Factor models

Fabio Canova
Norwegian Business School and CEPR
April 2023

Outline

- Introduction.
- Static and dynamic principal components.
- Static and dynamic factor models.
- Adding normality.
- Large data sets.
- Generalizations and structural factor models.
- FAVARs; Panel VARs.

References

Ahmadi, P. and Uhlig, H (2015) Sign Restrictions in a Bayesian FAVAR, with an application to Monetary Policy Shocks. NBER wp 21738.

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

Bai, J. and Ng, S. (2007) Determining the number of primitive shocks in factor models, Journal of Business and Economic Statistics, 25, 52-60.

Bai, J., Li, K. and L. Lu (2016) Estimation and Inference of FAVAR Models, Journal of Business Economic Statistics, 34(4), 620-641.

Bernanke, B. and Boivin, J.(2003) Monetary Policy in a data rich environment, Journal of Monetary Economics, 50, 525-546.

Bernanke, B., Boivin, J., and Elias, P.(2005)Measuring the effects of monetary policy: a factor augmented VAR approach, Quarterly Journal of Economics, 120, 387-422.

Canova, F. and Ciccarelli, M. (2009) Estimating multi-country VAR models, International Economic Review, 50, 929-960.

Doz, C. Fuleky, P. (2019) Dynamic factor models, Paris School of Economics, working paper 2019-45

Forni, M., Hallin, M., Lippi, M., Reichlin, L., 2000. The generalized dynamic factor model: identification and estimation. Review of Economics and Statistics 82, 540–554.

Forni, M. and Gambetti, L. (2010) The Dynamic Effects of Monetary Policy: A structural factor model approach, Journal of Monetary Economics, 57, 203-216.

Stock, J. and Watson, M. (1999) Forecasting Inflation, Journal of Monetary Economics, 44, 293-335.

Stock, J. and Watson, M. (2002) Macroeconomic Forecasting using Diffusion Indices, Journal of Business and Economic Statistics, 20, 147-162.

Stock J, and M. Watson. (2011). Dynamic Factor Models. In Oxford Handbook on Economic Forecasting, eds. Michael P. Clements and David F. Hendry. Oxford: Oxford University Press.

Stock, J. and M. Watson (2016). Dynamic Factor Models, Factor-Augmented Vector Autoregressions, and Structural Vector Autoregressions in Macroeconomics. Handbook of Macroeconomics, volume 2, 415-525.

1 Preliminaries

• Problem with VARs is that they have less variables than one would like to consider: trade-off dimensionality vs. estimability. One alternative: factor models.

• Idea: a few factors explain a large portion of the variability of macro and financial variables.

• Scope of factor models: given a set of cross sectionally correlated time series, find a few indicators which are responsible for their comovements, i.e. try to explain the data with a smaller set of explanatory variables.

• Observable vs. unobservable factors models (explanatory variables are observable or not).

Example 1 *i) From psychology: the IQ from a set of test scores.*

ii) From economics: a business cycle indicator from a vector of macroeconomic time series.

iii) From finance: the market portfolio in a CAPM model

Principal Components

• A technique related to factor models.

• Principal Components (PC) are linear combinations of observable variables maximizing the explained portion of the variance of the observables.

• PCs provide a way to reduce the dimensionality of the observables.

• PCs can be computed for every data set. There are assumptions that need to be satisfied to run factor models.

• No statistical model is needed to setup PCs.

2 Static Principal Components

• Let $X \sim (0, \Sigma)$ be a $p \times 1$ vector of iid random variables.

• Want to find the linear combination $\beta' X$, where β is a $p \times 1$ vector, such that $var(\beta' X)$ is maximized, i.e.

$$\max_{\beta}[var(\beta' X)] = \max_{\beta} \beta' \mathbf{\Sigma} \beta \quad \text{subject to } \beta' \beta = 1 \tag{1}$$

• If $\beta' \beta = 1$ is not used, the solution is $\beta = \infty$ (uninteresting).

• The Lagrangian is $L_1 = \beta' \mathbf{\Sigma} \beta - \lambda(\beta' \beta - 1)$.

• The first and second order conditions are

$$\frac{\partial L_1}{\partial \beta} = 2\boldsymbol{\Sigma}\beta - 2\lambda\beta \;=\; 0 \tag{2}$$

$$\frac{\partial^2 L_1}{\partial \beta \partial \beta} = -2\lambda \;<\; 0 \tag{3}$$

so that the solution must satisfies $(\boldsymbol{\Sigma} - \lambda I_p)\beta = 0$.

• (2) is a system of p linear equations in $p+1$ unknowns (β plus λ).

• Since it is a "homogenous system" (i.e. of the form $AX = 0$) and β can not be a zero vector, the FOC's have a non-trivial solution if and only if $\det(\boldsymbol{\Sigma} - \lambda I_p) = 0$.

Procedure to find PC:

- Find the p eigenvalues λ of $\mathbf{\Sigma}$ and order them decreasingly.
- Using $(\mathbf{\Sigma} - \lambda I_p)\beta = 0$ find the associated eigenvectors $(\beta^1, \ldots, \beta^p)$.
- Premultiplying the FOC by β' we get $\beta'\mathbf{\Sigma}\beta = \lambda\beta'\beta = \lambda$.
- Since $\beta'\mathbf{\Sigma}\beta$ is the variance of $\beta' X$:

i) The β that maximizes the objective function is the one associated with the largest λ, i.e. choose $\beta = \beta^1$ and PC_1 is $(\beta^1)'X$.

ii) $var((\beta^1)'X) = \lambda_1$, the first eigenvalue is the variance explained by the first PC.

• Suppose we want to find another linear combination of αX such that:

- it maximizes the variance of X_t explained.

- it is orthogonal to the first PC, i.e.$E(\alpha' X PC_1') = 0$.

• The latter condition implies $E(\alpha' X X' \beta^1) = \alpha' E(X X')\beta^1 = \alpha' \mathbf{\Sigma} \beta^1 = \alpha' \lambda \beta^1 = \lambda \alpha' \beta^1 = 0$. Since $\lambda_1 \neq 0$ then $\alpha' \beta^1 = 0$, i.e. α must be orthogonal to β^1.

- Lagrangian: $L_2 = \alpha' \mathbf{\Sigma} \alpha - \lambda(\alpha'\alpha - 1) - 2\eta(\alpha' \mathbf{\Sigma} \beta^1)$.

- FOC: $2\mathbf{\Sigma}\alpha - 2\lambda\alpha - 2\eta\mathbf{\Sigma}\beta^1 = 0$.

- Premultiplying by β^1 we have:

$$\beta^1 \mathbf{\Sigma} \alpha - \beta^1 \alpha - 2\eta \beta^1 \mathbf{\Sigma} \beta^1 = 0 - 0 - 2\eta\lambda_1 = 0 \tag{4}$$

where the first two zeros comes from orthogonality. Then, the maximum is obtained when $\eta = 0$ i.e. the second constraint is not binding. Thus the FOC are: $(\mathbf{\Sigma} - \lambda I_p)\alpha = 0$.

