Pisa. April 2009. PLAN OF THE TALK

The talk is limited to Analysis and Prediction of Macroeconomic Time Series.

Part I. Standard techniques (small-dimensional). ARMA and VAR models

Stationary processes.

Wold representation theorem.

ARMA models.

Vector processes. VAR models.

Spectral density, filters, principal components.

Part II. Factor models (large-dimensional).

Dynamic approach.

Static approach.

Work in progress.

I. Stationary processes.

Second-order, discrete-time stationary processes:

$$\mathsf{E}(x_t) = \mu, \qquad \mathsf{E}[(x_t - \mu)(x_{t-k} - \mu)] = \gamma_k$$

Both moments are independent of t.

In order to apply the theory of stationary processes it assumed that macroeconomic time series become stationary by suitable transformations, e.g.

$$x_t = a + bt + C_t, \qquad x_t - x_{t-1} = k + H_t$$

I. Stationary processes.

Examples:

u_t is a white noise if $\mu = 0$ and $\gamma_k = 0$ for $k \neq 0$.

Moving averages of a white noise: $x_t = \Sigma_{h=-\infty}^{\infty} a_k u_{t-k}$, under $\Sigma\, a_h^2 < \infty$.

Linearly deterministic processes, e.g. $x_t = A$, where A is a stochastic variable.

I. Stationary processes. Wold representation theorem

The best linear prediction of x_t, based on its past values, with respect to the mean square criterion is the projection of x_t on the space spanned by its past values:

$$x_t = \text{Proj}(x_t \,|\, x_{t-1},\ x_{t-2},\ \ldots) + u_t,$$

which, under reasonable assumption can be written as

$$x_t = [a_1 x_{t-1} + a_2 u_{t-2} + \cdots] + u_t.$$

The process u_t is called the innovation of x_t and is a white noise.

I. Stationary processes.

Again the projection:

$$x_t = [a_1 x_{t-1} + a_2 u_{t-2} + \cdots] + u_t.$$

Wold representation:

$$x_t = [u_t + b_1 u_{t-1} + b_2 u_{t-2} + \cdots] + d_t, \quad d_t \perp u_s, \quad d_t = \text{Proj}(d_t \,|\, d_{t-1},\, d_{t-2},\, \ldots)$$

the first being the linearly non-deterministic component, the second deterministic. It is assumed that d_t is zero for macroeconomic time series, so that we restrict our analysis to

$$x_t = u_t + b_1 u_{t-1} + b_2 u_{t-2} + \cdots$$

I. Stationary processes.

The most popular model in this class is the ARMA (AutoRegressiveMovingAverage):

$$x_t - \alpha_1 x_{t-1} - \cdots - \alpha_p x_{t-p} = u_t + \beta_1 u_{t-1} + \cdots + \beta_q u_{t-q}$$

also written as

$$(1 - \alpha_1 L - \cdots - \alpha_p L^p) x_t = (1 + \beta_1 L + \cdots + \beta_q L^q) u_t$$

or

$$\alpha(L) x_p = \beta(L) u_t$$

x_t is a moving average

$$x_t = \alpha(L)^{-1} \beta(L) u_t = u_t + \delta_1 u_{t-1} + \delta_2 u_{t-2} + \cdots$$

I. Stationary processes.

x_t is a moving average

$$x_t = \alpha(L)^{-1}\beta(L)u_t = u_t + \delta_1 u_{t-1} + \delta_2 u_{t-2} + \cdots$$

The coefficients δ_k decline geometrically.

Incidentally, the process

$$z_t = u_t + 2^{-1}u_{t-1} + 3^{-1}u_{t-2} + \cdots$$

is not an ARMA.

I. Stationary processes.

Inverting $\beta(L)$ we get

$$\beta(L)^{-1}\alpha(L)x_t = u_t$$

that is

$$x_t = [A_1 x_{t-1} + A_2 x_{t-2} + \cdots] + u_t$$

or

$$\hat{x}_t = A_1 x_{t-1} + A_2 x_{t-2} + \cdots$$

which is the prediction equation.

I. Stationary processes.

The second basic theorem is the spectral representation theorem

$$x_t = \lim_{n\to\infty} \sum_{k=0}^{n} A_k e^{i\theta_k t}$$

with $\theta_k \in [-\pi\ \pi]$. Moreover

$$\text{var}(A_k) = f^x(\theta_k)(\theta_{k+1} - \theta_k)$$

where

$$f^x(\theta) = \frac{1}{2\pi} \sum_{k=-\infty}^{\infty} \gamma_k^x e^{-ik\theta},$$

the Fourier transform of $\{\gamma_k^x\}$. Of course

$$\gamma_h^x = \int_{-\pi}^{\pi} e^{ih\theta} f^x(\theta) d\theta$$

I. Stationary processes.

Again:

$$x_t = \lim_{n\to\infty} \sum_{k=0}^{n} A_k e^{i\theta_k t}$$

with $\theta_k \in [-\pi\ \pi]$. Moreover

$$\text{var}(A_k) = f^x(\theta_k)(\theta_{k+1} - \theta_k)$$

Interpretation of the spectral density.

