can identify variables into three groups:

- slow – investment, GDP, unemployment; say n_s of them

- fast – prices, exchange rate; n_f of them

- identifying – interest rate; 1 of them

Assume there are n_s slow variables, and n_F fast ones, $n = n_S + n_F + 1$. Divide also shocks in to q_S shocks to slow, q_F shocks to fast, and monetary shock ($q = 1 + q_S + q_F$).

Then the identifying assumptions are:

$$B_0^* = \begin{bmatrix} B_{SS}^* & 0 & 0 \\ B_{RS}^* & B_{RR}^* & 0 \\ B_{FS}^* & B_{FR}^* & B_{FF}^* \end{bmatrix}$$

here B^*0 is $n \times q$.

Or we have a system

$$\varepsilon_t^S = B_{SS}^* A_S \eta_t + v_t^S \tag{2}$$

$$\varepsilon_t^R = B_{RS}^* A_S \eta_t + B_{RR}^* A_R \eta_t + v_t^R \tag{3}$$

$$\varepsilon_t^F = B_{FS}^* A_S \eta_t + B_{FR}^* A_R \eta_t + B_{FF}^* A_F \eta_t + v_t^F \tag{4}$$

here ε_t^S is reduced form error to slow variables in regression of each slow variable X_t^i on its own lags and F_t. In the same way we receive ε_t^R and ε_t^F - reduced form errors to Fed rate and fast variables. η_t- reduced form shock to dynamic factors(recovered before), $A_R \eta_t = \xi_t^R$ is monetary shock, $A_S \eta_t = \xi_t^S$ and $A_F \eta_t = \xi_t^F$ are slow and fast shocks (they are identified as sets, not labeled personally).

What can we do to recover ξ_t and B^*? It's called "reduced rank regression".

Reduced rank regression   Let's observe y_t and x_t, $t = 1...T$ such that

$$y_t = \alpha \beta' x_t + e_t$$

where y_t is n-dimensional, while x_t is k dimensional, β is $k \times q$-matrix, α is $n \times q$ matrix, and $q < k$. The idea is that the set of n variables y_t is spanned not by all x_t but by a q dimensional linear combination of them $\beta' x_t$.

Let $S_{..}$ be a set of sample covariance matrices. Define

$$S_{yy.x} = S_{yy} - S_{yx}S_{xx}^{-1}S_{xy}.$$

Find q eigenvectors $\{V_i\}_{i=1}^q$ corresponding to largest eigenvalues of matrix $S_{yy.x}^{-1/2}S_{yx}S_{xx}^{-1}S_{xy}S_{yy.x}^{-1/2}$. Then

$$\widehat{\alpha} = S_{yy.x}^{1/2}V, \quad \widehat{\beta} = S_{xx}^{-1}S_{xy}S_{yy.x}^{-1/2}V, \quad V = (v_1, ..., v_q)$$

Back to structural estimation So we have system (2)- (4).

- First we do reduced rank regression of ε_t^S on η_t imposing rank q_S. The resulting $\widehat{\beta}$ gives us the set of slow shocks $\xi_t^S = \widehat{\beta}\eta_t$.

- By definition $\xi_t^R = \widehat{P}(\varepsilon_t^R|\eta_t) - \widehat{P}(\varepsilon_t^R|\xi_t^S)$- monetary shock is a part of factor innovation unexplained by slow shocks

- We run regression of ε_t^R on η_t, get the explained part, then run regression of ε_t^R on ξ_t^S, get the explained part. Define monetary shock as the difference between the explained parts.

Remark 1. Impulse response of any variable to monetary shock can be found by regressing left-hand variables on our reconstruction of monetary shock.