- Same problem; same solution: choose α to be the eigenvector β^2; the variance explained by the second PC is λ_2, the second largest eigenvalue

- Conclusions:

1) The first $k < p$ principal components of X are $[(\beta^i)'X, (\beta^2)'X....(\beta^k)'X]$ and their variance is $[\lambda_i, \lambda_2, ...\lambda_k]$.

2) The proportion of the variance explained by the first k PC is $\frac{\sum_{i=1}^k \lambda_i}{trace(\mathbf{\Sigma})}$.

- Can see this by noting that:

i) $\mathbf{\Sigma} = B\Lambda B'$ where B is the matrix of eigenvectors, $\Lambda = diag(\lambda_i)$ is a matrix with eigenvalues.

ii) Define B_k the matrix with the first k eigenvectors and Λ_k the matrix with the first k eigenvalues. Then $\mathbf{\Sigma} = B_k\Lambda_k B_k' + B_{p-k}\Lambda_{p-k}B_{p-k}'$.

iii) Combining i) and ii) we have $\mathbf{\Sigma}_k = B_k\Lambda_k B_k' = \sum_i^k \lambda_i$.

- Easy to compute PCs: just find eigenvectors/eigenvalues of $X'X$, take as PC the first k eigenvectors β multiplied by X.

- PCs depend on the units in which variables are measured (it is better to standardize if variables are measured in different units).

- If the elements of X are nearly orthogonal (Σ close to diagonal) each PC explains about $1/p$ of the total variance.

- PC analysis produce a reduced rank representation of the data.

3 Dynamic Principal Components

• What if X is not iid? Since static PCs do not take into account autocovariances, they will not maximize the portion of the variance explained.

• What can we do in this case?

a) Filter X until its autocovariances are all zero (i.e. run a VAR(q) on X). Apply static PC to the residuals.

b) Use dynamic principal components (DPC): find the β that maximize the spectral density (a transformation of the autocovariance function).

- Mechanics:

i) Compute spectral density of X at Fourier frequencies ω (the components of the spectral density are orthogonal so we are back to a situation similar to iid X).

ii) Compute PC at each ω. We can do this since, given ω, the spectral density is like a covariance matrix. Hence, find eigenvalues and eigenvectors at each frequency ω.

Let the first principal component at frequency ω be $DPC_1(\omega) = r_1(\omega)x(\omega)$, where $x(\omega)$ is the Fourier transform of X. Its spectral density is $f_{DPC_1}(\omega) = r_1(\omega)f_x(\omega)r_1(\omega)$, where $r_1(\omega)$ is the first dynamic eigenvector at ω and $f_x(\omega)$ the spectral density of X at frequency ω.

Note that $f_{DPC_1}(\omega) = \lambda_1(\omega)$, i.e. the spectral density of the first dynamic PC is the first eigenvalue f the spectral matrix at ω.

• The time domain representation of the first dynamic principal component is: $DPC_1 = r_1(\ell)'X_t$ where $r_{1i}(h) = \int_{-\pi}^{\pi} r_{1i}(\omega)e^{i\omega h}d\omega \; i = 1, \ldots, n, h = -w, , \ldots, 0, , \ldots, w$.

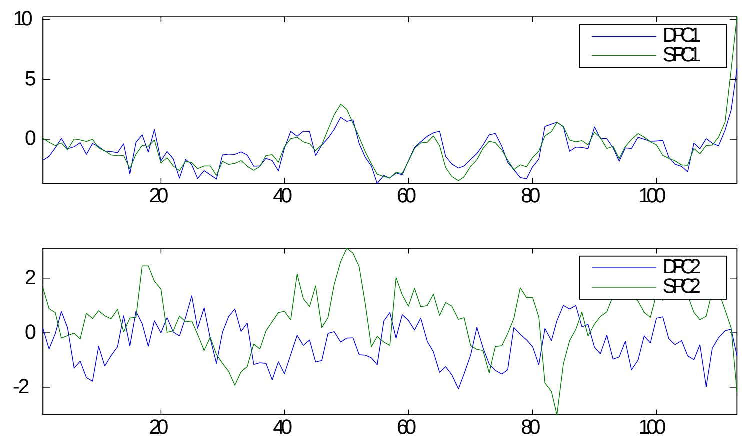
• $r_1(\ell)$ is a two-sided polynomial of infinite length. Problem in forecasting and policy exercises! Possibility of making $r_1(\ell)$ one-sided.

• DPC_j and $DPC_{j'}$ are orthogonal at all leads and lags, $j \neq j'$.

• What if we are interested in the PCs that maximize the explained portion of X_t in a band of frequencies? i.e., what if we are interested in finding the principal components that maximize the variance of X in at business cycle frequencies.

• Since the elements of the spectral density at Fourier frequencies are orthogonal, DPC_1 in a band of frequencies is the sum of the first dynamic principal components for the frequencies of the band.

Example 2 *Use yearly growth rate of IP from 1980:1 to 2008:4 for France, Turkey, Italy, Portugal, Spain, Cyprus and Israel. The first static PC explains 37 percent of the covariations, the second 22 percent and the third 10 percent. The first static and dynamic PC have similar movements. Not the case of the second PCs - they load on different frequencies.*



4 Static Factor Models

$$X = \Lambda F + U \tag{5}$$

where $E(U) = E(F) = 0$, $E(UU') = \Psi$ (a diagonal matrix), $E(FF') = \Phi$, $E(FU') = 0$.

• Λ are the factor loadings $(n \times m)$, F are common factors $(m \times 1)$, $m < n$; ΛF is the common component, U $(n \times 1)$ the idiosyncratic component.

• All covariations in X are due to the common component.

• F could be observable or latent.

4.1 Observable factors

• Often in finance want to find the "factors" that drive stock returns.

• Look for macroeconomic factors: real vs. nominal, domestic vs. international.

• Use proxies. Estimate the loadings on the factors. See how much they explain of the endogenous variables.

• Careful: proxies generate **error-in-variables/ generated regressors**.Need to adjust standard errors of estimates.

• A panel VAR with some reasonable (shrinkage) restrictions induce observable factor models (see later).

4.2 Unobservable factors

• There is an identification problem (both Λ and F are unknown). Consider a non-singular $C \neq I$ of dimension $m \times m$. Then

$$\begin{aligned} X &= \Lambda F + U = \Lambda C C^{-1} F + U \\ &= \Lambda^* F^* + U \end{aligned} \tag{6}$$

(5) and (6) are observational equivalent!! Restrict Λ, Ψ, Φ to avoid this.

• Order condition: want the number of free parameters in the covariance of X_t to be greater or equal to the number of parameters to be estimated. Notice:

$$\frac{1}{T}\sum_t (X_t' X_t) \equiv S \xrightarrow{P} \Sigma \equiv \Lambda \Phi \Lambda' + \Psi \tag{7}$$

where S empirical covariance matrix and Σ theoretical covariance matrix of X and $\xrightarrow{P}$ indicates convergence in probability.