I. Stationary processes.

Linear filtering:

$$y_t = b(L)x_t = \left[\sum b_k L^k\right] x_t = \sum b_k x_{t-k}.$$

Then

$$f^y(\theta) = |b(e^{-i\theta}|^2 f^x(\theta)$$

This is the basis for construction of filters in the frequency domain. The most famous is the band-pass filter

$$b(\theta) = \begin{cases} 1 \text{ if } |\theta| \leq \theta^* \\ 0 \text{ if } |\theta| > \theta^* \end{cases}$$

Take the Fourier expansion of b

$$b(\theta) = \sum_{k=-\infty}^{\infty} b_k e^{-ik\theta}, \quad b_h = \frac{1}{2\pi} \int_{-\pi}^{\pi} e^{ih\theta} b(\theta) d\theta = \frac{1}{2\pi} \int_{-\theta^*}^{\theta^*} e^{ih\theta} d\theta$$

I. Stationary processes.

Take the Fourier expansion of b

$$b(\theta) = \sum_{k=-\infty}^{\infty} b_k e^{-ik\theta}, \qquad b_h = \frac{1}{2\pi}\int_{-\pi}^{\pi} e^{ih\theta} b(\theta) d\theta = \frac{1}{2\pi}\int_{-\theta^*}^{\theta^*} e^{ih\theta} d\theta$$

The band pass filter is then obtained as

$$\sum b_k L^k$$

I. Stationary processes.

Extension to n-dimensional processes $x_t = (x_{1t}\ x_{2t}\ \cdots\ x_{nt})$.

Definition. Autocovariances are $n \times n$ matrices Γ_k.

White noise: $\Gamma_k = 0$ if $k \neq 0$.

Moving averages: $\Sigma\, a_k u_{t-k}$.

Wold Theorem.

ARMA modeling.

VAR modeling:

$$\begin{aligned} x_{1t} &= b_{11,1}x_{1,t-1} + b_{11,2}x_{1,t-2} + \cdots + b_{11,p}x_{1,t-p} + \cdots + b_{1n,1}x_{n,t-1} + b_{1n,2}x_{n,t-2} + \cdots + b_{1n,p}x_{n,t-p} + u_{1t} \\ &\vdots \\ x_{nt} &= b_{n1,1}x_{1,t-1} + b_{n1,2}x_{1,t-2} + \cdots + b_{n1,p}x_{1,t-p} + \cdots + b_{nn,1}x_{n,t-1} + b_{nn,2}x_{n,t-2} + \cdots + b_{nn,p}x_{n,t-p} + u_{nt} \end{aligned}$$

I. Stationary processes.

Spectral density:

$$f^x(\theta) = \frac{1}{2\pi} \sum_{k=-\infty}^{\infty} \Gamma_k^x e^{-ik\theta},$$

the Fourier transform of $\{\Gamma_k^x\}$. Of course

$$\Gamma_h^x = \int_{-\pi}^{\pi} e^{ih\theta} f^x(\theta) d\theta$$

If $B(L)$ is a $m \times n$ matrix filter

$$y_t = B(L)x_t,$$

then

$$f^y(\theta) = B(e^{-i\theta}) f^x(\theta) B'(e^{i\theta})$$

I. Stationary processes.

An interesting prediction problem. Back to the band-pass filter

$$y_t = b(L)x_t,$$

where x_t is for example the industrial production index.

Since the filter $b(L)$ is infinite and symmetric, y_t is less and less accurate when we approach the end of the sample. But we may be desperately interested in the end-of-sample values of y_t: what is happening to the economy, meaning what is happening to medium-long run component of the economy? We do not care for short-run oscillations. This is of course a prediction problem. It can be solved either by an ARMA model for x_t, or by predicting x_t by means of other macroeconomic variables. These variables should be leading with respect to x_t, so that they proxy future values of x_t. Etc.

II. The Dynamic Factor Model

This part is based upon:

Forni, M., D. Giannone, M. Lippi and L. Reichlin "Opening the Black Box: Structural Factor Models vs Structural VAR's, Mimeo, 2006.

Forni, M., M. Hallin, M. Lippi and L. Reichlin "The Generalized Dynamic Factor Model: One-Sided Estimation and Forecasting", JASA, 2005.