Number of Factors

- **Static Factors**: Information Criteria

$$\log \hat{\sigma}_k^2 + kC_{NT}$$

where $C_{NT} \to 0$ and $\min\{N,T\}C_{NT} \to \infty$ as both $T \to \infty$ and $N \to \infty$. Bai and Ng suggest $C_{NT} = \frac{\log(\min\{N,T\})}{NT/(N+T)}$

- **Dynamic Factors**: Compute covariance matrix of $u_t = R\epsilon_t$ and look at eigenvalues – only q should be non-zero in population. Bai and Ng (2005) discuss how to choose q in samples. The idea is to estimate covariance matrix of u_t and let $c_1 \geq c_2 \geq ... \geq c_r \geq 0$ be eigenvalues. We need to define where to say that eigenvalues are small enough to assume that they are zeros. One may consider statistics of the form

$$D_{k,1} = \left(\frac{c_{k+1}^2}{\sum_{j=1}^r c_j^2}\right)^{1/2}$$

or

$$D_{k,2} = \left(\sum_{j=k+1}^r \frac{c_j^2}{\sum_{j=1}^r c_j^2}\right)^{1/2}$$

in population they should be 0 for any $k > q$. The suggestion is to choose $\hat{q}$ as minimal k for which $D_{k,i} < \frac{m}{\min\{N^{1/2-\delta}, T^{1/2-\delta}\}}$. The formal statement is that for any $m > 0$ and $\delta \in (0, 1/2)$ the estimate of q is consistent. A practical suggestion is to take $m \approx 1$ and δ close to 0.

Factor IV

We will motivate this with an example. Suppose inflation is given by a Philips' curve:

$$\pi_t = \gamma_f E_t \pi_{t+1} + \gamma_b \pi_{t-1} + \lambda x_t + \epsilon_t$$

where x_t might be a proxy for marginal costs, say the output gap. The usual way to estimate Phillips curve is IV. That is, to use instruments z_t such that:

$$E\left[z_t(\pi_t - \gamma_f \pi_{t+1} - \gamma_b \pi_{t-1} - \lambda x_t)\right] = 0$$

and estimate $\beta = [\gamma_f\, \gamma_b\, \lambda]$ by GMM. The requirements for exogeneity of z_t is that it belongs to information set available at time t (measurable with respect to $\mathfrak{F}_t$). However, there are many candidates for z_t, lags of inflation, various price indices, and other variables and their lags. We have a problem because we have many/weak instruments. Newey and Smith(2004) show that bias of GMM is linear in the number of over-identifying restrictions. We could throw out some of the instruments, but then we'll lose some information. And the estimation of Phillips curve usually show weak identification (means information is valuable).

The following idea is in Bai and Ng(2006). If we assume a factor model as before, that is, there are just few structural shocks that span the space of macro variables, then the factors can be our instruments. That is, assume that

$$y_t = \beta' x_t + \epsilon_t$$

is a regression of interest and all regressors can be divided in to two groups: $x_t = [x_{1t}\, x_{2t}]$: x_{1t} are exogenous, x_{2t} are endogenous ($E[x_{2t}\epsilon_t] \neq 0$). The method of estimation is IV, and there tons of instruments z_t. We assume a factor structure:

$$z_{it} = \lambda_i' F_t + e_{it}$$
$$E[e_{it}\epsilon_t] = 0$$
$$x_{2t} = \phi F_t + u_t$$
$$E[u_t\epsilon_t] \neq 0$$

In this model z_{it} would be valid instruments. However, if the variance of e_{it} is large relative to the variance of $\lambda_i' F_t$, then z_{it} would be weak instruments, while F_t would be a good instrument. This suggests the following FIV procedure:

1. Estimate $\hat{F}$ as eigenvectors of $\frac{zz'}{NT}$

2. Do IV with $\hat{F}$ as instruments

Results

- If $\frac{\sqrt{T}}{N} \to 0$ as $N, T \to 0$, then the asympototic distribution of $\hat{\beta}_{FIV}$ is the same as if we observed F (Bai and Ng 2006)

- Let we run a regular IV using any subset of instruments z_t and get $\hat{\beta}_{IV}$, then the asymptotic variances have the following order:
$$AVar(\hat{\beta}_{FIV}) \leq AVar(\hat{\beta}_{IV})$$

Factor models in Finance.

Refer to Kleibergen (2010) "Reality checks for and of factor pricing".