• Number of estimated parameters (the RHS of (7)): mn in Λ, n in Ψ, $m(m+1)/2$ is Φ (symmetric). The total is $(mn + n + m(m+1)/2)$.

• Number of free parameters in the data (the LHS of (7)): $n(n+1)/2$.

• If we had m^2 (additional) constraints

$d = n(n+1)/2 + m^2 - (mn + n + m(m+1)/2) = 0.5((n-m)^2 - m - n)$

There are three possibilities:

- $d < 0$ (no identification is achieved, infinite number of solutions).

- $d = 0$ (just identification, one solution).

- $d > 0$ (over-identification, no exact solution).

• The interesting case is $d > 0$ since here the model provides a simplified representation of the data, i.e. $\Lambda\Phi\Lambda' + \Psi \approx S$.

- Where do we get the m^2 restrictions?. Typical scheme:

i) E(FF')=$\Phi = I$ ($m(m+1)/2$ restrictions), i.e. sources of comovements are orthogonal among each other.

ii) $\Gamma = \Lambda'\Psi^{-1}\Lambda$ is diagonal ($m(m-1)/2$ restrictions), i.e. the columns of Λ are orthogonal in the metric of Ψ^{-1} (similar in spirit to Choleski decomposition in VARs). See later(adding normal error) for further intuition.

4.2.1 Estimation

- Use Principal Factors (PF), a technique related to PC.

Algorithm 4.1 *Suppose we have an initial* Ψ^0.

1) Compute $\Sigma - \Psi^0$. *Decompose* $\Sigma - \Psi^0 = ABA'$, *where* A *are the eigenvectors and* B *the matrix of eigenvalues. Take the* k *largest eigenvalues and the associated eigenvectors and construct* $\Lambda^1 = AB^{0.5}$.

2) Compute $\Psi^1 = \Sigma - \Lambda^1\Phi(\Lambda^1)' = \Sigma - \Lambda^1(\Lambda^1)'$ *since* Φ *is the identity matrix.*

3) Repeat steps 2)-3) until convergence, i.e. until $||\Psi^l - \Psi^{l-1}|| < \iota$.

- After the iterations end, compute the factors as $F_t = \Lambda X_t$.

• How do we choose Ψ^0?

i) Use $\Psi^0 = I$

ii) Use $\Psi_i^0 = 1 - max_{j \neq i} |corr(X_i, X_j)|$.

• As in PCs, results are sensitive to the scaling of variables, so need to standardize (use correlations not covariances),

• Finding PF is equivalent to minimizing: trace$(S - \Sigma)^2$. Other possible criterion functions: trace$((S - \Sigma)\Sigma^{-1})^2$, or trace$((S - \Sigma)\Psi^{-1})^2$.

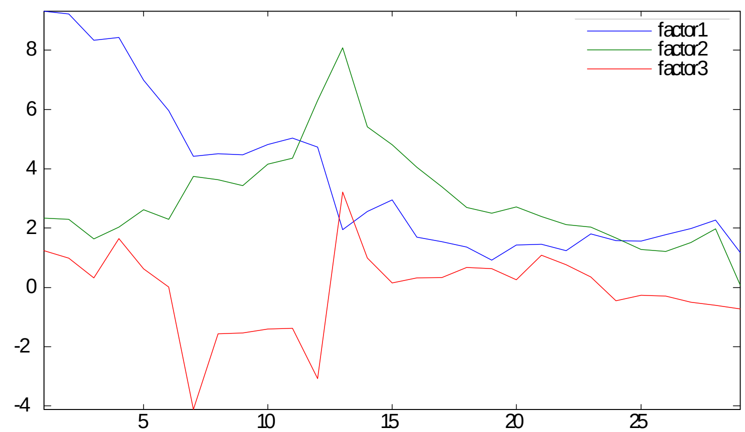
Comparison PC-PF:

• If the variables have the same scale and $\Psi = 0$, PC=PF. If $\Psi \neq 0$, idiosyncratic component contaminates all PFs.

• In PCs, the leftover is non-diagonal and has rank $n-k$. In PF leftover (i.e. U_t) has diagonal covariance matrix with rank n.

• If $\Psi = \sigma^2 I$, $\Lambda_{PF} = \Lambda_{PC} G$, where G is a non-singular matrix.

Example 3 *Use annual inflation rates of 17 mediterranean countries constructed using GDP deflators from 1980 to 2009. Estimate three factors and use PF analysis to get the loadings. The inflation rates are quite idiosyncratic: the three factors explain only 35, 10 and 5 percent on the standardized variations. Below the time series of the factors.*



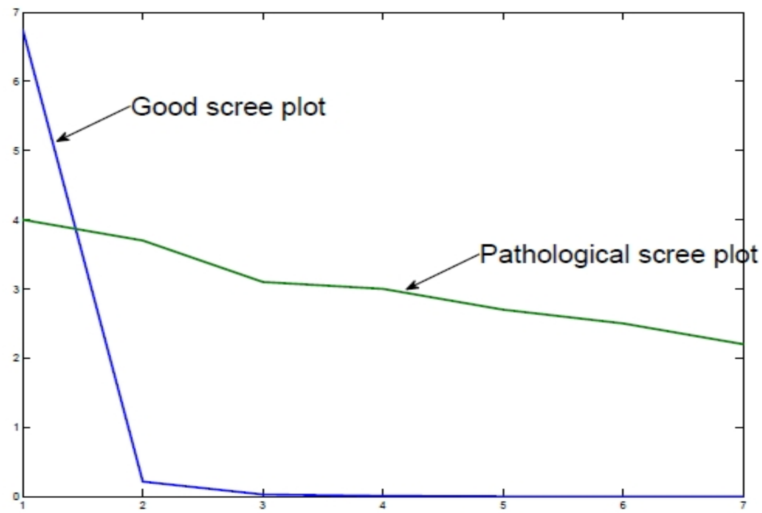
4.3 Determining the number of static factors

1) Informal procedure. Since eigenvalues in decreasing order, choose a fraction of total variance you want to explain and include as many factors to reach that level (recall fraction of variance explained by first k factors is $\frac{\sum_i \lambda_i}{trace(\Sigma)}$.

• No general accepted level for how much variance should be explained (usually 80, 90%, but at times 50% is used).

• Sometimes recommended to include those factors with $\lambda_i > 1$ since these factors may explain more than the "average factor". But, what is an "average factor?"

• Use scree plots: exploits the idea that if $p > k$, the covariance of X should have a special structure (only a few large eigenvalues).



(eigenvalues on the y-axis, number of factors on the x-axis). The point where the slope of the curve levels off (the elbow of the curve) indicates the number of factors.

2) Correlation test. If enough factors are included idiosyncratic component should be uncorrelated across i. Score test: compute the sum of all N(N-1) squared correlations, compare it to a χ^2. Test is asymptotically equivalent to a LR where H_0 is a model with r factors and H_1 a model with N factors.