Forni, M., M. Hallin, M. Lippi and L. Reichlin "The Generalized Dynamic Factor Model: Consistency and Rates" , Journal of Econometrics, 2004, 119, 231-255.

The talk is based upon

Forni M., M. Hallin, M. Lippi and L. Reichlin “Coincident and Leading Indicators for the EURO Area” , The Economic Journal, 2001, 111 (471).

Forni M. and M. Lippi“The Generalized Dynamic Factor Model: Representation Theory”, Econometric Theory, 2001, 17.

Forni M., M. Hallin, M. Lippi and L. Reichlin “The Generalized Dynamic Factor Model: Identification and Estimation”, Review of Economics and Statistics, 2000, 82(4).

Forni, M. and L. Reichlin, “Let’s get real: a factor analytic approach to disaggregated business cycle dynamics”, Review of Economic Studies, 1998.

Forni, M. and M. Lippi, Aggregation and the Microfoundations of Dynamic Macroeconomics, Oxford University Press, 1997.

II. The Dynamic Factor Model

$$\begin{aligned} x_{it} &= \chi_{it} + \xi_{it} = \mathbf{b}_i(L)\mathbf{u}_t + \xi_{it} \\ &= b_{i1}(L)u_{1t} + b_{i2}(L)u_{2t} + \cdots + b_{iq}(L)u_{qt} + \xi_{it}. \end{aligned}$$

– for $i = 1, 2, \ldots, n$; typically n is huge; consistency results are obtained for T, the number of observations for each series, and n, tending to infinity

– q is very small as compared to n in empirical applications

– χ_{it} are the common components; u_{jt} the common shocks; the vector $\mathbf{u}_t$ is an orthonormal white noise

– ξ_{it} are the idiosyncratic components; $\xi_{it} \perp u_{j\tau}$ for all $i,\ j,\ t,\ \tau$

– the filters $b_{ij}(L)$ are square summable

– the vectors $\boldsymbol{\chi}_{nt} = (\chi_{1t} \ \cdots \ \chi_{nt})$ and $\boldsymbol{\xi}_{nt} = (\xi_{1t} \ \cdots \ \xi_{nt})$ are stationary

Common components

$$\chi_{it} = b_{i1}(L)u_{1t} + b_{i2}(L)u_{2t} + \cdots + b_{iq}(L)u_{qt}$$

We assume that the common components are pervasive. This will be formalized in an assumption. For the moment, suppose for example that

$$\chi_{it} = b_i u_t.$$

Then pervasiveness means that

$$\sum_{i=1}^{\infty} b_i^2 = \infty.$$

Common components. Technical Assumption

Let $\Sigma^{\boldsymbol{\chi}}{}_n(\theta)$ be the spectral density of $\boldsymbol{\chi}_{nt}$ and

$$\lambda^{\chi}_{1n}(\theta)\ \lambda^{\chi}_{2n}(\theta)\ \ \cdots\ \lambda^{\chi}_{nn}(\theta)$$

its eigenvalues in decreasing order of magnitude. Then

$$\lambda^{\chi}_{1n}(\theta)\ \lambda^{\chi}_{2n}(\theta)\ \ \cdots\ \lambda^{\chi}_{qn}(\theta)$$

diverge as $n \to \infty$ for almost all $\theta \in [-\pi\ \pi]$.

(Note however that all the eigenvalues

$$\lambda^{\chi}_{q+1,n}(\theta)\ \lambda^{\chi}_{q+2,n}(\theta)\ \ \cdots\ \lambda^{\chi}_{nn}(\theta)$$

vanish for all θ)

In the example $\chi_{it} = b_i u_t$ we have $\lambda_{1n} = \Sigma_{i=1}^{n} b_i^2$.

Idiosyncratic components

Strictly idiosyncratic

$$\xi_{it} \perp \xi_{j\tau} \quad \text{for all } i,\, j,\, i \neq j,\, t,\, \tau.$$

We do not need so much. Let $\Sigma_n^{\xi}(\theta)$ be the spectral density of $\boldsymbol{\xi}_{nt}$. We assume that $\lambda_{1n}^{\xi}(\theta) < \Lambda$ for all n.

Local correlations among idiosyncratic variables are allowed.

Estimation. Example (all you need to know to understand everything)

Assume the elementary example

$$x_{it} = b_i u_t + \xi_{it}.$$

Take the average

$$\frac{1}{n}\sum_{i=1}^{n} x_{it} = \left(\frac{1}{n}\sum_{i=1}^{n} b_i\right) u_t + \frac{1}{n}\sum_{i=1}^{n} \xi_{it}$$

The variances are

$$\mathrm{var}\frac{1}{n}\sum_{i=1}^{n} x_{it} \leq \left(\frac{1}{n}\sum_{i=1}^{n} b_i\right)^2 \sigma_u^2 + \frac{1}{n^2} n \max_i \mathrm{var}\xi_{it} = \bar{b}_n^2 \sigma_u^2 + \frac{1}{n} M$$

Thus the average of the x's converges in mean square to u_t.