Many theories of stochastic discount factor results in a statement that portfolio returns exhibit a (unobserved) factor structure (starting from Merton (1973) and Ross (1976))

$$r_{it} = \mu_i + \beta_{i1} f_{1t} + ... + \beta_{ik} f_{kt} + \varepsilon_{it} = \mu_i + \beta_i F_t + \varepsilon_{it}, \quad i = 1, ..., N; t = 1, ..., T \tag{5}$$

here $F_t = (f_{1t}, ..., f_{kt})'$ is $k \times 1$ vector of factors at time t and β_i is $1 \times k$ factor loadings for portfolio i, ε_{it} is idiosyncratic disturbances.

Empirically, in many data sets on quarterly return of different portfolios people find around 3 factors. Example, for the data set on 25 book-to-market sorted portfolios from Ken French's web-site ($N = 25, T = 200$) 3 principle component explains 25% variation. Another example is Jagannathan and Wang set of monthly returns of 100 portfolios ($T = 330, N = 100$), the first 3 principle components explain 86% of variation.

There is a long list of empirical papers suggesting a set of **observed** factors that explains portfolio returns. That is,

$$r_{it} = \tilde{\mu}_i + B_i G_t + e_{it} \tag{6}$$

where $G_t = (g_{1t}, ..., g_{mt})'$ is $m \times 1$ vector of observed factors. Examples of such papers: Fama and French (1993)- where factors are market portfolio, "small-minus-big" (the difference in returns between portfolios of stock with small vs large market capitalization) and "high minus low"(the difference in returns between portfolios of stocks with high vs low book-to-market ratios). Another example, Lettau and Ludvigson (2001)- where factors are market portfolio, the consumption-wealth ratio (cay), consumption growth and income growth.

Stochastic discount factor models imply that

$$Er_{it} = \lambda_0 + B_i \lambda_G,$$

where λ_G is $m \times 1$ vector of factor risk premia, while B_i is $1 \times m$ factor loadings for portfolio i (corresponds to risks associated with the factors). To estimate factor risk premia, there exists a two-pass procedure, sometimes called Fama-MacBeth method:

1. For each portfolio i run a time series regression of r_{it} on G_t to uncover coefficients B_i (estimates called $\hat{B}_i$):

$$\hat{B}_i = \sum_{t=1}^{T} r_{it} (G_t - \overline{G})' \left(\sum_{t=1}^{T} (G_t - \overline{G})(G_t - \overline{G})' \right)^{-1},$$

where $\overline{G} = \frac{1}{T} \sum_{t=1}^{T} G_t$.

2. Regress average(over time) returns $\overline{r}_i = \frac{1}{T} \sum_{t=1}^{T} r_{it}$ on constant and $\hat{B}_i$ to get λ's (factor risk premia)

Problem set 2 contains a problem that discusses this procedure in detail.

To demonstrate the adequacy of the suggested observed factors, researchers usually report second pass R^2 and t-stats for λ's. A large value of R^2 is typically seen as an indicator that the observed factors explain a large part of the variation of the average portfolio returns. The logic is based on the following theoretical and simulation exercise: imagine that the returns have unobserved factor model (5) with 3 factors, but you are trying to fit observed factor model (6) where your observed factors G_t will be exactly equal to: (i) first true factor f_{1t}, (ii) two first factors f_{1t}, f_{2t} and (iii) to all true factors F_t. If you simulate situations (i)-(iii), you will see that the distribution of R^2 is in case (i) is strictly to the left from that in case (ii), and they bothe strictly to the left of that in case (iii).

A new paper by Kleibergen(2010) shows that despite of this, R^2 is a bad indicator of the adequacy of the observed factor structure. He simulate a situation when observed factors G_t have nothing to do with the true unobserved factors F_t and calibrated this situation to a typical asset pricing setup. He showed that R^2 has a non-trivial distribution that has a big chanc of probability located on large values of R^2 (close to 1). This might be called a "spurious factor structure". He also demonstrated that the second run t-statistics are very misleading as well in this case.

Can you see if the factor structure you find is spurious? Kleibergen suggests to look at the residuals e_{it} and check whether they have significant remaining factor structure, if they do, the results you are getting are likely misleading. Kleibergen showed that Fama-French likely correspond to a relevant factor structure, while many other empirical asset pricing factor models are spurious (including Lettau and Ludvigson (2001)).

MIT OpenCourseWare

14.384 Time Series Analysis
Fall 2013