- Problem: it requires $T >> N >> r$. Impossible to use when N is large.

3) Bai-Ng (2002) tests. Choose the k that minimizes:

$$
\begin{aligned}
IC_{p1}(K) &= \ln V(k,\hat{F}) + k\left(\frac{N+T}{NT}\right)\ln\left(\frac{N+T}{NT}\right) && (8)\\
IC_{p2}(K) &= \ln V(k,\hat{F}) + k\left(\frac{N+T}{NT}\right)\ln(\max\{N,T\}) && (9)\\
IC_{p3}(K) &= \ln V(k,\hat{F}) + k\frac{\ln(\max\{N,T\})}{\max\{N,T\}} && (10)\\
& && (11)
\end{aligned}
$$

where $V(k,\hat{F}) = argmin_{\Lambda_i}\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}(X_{it}-\Lambda_i F_t)^2$.

• All criteria consistent i.e. as $N,T \to \infty$, $k \to r$. Need to choose a $kmax$ to apply the criteria.

• If we use

$$BIC(k) = \ln V(k, \hat{F}) + k\left(\frac{\ln T}{T}\right) \tag{12}$$

$$AIC(k) = \ln V(k, \hat{F}) + k\left(\frac{2}{T}\right) \tag{13}$$

we DO NOT get consistent estimates of r.

• If we use $BIC(k)$ to estimate the number of factors when the factors are observables we DO get a consistent estimate ($AIC(k)$ overestimates with positive probability).

4.3.1 Adding Normality

- Suppose:

$$\begin{bmatrix} X \\ F \end{bmatrix} = N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \Lambda\Lambda' + \Psi & \Lambda \\ \Lambda' & I \end{bmatrix}\right.$$

- Then, using the properties of normal variables:

$$X|F \sim N(\Lambda F, \Psi) \tag{14}$$

$$F|X \sim N(\Lambda'\Sigma^{-1}X, I - \Lambda'\Sigma^{-1}\Lambda) \tag{15}$$

• Adding normality is equivalent to assume that F is a random vector. From (9) it is combination of observables, the X's, and of unobservables, $z \sim N(0, I - \Lambda\Sigma^{-1}\Lambda)$, where $\Sigma = \Lambda\Lambda' + \Psi$.

• Note 1): $\text{var}(F|X) = (1 - \Lambda\Sigma^{-1}\Lambda) \equiv (I + \Gamma)^{-1}$ where $\Gamma = \Lambda^{-1}\Psi(\Lambda^{-1})'$. By identification assumption we assume that Γ is diagonal

• Note 2) $\text{var}(F|X)$ is minimized when trace(Γ) is largest. In turns, this occur when the variance of u_i is small relative to the variance of the common factor $\lambda_i F$.

4.3.2 ML estimation: EM algorithm

• Idea:

- treat the factor F as latent variables and compute its expected value conditional on X_t.

- Conditional on the expected value of F, maximize the likelihood function.

E-step: takes conditional expectations, given the data.

M-step: maximizes the likelihood function.

- If F was observable the likelihood function would be:

$$logL \propto -0.5N\sum_{j=1}^{p}\log\Psi_{jj} - 0.5\sum_{i=1}^{N}\sum_{j=1}^{p}\frac{(X_{ij}-\lambda_j F_i)^2}{\Psi_{jj}} - 0.5\sum_{i=1}^{N}F_i'F_i \tag{16}$$

When F is unobservable, (16) can not be compute. Compute instead:

$$\begin{aligned} E(logL|X,\Psi,\Lambda) &\propto -0.5N\sum_{j=1}^{p}\log\Psi_{jj} \\ &- 0.5\sum_{i=1}^{N}\sum_{j=1}^{p}\frac{(E(X_{ij}-\lambda_j F_i)^2|X,\Psi,\Lambda)}{\Psi_{jj}} \\ &- 0.5\sum_{i=1}^{N}E(F_i'F_i|X,\Psi,\Lambda) \end{aligned} \tag{17}$$

In (17) we have substituted a latent variable F with its expected value and its expected variance conditional on X,Ψ,Λ.

Algorithm 4.2 - *(E-Step) Given* Λ, Ψ, X *use* $F = AX+e$ *to find* $E(F|X, \Psi, \Lambda)$, $cov(F|X, \Psi, \Lambda)$, $cov(e|X, \Psi, \Lambda)$.

Since $\hat{A} = FX'(XX')^{-1} = \Lambda'(\Lambda\Lambda' + \Psi)^{-1}$, *(using the representation for* X*) we have*

$$\begin{aligned}
E(F|X, \Psi, \Lambda) &= \Lambda'(\Lambda\Lambda' + \Psi)^{-1}X \\
cov(e|X, \Psi, \Lambda) &= F'F - \hat{A}XX'\hat{A}' \\
&= I - \Lambda'(\Lambda\Lambda' + \Psi)^{-1}\Lambda \equiv \Delta \\
E(X^2|X, \Psi, \Lambda) &= X'X \\
E(FF'|X, \Psi, \Lambda) &= \Delta + \hat{A}XX'\hat{A}' \\
E(XF'|X, \Psi, \Lambda) &= XE(F'|X, \Psi, \Lambda) = XX'\hat{A}' \qquad (18)
\end{aligned}$$

These are the elements needed to evaluate the Likelihood function.

- *(M-Step) Maximize the conditional log likelihood with respect to* Λ, Ψ*. Since the model is linear this is equivalent to computing the coefficients of the linear projection* $F = \Lambda X + u$ *and the variance of the prediction error, i.e. at iteration* $i+1$

$$
\begin{aligned}
\hat{\Lambda}^{i+1} &= E(XF'|X,\Psi,\Lambda)E(FF'|X,\Psi,\Lambda)^{-1} \\
&= XX'\hat{A}(\Delta + \hat{A}XX'\hat{A}')^{-1} \\
\hat{\Psi}^{i+1} &= \mathrm{diag}(XX' - (\hat{\Lambda}^{i+1}(\hat{\Lambda}^{i+1})') \qquad (19)
\end{aligned}
$$

- *Iterate on the E-M steps until convergence.*

Bayesian estimation:

• Assume priors on Λ, Ψ, F. Then (14) and (15) are the conditional of X and F, given Λ, Ψ.

• Under suitable prior assumptions, the conditional posterior of Λ is normal with mean Λ_{ols} and variance $(F'\Psi^{-1}F)^{-1}$

• The conditional posterior of Ψ is $IW((X - \Lambda F)'(X - \Lambda F), T - m)$.

• The conditional posterior of F is normal.

• Therefore put these conditionals into a Gibbs sampler. As the number of draws grows, the resulting draws are from the marginal posteriors.