Example (continued)

Back to

$$\mathrm{var}\frac{1}{n}\sum_{i=1}^{n} x_{it} \leq \left(\frac{1}{n}\sum_{i=1}^{n} b_i\right)^2 \sigma_u^2 + \frac{1}{n^2} n \max_i \mathrm{var}\xi_{it} = \bar{b}_n^2 \sigma_u^2 + \frac{1}{n} M$$

What if

$$\bar{b}_n \to 0$$

This problem is solved using principal components of the x's

Assumptions so far are:

1. The first q eigenvalues of $\Sigma_n^{\chi}(\theta)$ diverge
2. The first eigenvalue of $\Sigma_n^{\xi}(\theta)$ is bounded

Note that 1 and 2 refer to unobservable variables. The following assumption refers directly to the x's

A. The first q eigenvalues of $\Sigma_n^{x}(\theta)$ diverge, whereas the $(q+1)$-th is bounded.

A. The first q eigenvalues of $\Sigma_n^x(\theta)$ diverge, whereas the $(q+1)$-th is bounded.

Theorem

If Assumption A holds then x_{it} has a factor structure with q common shocks. The converse also holds.

Thus the spectral density of the x's is sufficient to reveal all about the factor structure.

Moreover, The common components χ_{it} are identified. Not the shocks of course.

Estimation

Principal components in the frequency domain and their corresponding filters

$$\mathbf{p}_{1n}(\theta),\ \mathbf{p}_{2n}(\theta),\ \ldots,\ \mathbf{p}_{qn}(\theta)$$

Inverse Fourier transforms

$$\underline{\mathbf{p}}_{1n}(L),\ \underline{\mathbf{p}}_{2n}(L),\ \ldots,\ \underline{\mathbf{p}}_{qn}(L)$$

As $n \longrightarrow \infty$,

$$\underline{\mathbf{p}}_{jn}(L)\mathbf{x}_{nt}$$

converges to the space spanned by $\mathbf{u}_t$.

Estimation (continued)

The common components are estimated by projecting the x's on the principal components

$$\hat{\chi}_{it}^{(n)} = \mathrm{proj}(x_{it} | \text{space spanned by } \underline{\mathbf{p}}_{jn}(L)\mathbf{x}_{n\tau}, \text{ for } j = 1, \dots, q,\ \tau \in \mathbb{Z})$$

The following is the implicit estimate of the spectral density matrix of the common components:

$$\hat{\mathbf{\Sigma}}_n^{\chi}(\theta) = \lambda_{1n}^x(\theta)\overline{\mathbf{p}}'_{1n}(\theta)\mathbf{p}_{1n}(\theta) + \lambda_{2n}^x(\theta)\overline{\mathbf{p}}'_{2n}(\theta)\mathbf{p}_{2n}(\theta) + \cdots + \lambda_{qn}^x(\theta)\overline{\mathbf{p}}'_{qn}(\theta)\mathbf{p}_{qn}(\theta)$$

Main problem with frequency domain approach

Some more detail on the eigenvectors and their Fourier series

$$\mathbf{p}_{1n}(\theta) = \sum_{k=-\infty}^{\infty} a_{1n,k} e^{-ik\theta}, \quad a_{1n,k} = \frac{1}{2\pi} \int_{-\pi}^{\pi} e^{ik\theta} \mathbf{p}_{1n}(\theta) d\theta$$

$$\underline{\mathbf{p}}_{1n}(L) = \sum_{k=-\infty}^{\infty} a_{1n,k} L^k$$

In general the filter $\underline{\mathbf{p}}_{1n}(L)$ is two sided. The estimates are very good within the sample, but not suited for forecasting.

The static method

Rewrite the model

$$\mathbf{x}_{nt} = \mathbf{b}_n(L)\mathbf{u}_t + \boldsymbol{\xi}_{nt}.$$

Suppose that

$$\mathbf{b}_n(L) = \mathbf{b}_{0n} + \mathbf{b}_{1n}L + \cdots + \mathbf{b}_{sn}L^s,$$

the matrix $\mathbf{b}_{nj}$ being nested in $\mathbf{b}_{mj}$ for $m > n$. A finite moving average.