5 Dynamic Factor Models

• Two versions

i) Parametric model: Stock and Watson (1989, KF), Ireland (2001, KF), Otrok-Whiteman (1998, Bayesian).

$$\begin{aligned} X_t &= \gamma F_t + u_t \\ \phi(\ell) F_t &= \eta_t \\ d(\ell) u_t &= e_t \end{aligned} \tag{20}$$

where X_t is $n \times 1$, F_t is a $m < n \times 1$,η_t and e_t are jointly normal.

• To estimate (20) use the EM algorithm, or the Gibbs sampler.

• Can use the KF to evaluate the likelihood at each step since the model it is a state space model.

ii) Nonparametric (Frequency domain) Geweke (1977), Sargent and Sims (1977), Altug (1989).

$$X_t = \sum_{s=-\infty}^{\infty} \Lambda_s F_{t-s} + \epsilon_t \tag{21}$$

X_t is a $n \times 1$ vector, F an $m < n \times 1$ vector, $E(F_t) = E(\epsilon_t) = 0$ and

- $E(F_{t-s}\epsilon_{t-r}) = 0, \ \forall s, r$.

- The elements of F_t may be mutually correlated.

- Both F_t and ϵ_t may be serially correlated.

- The assumptions made imply that all covariations in X_t are due to $\sum_{s=-\infty}^{\infty} \Lambda_s F_{t-s}$.

- If we represent $F_t = \sum_{j=0}^{\infty} B_j e_{t-j}$ then

$$\begin{aligned} X_t &= \Lambda(\ell) F_t + \epsilon_t \\ &= \sum_{s=-\infty}^{\infty} \Lambda_s \left(\sum_{j=0}^{\infty} B_j e_{t-j-s} \right) + \epsilon_t \\ &= \sum_{s=-\infty}^{\infty} D_s e_{t-s} + \epsilon_t = D(\ell) e_t + \epsilon_t \end{aligned} \tag{22}$$

- Therefore, the spectral density of X at frequency ω is

$$S_X(\omega) = D(\omega) S_e(\omega) D(\omega)' + S_\epsilon(\omega) \tag{23}$$

If this is calculated at Fourier frequencies, $S_X(\omega_j)$ is orthogonal to $S_X(\omega_{j'})$, for $\omega_j \neq \omega_{j'}$. Then, for a given ω, we need to solve the same problem as with iid data.

• A factor model with time series correlation is transformed into a vector of (iid) factor models, one per frequency.

• We have the same identification problems as in static factor models (little more subtle here since we need to deal with complex numbers and the fact that the problem appears at all frequencies).

• Estimation approach: EM algorithm, frequency by frequency, under normality of both the e_t and the ϵ_t.

Few issues

• EM works better if the spectral density is roughly constant in an interval: better to prewhiten the data first.

• Can approximate the likelihood function in frequency domain.

• To test overidentifying restrictions use a LR test.

6 From small to large n

- If n is large, the LLN implies that idiosyncratic components will be averaged out. Hence, we can estimate the factors by large sample (cross sectional) averages.

Example 4 *Suppose* $X_i = a_i f + u_i,\ i = 1, \ldots, n$. *Note*

$$\begin{aligned} var(\bar{X}) &= var(n^{-1}\sum_i X_i) = var(n^{-1}(\sum_i a_i f)) + var(n^{-1}(\sum_i u_i)) \\ &\leq n^{-2}\sum_i a_i^2 + n^{-2}(n * \sigma^2_{u,\mathsf{max}}) = n^{-2}\sum_i a_i^2 \end{aligned} \tag{24}$$

Then if follows that

$$\bar{X}_\infty = \lim_{n\to\infty} n^{-1}\sum_i X_i = \lim_{n\to\infty}(n^{-1}\sum_i a_i)f + \lim_{n\to\infty}(n^{-1}\sum_i u_i) = \bar{a}f \tag{25}$$

so that $\bar{X}_\infty$ spans the same space as f! It is a linear transformation of f.

• Conclusion: averages can be used to proxy for the unobservable f, so the procedure becomes:

Step 1: Recover the space of factors (via cross sectional aggregation).

Step 2: Project X_i over the space of factors, gets the loadings.

$$X_i = \beta_i \bar{X}_\infty + u_i \approx \beta_i \bar{a} f + u_i \tag{26}$$

Note: we can use any average, not necessarily the arithmetic one.

For example, if the cross section is large, we can simplify algorithm 4.1 as follows:

Algorithm 6.1 - *Let Σ be the covariance of the endogenous variables.*

- Decompose $\Sigma = ABA'$, where A are the eigenvectors and B the matrix of eigenvalues. Take the k largest eigenvalues and the associated eigenvectors and construct $\Lambda^1 = AB^{0.5}$.

- Construct $F_t = \Lambda X_t$

- No iteration needed. Just one step. Disregard the presence of Ψ.

7 Generalizations

7.1 The Generalized Static Factor Model

• What if the idiosyncratic terms are cross sectionally correlated?

- Procedure breaks down since it is impossible to separate common and idiosyncratic component.

- Solution: the Generalized factor model (Chamberlain (1983) and Chamberlain and Rothschild (1983)).

$$X_i = \Lambda_i f + \xi_i \equiv \chi_i + \xi_i \quad i = 1, \ldots, n \tag{27}$$

so $\Sigma_x = \Lambda\Lambda + \Psi$.

Let $\Gamma_n^{\chi}, \Gamma_n^{\xi}$ be the covariances of χ_n, ξ_n, respectively and let $\mu_{nh}^{\chi}, \mu_{nh}^{\xi}$ be the h-th eigenvalue of $\Gamma_n^{\chi}, \Gamma_n^{\xi}$ where eigenvalues are ordered decreasingly. Let the dimension of f_t be r. Assume:

- The components of χ_j and ξ_i are mutually orthogonal, $\forall i, j$.

- $\mu_{nr}^{\chi} \to \infty$ as $n \to \infty$.

- There exists a real M such that $\mu_{n1}^{\xi} \leq M$, for any n.

With the third assumption, as n increases the variance of X captured by the largest eigenvalue of the idiosyncratic component remains bounded, while with the second assumption, the variance explained by the first r eigenvalues of the common component grows to infinity.

As $n \to \infty$ the importance of the idiosyncratic component in explaining the variance of X becomes negligible. Hence, by the LLN, as $n \to \infty$, r appropriately chosen linear combinations of X become increasingly collinear with the r common factors.

To estimate the common component we can then use:

Algorithm 7.1 - *Find the aggregates which lie in the space of factors. Analogously to the case where* Ψ *is diagonal, as* $n \to \infty$, *PC=PF so that the first* k *principal components are consistent estimators of the* k *common factors,* $\forall k$

- Project X *on the first* r *principal components found in the first step to find the loadings.*

7.2 The Generalized Dynamic Factor Model

$$X_{it} = b_i(\ell)u_t + \xi_{it} \equiv \chi_{it} + \xi_{it}$$

so that $\Sigma_x(\omega) = \Sigma_\chi(\omega) + \Sigma_\xi(\omega)$,where rank($\Sigma_x(\omega)$) =rank($\Sigma_\xi(\omega)$) $= n$, rank($\Sigma_\chi(\omega)$) $= m$ and no restrictions are imposed in $b_i(\ell)$ (in particular, it could be two sided).

Let $\mu^\chi_{nh}(\omega)$, $\mu^\xi_{nh}(\omega)$ be the h-th dynamic eigenvalue of $\Sigma_\chi(\omega)$, $\Sigma_\xi(\omega)$ where eigenvalues are ordered decreasingly.

As in the static factor model we assume:

- The components of u_t and ξ_{it} are mutually orthogonal at all leads and lags.

- $\mu^{\chi}_{nq}(\omega) \to \infty$ as $n \to \infty$, almost everywhere in $[-\pi, \pi]$.

- There exists a real M such that $\mu^{\xi}_{n1}(\omega) \leq M$, for all n and all $\omega \in [-\pi, \pi]$

Then same logic of the generalized static model applies, frequency by frequency, to the generalized dynamic model.

Problem: the factors extracted this way are useful only in certain situations since at each t they reflect information from $t-q$ up to $t+q$ with $q \to \infty$.

7.3 The MA generalized factor model

$$X_{it} = b_i(\ell)u_t + \xi_{it} \tag{28}$$

Assume:

- $b_i(\ell)$ is **one sided** and of finite order s.

- u_{t-j} and ξ_{it-h} are orthogonal all j, h.

- $\mu_{nr}^{\chi} \to \infty$ as $n \to \infty$, for $r = q(s+1)$.

- There exists a real M such that $\mu_{n1}^{\xi}(\omega) \leq M$, for all n and all $\omega \in [-\pi, \pi]$.

The generalized MA model satisfies the conditions of the generalized static factor model and can be written as

$$X_{it} = B_i F_t + \xi_t \equiv \chi_{it} + \xi_{it}$$

where $B = [b_{i,o}, \ldots, b_{i,s}], F_t = [u_t, \ldots, u_{t-s}]'$. This is the model proposed by Stock and Watson (2002). Note :

- Static factors are r elements of F_t (number of static factors is the rank of cov(χ_t)).

- Dynamic factors are q elements of u_t (number of dynamic factors is the rank of the spectral density of χ_t).

The match can be made if and only if $s < \infty$.

7.4 Structural factor models

$$\begin{aligned} X_{it} &= c_i(\ell)f_t + \xi_{it} && (29) \\ a(\ell)f_t &= du_t && (30) \end{aligned}$$

where $c_i(\ell)$ is of order s for all i and $a(\ell)$ is of order p. Combining the two equations we have $X_{it} = b_i(\ell)u_t + \xi_{it}$ which is the MA model we have considered before.

Stacking: $F_t = [f_t', f_{t-1}', \dots f_{t-s}']$ which is of dimension $r = q(s+1)$. If $p < s+1$ we can rewrite the system as:

$$\begin{aligned} x_{it} &= C_iF_t + \xi_{it} && (31) \\ F_t &= AF_{t-1} + DU_t && (32) \end{aligned}$$

where $D = (b, 0, \ldots, 0)'$ $U_t = (u_t, 0, \ldots, 0)'$ and

$$A = \left[\begin{array}{cccc} a_1 & a_2 & \ldots & a_p \\ I & 0 & \ldots & 0 \\ 0 & I & \ldots & 0 \\ \ldots & \ldots & \ldots & \ldots \\ 0 & 0 & \ldots & I \\ \end{array} , \right], \quad C_i = \left[\begin{array}{cccc} c_{i1} & c_{i2} & \ldots & c_{is} \\ I & 0 & \ldots & 0 \\ 0 & I & \ldots & 0 \\ \ldots & \ldots & \ldots & \ldots \\ 0 & 0 & \ldots & I \\ \end{array} , \right]$$

In this model, consistent estimates of the loadings C_i, of the factors F_i and of the responses to shocks in the factors are obtained with standard methods.

- A linearized DSGE model has the same format as (31) and (32):

$$
\begin{aligned}
y_t &= C x_t + e_t && (33)\\
x_t &= A x_{t-1} + v_t && (34)
\end{aligned}
$$

where y_t are the controls and x_t are the states.

Note: (31) and (32) define a structural factor model. We can do inference here in the same way as in SVAR as the next algorithm shows.

Algorithm 7.2 - *Let* $\Sigma_x = E(X_t X_t')$. *Find* R, V, *such that* $RVR' = \Sigma_x - \Psi$.

- V *is an* $r \times r$ *diagonal matrix with diagonal elements given by the* r *largest eigenvalues of* Σ_x *and* R *the* $n \times r$ *matrix of eigenvectors;*

- *An estimate of the loadings is* $\hat{C} = R$.

- *An estimate of the factors is* $\hat{F}_t = R'X_t$ *and an estimate of* Ψ *is* $E(X_t - \hat{C}\hat{F}_t)(X_t - \hat{C}\hat{F}_t)'$

- *An estimate of the autoregressive matrix* $\hat{A} = (\hat{F}_{t-1}\hat{F}_{t-1}')^{-1}(\hat{F}_{t-1}'\hat{F}_t)$.

- *An estimate of* $e_t = Du_t$ *is* $\hat{e}_t = \hat{F}_t - \hat{A}\hat{F}_{t-1}$ *and an estimate of* Ω *is* $\hat{\Omega} = E(\hat{e}_t\hat{e}_t')$.

Identify shocks to the factors: find M, P satisfying $MPM' = \Omega$.

- P is the $q \times q$ diagonal matrix with diagonal elements given by the q largest eigenvalues of Ω and M is the $n \times q$ matrix of the corresponding eigenvectors.

- Compute $\hat{B} = MP^{-0.5}$. Verify if identification restrictions are satisfied(i.e. check if the responses of F_t (or of X_t) to shocks in u_t follow the required pattern).

- Use $HH' = I$ to generate an alternative $\hat{B}^* = MP^{-0.5}H$. Verify if identification restrictions are satisfied. Continue.

The steps of the algorithm can be iterated and the factors can be recursively obtained with the Kalman Filter (using a version of the EM algorithm) if N is small.

7.5 Determining the number of dynamic factors

- Given that some information criteria (see Bai and Ng(2002)) has chose some r static factors, how do we know how many dynamic factors there are?

• Bai and Ng (2007). Compute the spectral density of f_t. Its rank is the number of dynamic factors in the model.