Then, defining

$$\mathbf{f}_t = (\mathbf{u}_t' \; \mathbf{u}_{t-1}' \; \cdots \; \mathbf{u}_s')', \quad \mathbf{B}_n = (\mathbf{b}_{0n} \; \mathbf{b}_{1n} \; \cdots \mathbf{b}_{sn}),$$

the model can be written in static form

$$\mathbf{x}_{nt} = \mathbf{B}_n\mathbf{f}_t + \boldsymbol{\xi}_{nt}$$

More in general, the static method applies when the space spanned by the variables χ_{it} is finite dimensional.

The static method (continued)

$$\mathbf{x}_{nt} = \mathbf{B}_n \mathbf{f}_t + \boldsymbol{\xi}_{nt}$$

It is important to distinguish between the static method and the static model.

If the model is static then $\mathbf{f}_t$ is a white noise. Otherwise the spectral density of $\mathbf{f}_t$ is not trivial.

The static method. Estimation.

Estimation based on the static method employs the principal components of the x's in the time domain. More precisely, consider the first r eigenvalues and corresponding eigenvectors of the variance-covariance matrix of the x's, where r is the dimension of $\mathbf{f}_t$.

$$\mathbf{P}_{1n},\ \mathbf{P}_{2n},\ \ldots,\ \mathbf{P}_{rn}$$

The principal components

$$\mathbf{P}_{jn}\mathbf{x}_{nt}$$

for $j = 1, \ldots, r$, converge to the space spanned by the factors $\mathbf{f}_t$.

The static method. Estimation.

Moreover, the common components are estimated as

$$\chi_{it}^{(n)} = \mathrm{proj}(x_{it}|\text{space spanned by } \mathbf{P}_{jn}\mathbf{x}_{nt}, \text{ for } j = 1, 2, \dots, r)$$

and the estimated variance-covariance matrix of the common components is

$$\hat{\mathbf{\Gamma}}_n^{\chi} = \mu_{1n}^x \mathbf{P}'_{1n}\mathbf{P}_{1n} + \mu_{2n}^x \mathbf{P}'_{2n}\mathbf{P}_{2n} + \cdots + \mu_{rn}^x \mathbf{P}'_{rn}\mathbf{P}_{rn}$$

The static method (continued)

Example

$$x_{it} = b_{i0}u_t + b_{i1}u_{t-1} + \xi_{it}, \quad q = 1, \quad r = 2$$

In this case, the dynamic method employs the first eigenvector of the spectral density. The static method employs the first two eigenvectors of the variance-covariance matrix.

Literature

Static method

Stock, J.H. and M.W. Watson (2002a). Forecasting using principal components from a large number of predictors, *Journal of the American Statistical Association* **97**, 1167-79.

Stock, J.H. and M.W. Watson (2002b). Macroeconomic forecasting using diffusion indexes. *Journal of Business and Economic Statistics* **20**, 147-162.

Kapetanios, G. and M. Marcellino (2005). A Comparison of Estimation Methods for Dynamic Factor Models of Large Dimension, Mimeo.

Determining the number of factors

Bai, J. and S. Ng (2002). Determining the number of factors in approximate factor models. *Econometrica* **70**, 191-221.

Hallin, M. and R. Liska (2006) Dynamic factor Analysis: The number of Factors and Related Issues, ULB Working Paper.

The two-step method

Forni, M., M. Hallin, M. Lippi and L. Reichlin "The Generalized Dynamic Factor Model: One-Sided Estimation and Forecasting", JASA, 2005.

Same finite-dimension assumptions as above. Consider again the static projection

$$\chi_{it}^{(n)} = \mathrm{proj}(x_{it}|\text{space spanned by } \mathbf{P}_{jn}\mathbf{x}_{nt}, \text{ for } j = 1, 2, \ldots, r)$$

We need the covariances between x_{it} and the estimated factors $\mathbf{P}_{jn}\mathbf{x}_{nt}$. These are trivially obtained from the covariance matrix of the x's.

The two-step method employs the covariances of the common components that result from $\hat{\mathbf{\Sigma}}_n^{\chi}(\theta)$, i.e. the spectral density matrix of the common components.

The two-step method (continued)

Back to the dynamic method:

$$\hat{\mathbf{\Sigma}}^{\chi}_{n}(\theta) = \lambda^{x}_{1n}(\theta)\overline{\mathbf{p}}'_{1n}(\theta)\mathbf{p}_{1n}(\theta)+\lambda^{x}_{n2}(\theta)\overline{\mathbf{p}}'_{2n}(\theta)\mathbf{p}_{2n}(\theta)+\cdots+\lambda^{x}_{qn}(\theta)\overline{\mathbf{p}}'_{qn}(\theta)\mathbf{p}_{qn}(\theta)$$

The covariances of the common components result from

$$\hat{\mathbf{\Gamma}}^{\chi}_{n,k} = \int_{-\pi}^{\pi} e^{ik\theta}\hat{\mathbf{\Sigma}}^{\chi}_{n}(\theta)d\theta.$$

The covariances of the idiosyncratic components result as differences $\hat{\mathbf{\Gamma}}^{\xi}_{n,k} = \hat{\mathbf{\Gamma}}^{x}_{n,k} - \hat{\mathbf{\Gamma}}^{\chi}_{n,k}$.

Now construct generalized principal components

$$\mathbf{Q}_{1n},\ \mathbf{Q}_{2n},\ \ldots,\ \mathbf{Q}_{rn}$$

where $\mathbf{Q}_{kn}$ is obtained as the solution of

$$\max_{\mathbf{a}}\ \mathbf{a}\hat{\mathbf{\Gamma}}^{\chi}_{n}\mathbf{a}' \ \text{ subject to }\ a\hat{\mathbf{\Gamma}}^{\xi}_{n}\mathbf{a}' = 1 \text{ and } \mathbf{a}\hat{\mathbf{\Gamma}}^{\xi}_{n}\mathbf{Q}_{hn} = 0 \text{ for } h < k$$

The two-step method (continued)

Once the generalized principal components are obtained we can project

$$\chi_{it}^{(n)} = \mathrm{proj}(x_{it}|\text{space spanned by } \mathbf{Q}_{jn}\mathbf{x}_{nt}, \text{ for } j = 1, 2, \ldots, r)$$

This projection can be computes by replacing the covariances between the x's with the estimated covariance between the χ's, which proves to be a consideable advantage.

Forecast, both with the static and the two-step method is obtained by projecting $x_{i,t+h}$ on the space spanned by the static factors at time t.

Dynamic vs static method

Specify the example

$$
\begin{aligned}
x_{1t} &= u_t + \xi_{1t} \\
x_{2t} &= u_{t-1} + \xi_{2t} \\
&\vdots
\end{aligned}
$$

With dynamic method we use the filter

$$(1 \ F \ 1 \ F \ \cdots)$$

With the static method we use the two averages

$$(1 \ 0 \ 1 \ 0 \ \cdots) \qquad (0 \ 1 \ 0 \ 1 \ \cdots)$$

Different efficiency.

Dynamic vs static method (continued)

The static method can be applied to models that are more complicated than the moving average. However, we should keep in mind that a model as simple as

$$x_{it} = \frac{1}{1 - \alpha_i L} u_t + \xi_{it}$$

cannot be put in static form. This motivates an attempt to determine a one-sided representation of the common components in the general case. This will be the last part of the talk.

Forecast. Empirical results using the static and the two-step method.

Boivin, J. ans S. Ng (2005) Understanding and Comparing Factor-Based Forecasts, International Journal of Central Banking (forthcoming)

D'Agostino, A. and D. Giannone (2005) Comparing alternative predictors based on large-panel dynamic factor models, ECARES, ULB, Working Paper.

Reijer, A. den (2005) Forecasting Dutch GDP using Large Scale Factor Models, De Nederlandsche Bank, Research Division.

One-sided representation in the general case.

We start with a consistent estimate of the spectral density matrix of the common components

$$\hat{\mathbf{\Sigma}}_n^\chi(\theta)$$

This is not a parametric estimation. We can have it for each point in $[-\pi \ \pi]$.

Suppose you want to obtain the Wold representation. You may think of factorizing the spectral density following Wiener and Masani, but that method is confined to non-singular spectral densities, whereas our case is one of extreme singularity: a big dimension n with a small rank q.

Or, you may think of using the rational-spectrum factorization technique. But you should first fit a rational spectrum to $\hat{\mathbf{\Sigma}}_n^\chi(\theta)$.

One-sided representation

Alternatively, you may think of approaching the problem by estimating a VAR

$$\boldsymbol{\chi}_{nt} = \mathbf{A}_1 \boldsymbol{\chi}_{n,t-1} + \mathbf{A}_2 \boldsymbol{\chi}_{n,t-2} + \cdots + \mathbf{w}_t$$

All the covariances nedeed to compute the projection can be obtained from $\hat{\boldsymbol{\Sigma}}_n^{\chi}$. However, we must insist that $\boldsymbol{\chi}_{nt}$ is a highly singular vector, so that the projection above is typically not unique. For example, let

$$\chi_{it} = u_t \text{ for } i \text{ odd}, = u_{t-1} \text{ for } i \text{ even}$$

(we have already seen this example). In this case

$$\begin{aligned}
\chi_{1t} &= u_t \\
\chi_{2t} &= \chi_{1,t-1} \\
\chi_{3t} &= u_t \\
\chi_{4t} &= \chi_{1,t-1} \text{ but also } = \chi_{3,t-1}
\end{aligned}$$