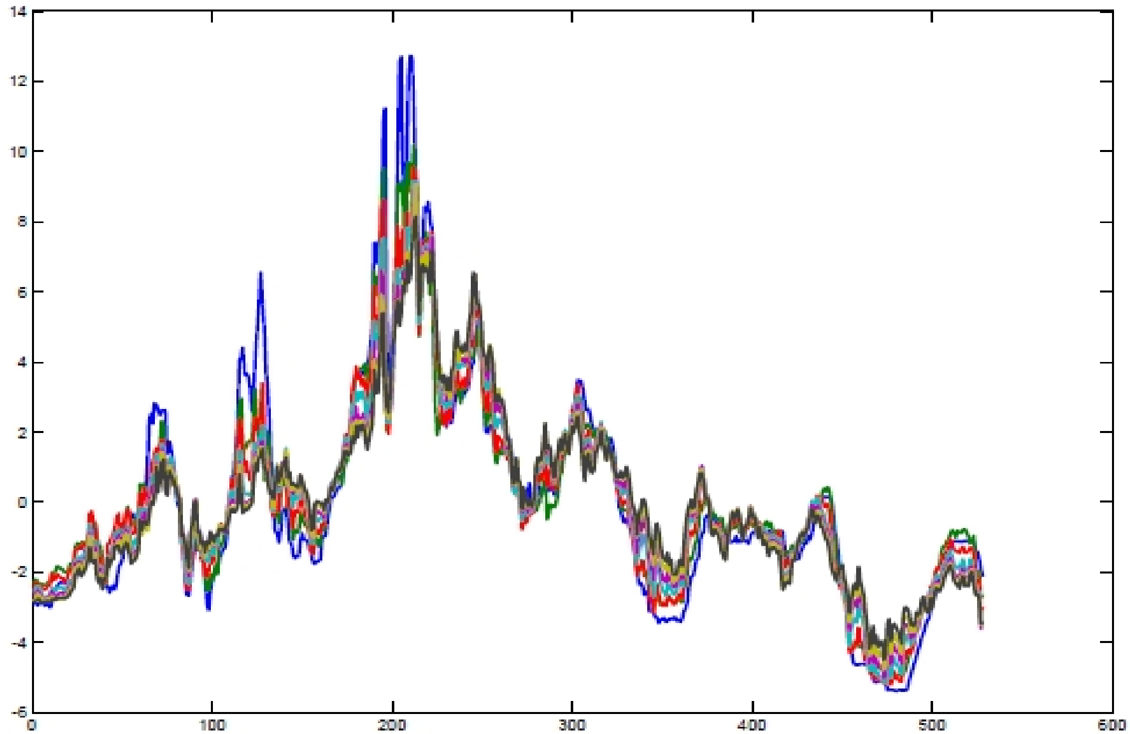
• Bai and Ng (2002) criteria can be used to find the number of dynamic factors (if applied to a system where F is less than full rank the method of principal components will find the number of factors $k < r$ which span the space of dynamic factors).

- Structural factor models impose an important restriction: the factor F_t are exogenous with respect to X_t.

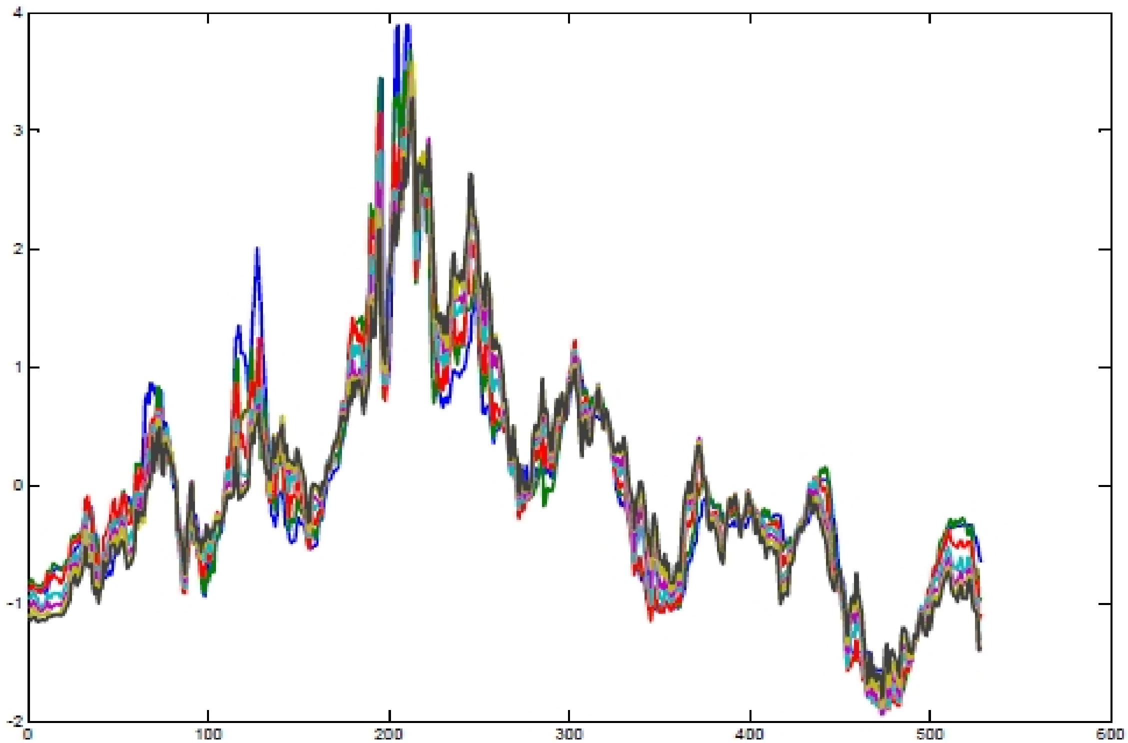
- This is the same setup used in stress testing exercises.

- Assumption may be appropriate in some cases. What to do when F_t are endogenous?

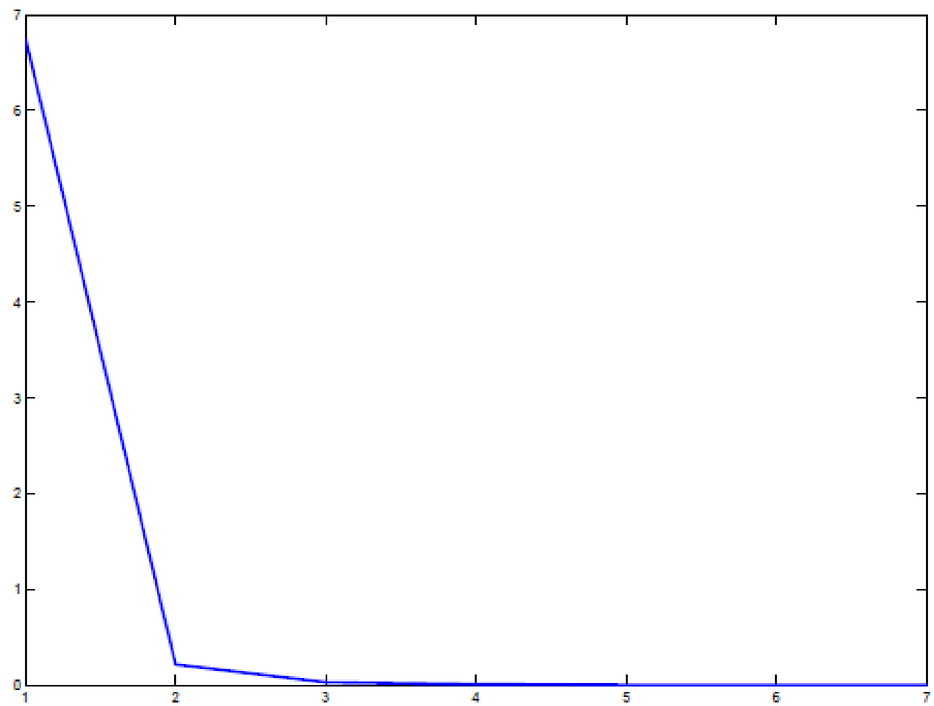
Example 5 *US Bond (Fed Funds) yields:3,12,24,36,48,60 months.*