One-sided representation. A natural assumption

We need a method to select one among all the autoregressive representations. Some considerations preliminary to the introducion of an assumption. Suppose for the moment that $q = 1$. Consider the projection

$$\boldsymbol{\chi}_{nt} = \text{proj}(\boldsymbol{\chi}_{nt} | \boldsymbol{\chi}_{n,t-1}, \ \boldsymbol{\chi}_{n,t-2}, \ \dots) + \mathbf{w}_{nt}, \qquad \mathbf{w}_{nt} = (c_1 \ c_2 \ \cdots \ c_n)' u_t$$

The Wold representation of $\boldsymbol{\chi}_{nt}$ is

$$\boldsymbol{\chi}_{nt} = \mathbf{b}_n(L) u_t$$

with u_t identified up to a constant multiplicative term. Consider the rational-spectrum case

$$\begin{aligned} \chi_{1t} &= \frac{\alpha_1(L)}{\beta_1(L)} u_t \\ \chi_{2t} &= \frac{\alpha_2(L)}{\beta_2(L)} u_t \\ &\vdots \end{aligned}$$

One-sided representation. Assumption

(continued) Consider the rational-spectrum case

$$\begin{aligned} \chi_{1t} &= \frac{\alpha_1(L)}{\beta_1(L)} u_t \\ \chi_{2t} &= \frac{\alpha_2(L)}{\beta_2(L)} u_t \\ &\vdots \end{aligned}$$

If $\alpha_1(L)$ and $\alpha_2(L)$ have no root in common within the unit disk, then u_t is fundamental for the vector $(\chi_{1t}\ \chi_{2t})$. This is tantamount to saying that if we take the projection

$$\chi_{jt} = \text{proj}(\chi_{jt} | \chi_{1,t-1},\ \chi_{2,t-1},\ \chi_{1,t-2},\ \chi_{2,t-2},\ \ldots) + z_{jt}, \quad \text{for } j = 1, 2$$

we have $z_{jt} = d_j u_t$, that is

The space spanned by past values of χ_{1t} and χ_{2t} coincides with the space spanned by past values of all the χ's.

Assumption F. The space spanned by past values of χ_{it} and χ_{jt}, any $i \neq j$, coincides with the space spanned by past values of all the χ's.

The typical example is the MA(1) model

$$\chi_{jt} = b_{j0}u_t + b_{j1}u_{t-1},$$

under reasonable heterogeneity of the couples $(b_{j0}\ b_{j1})$.

Note that we are not assuming fundamentalness of any single-component representation

$$\chi_{jt} = b_{j0}u_t + b_{j1}u_{t-1} + \cdots$$

We only require that as soon as the number of the selected χ's exceeds the dimension q, which is 1 for the moment, the heterogeneity of the coefficients allows recovering u_t by means of present and past of the selected χ's.

One-sided representation. We obtain a piecewise VAR:

$$\begin{pmatrix} c_{11}(L) & c_{12}(L) & 0 & 0 & \cdots & 0 \\ c_{21}(L) & c_{22}(L) & 0 & 0 & \cdots & 0 \\ 0 & 0 & c_{33}(L) & c_{34}(L) & \cdots & 0 \\ 0 & 0 & c_{43}(L) & c_{44}(L) & \cdots & 0 \\ \vdots & & & & & \\ 0 & 0 & \cdots & \cdots & c_{n-1,n-1}(L) & c_{n-1,n}(L) \\ 0 & 0 & \cdots & \cdots & c_{n,n-1}(L) & c_{nn}(L) \end{pmatrix} \begin{pmatrix} \chi_{1t} \\ \chi_{2t} \\ \chi_{3t} \\ \vdots \\ \chi_{nt} \end{pmatrix} = \begin{pmatrix} w_{1t} \\ w_{2t} \\ w_{3t} \\ \vdots \\ w_{nt} \end{pmatrix}$$

with $w_{it} = c_i u_t$.

Obviously the autoregressive representation above is not unique.

However, the resulting moving average representation is unique, up to a multiplicative constant. Denoting the autoregressive representation by $\mathbf{C}_n \boldsymbol{\chi}_{nt} = \mathbf{c}_n' u_t$,

$$\boldsymbol{\chi}_{nt} = \mathbf{c}_n' u_t + \mathbf{C}_n \mathbf{c}_n' u_{t-1} + \mathbf{C}_n^2 \mathbf{c}_n' u_{t-2} + \cdots = \mathbf{D}_{0n} u_t + \mathbf{D}_{1n} u_{t-1} + \mathbf{D}_{2n} u_{t-2} + \cdots$$

One-sided Representation

Summing up, under Assumption F:

1. Estimation of the spectral density $\hat{\mathbf{\Sigma}}_n^{\chi}$.

2. Fitting 2-dimensional VAR's to the couples (1,2), (3,4), etc., so obtaining $\mathbf{C}_n \boldsymbol{\chi}_{nt} = \mathbf{c}_n' u_t$, and therefore

$$\boldsymbol{\chi}_{nt} = \mathbf{c}_n' u_t + \mathbf{C}_n \mathbf{c}_n' u_{t-1} + \mathbf{C}_n^2 \mathbf{c}_n' u_{t-2} + \cdots = \mathbf{D}_{0n} u_t + \mathbf{D}_{1n} u_{t-1} + \mathbf{D}_{2n} u_{t-2} + \cdots$$

3. Note that the χ's are not observed, so the fitting above means fitting $\mathbf{C}_n$ and $\mathbf{c}_n$ using the covariances of the χ's obtained from the estimated $\hat{\mathbf{\Sigma}}_n^{\chi}$. We do not estimate u_t here.

One-sided Representation

4. Lastly, since $\chi_{nt} = \mathbf{x}_{nt} - \boldsymbol{\xi}_{nt}$, we have

$$\mathbf{C}_n \mathbf{x}_{nt} = \mathbf{c}_n' u_t + \mathbf{C}_n \boldsymbol{\xi}_{nt}.$$

An estimate of u_t is obtained by

$$\mathbf{c}_n \mathbf{C}_n \mathbf{x}_{nt} = (\mathbf{c}_n \mathbf{c}_n') u_t + \mathbf{c}_n \mathbf{C}_n \boldsymbol{\xi}_{nt}.$$

One-sided Representation. Generalization to the case $q > 1$

Assumption F. The space spanned by past values of $\chi_{i_1,t},\ \chi_{i_2,t},\ \ldots,\ \chi_{i_{q+1},t}$ coincides with the space spanned by past values of all the χ's.

The motivation goes as above. If

$$\boldsymbol{\chi}_{nt} = \mathbf{B}_n(L)\mathbf{c}_n'\mathbf{u}_t,$$

where $\mathbf{c}_n$ is $q \times n$, is the Wold representation of $\boldsymbol{\chi}_{nt}$, then considering the vector $\mathbf{A}_t$, which contain, say, the first $q+1$ of the χ's, the vector $\mathbf{u}_t$ is not fundamental for $\mathbf{A}_t$ only if the $(q+1) \times q$ matrix

$$(b_{ij}(L))_{i=1,\ldots,q+1;j=1,\ldots,q}$$

has rank less than q somewhere within the unit disk. This means that all the $q \times q$ subamatrices should be singular for at the same z and $|z| < 1$.

One-sided Representation. Generalization to the case $q > 1$

Generalizing the remaining steps is fairly obvious.

1. Estimation of $\hat{\mathbf{\Sigma}}_n^{\chi}$.

2. Fitting $(q+1)$-dimensional VAR's to the $(q+1)$-tuples $(1,\ 2,\ \ldots,\ q+1)$, etc., thus obtaining $\mathbf{C}_n \boldsymbol{\chi}_{nt} = \mathbf{c}_n' \mathbf{u}_t$, and therefore

$$\boldsymbol{\chi}_{nt} = \mathbf{c}_n' \mathbf{u}_t + \mathbf{C}_n \mathbf{c}_n' \mathbf{u}_{t-1} + \mathbf{C}_n^2 \mathbf{c}_n' \mathbf{u}_{t-2} + \cdots = \mathbf{D}_{0n} \mathbf{u}_t + \mathbf{D}_{1n} \mathbf{u}_{t-1} + \mathbf{D}_{2n} \mathbf{u}_{t-2} + \cdots$$

3. As above.

4. Lastly, since $\chi_{nt} = \mathbf{x}_{nt} - \boldsymbol{\xi}_{nt}$, we have

$$\mathbf{C}_n \mathbf{x}_{nt} = \mathbf{c}_n' \mathbf{u}_t + \mathbf{C}_n \boldsymbol{\xi}_{nt}.$$

An estimate of u_t is obtained by

$$\mathbf{c}_n \mathbf{C}_n \mathbf{x}_{nt} = (\mathbf{c}_n \mathbf{c}_n') \mathbf{u}_t + \mathbf{c}_n \mathbf{C}_n \boldsymbol{\xi}_{nt}.$$

One-sided Representation.

The procedure described above has been applied to simulated data.

-$q = 1, 2$.

-Rational case, low order.

-The estimates of the common component are worse than those obtained using the dynamic method (two-sided filters), but better than those obtained with the static method.