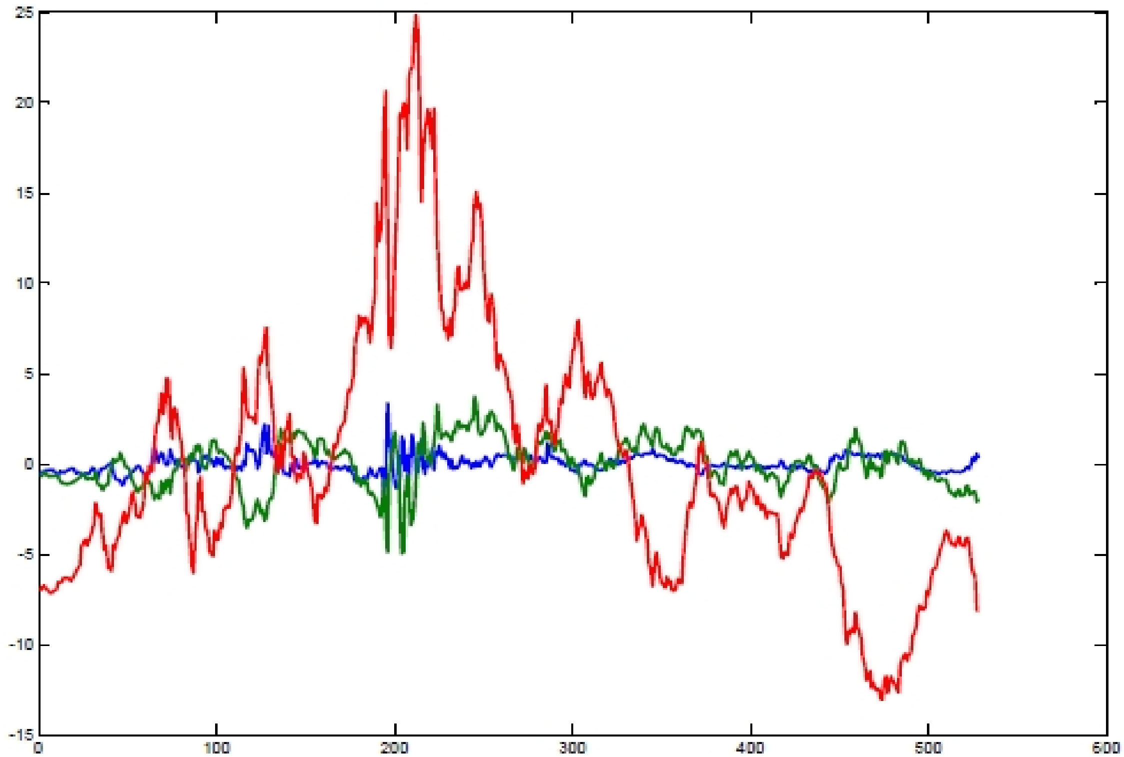
• 1) Normalize variances to make sure unit of measurements do not matter:
$\hat{x}_{it} = \frac{x_{it}}{E(x_{it}x'_{it})^{0.5}}, i = 1, 2, \ldots, n.$



• 2) Do eigenvalue decomposition $E(\hat{x}_t \hat{x}_t') = \mathbf{\Sigma} = W \Lambda W'$ (Matlab:$[W, \Lambda] = eig(\mathbf{\Sigma})$). Scree plot:

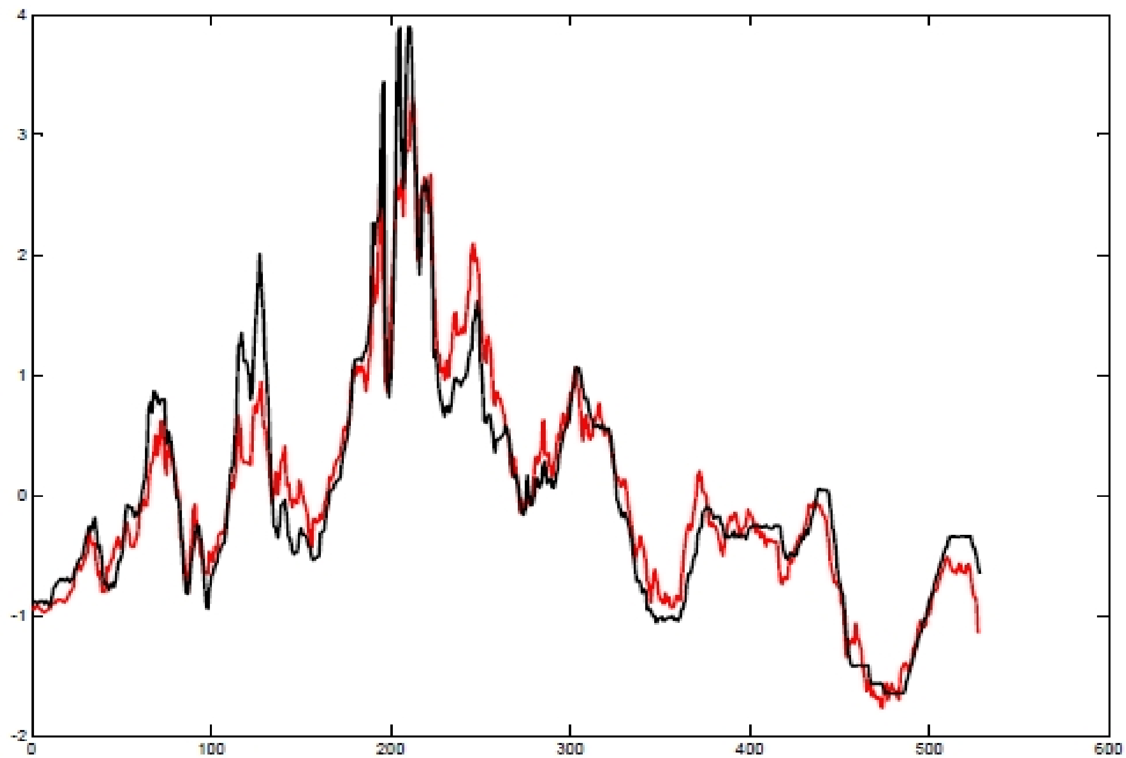


- 3) Get the factors: $\hat{F}_t = W'\hat{x}_t$.

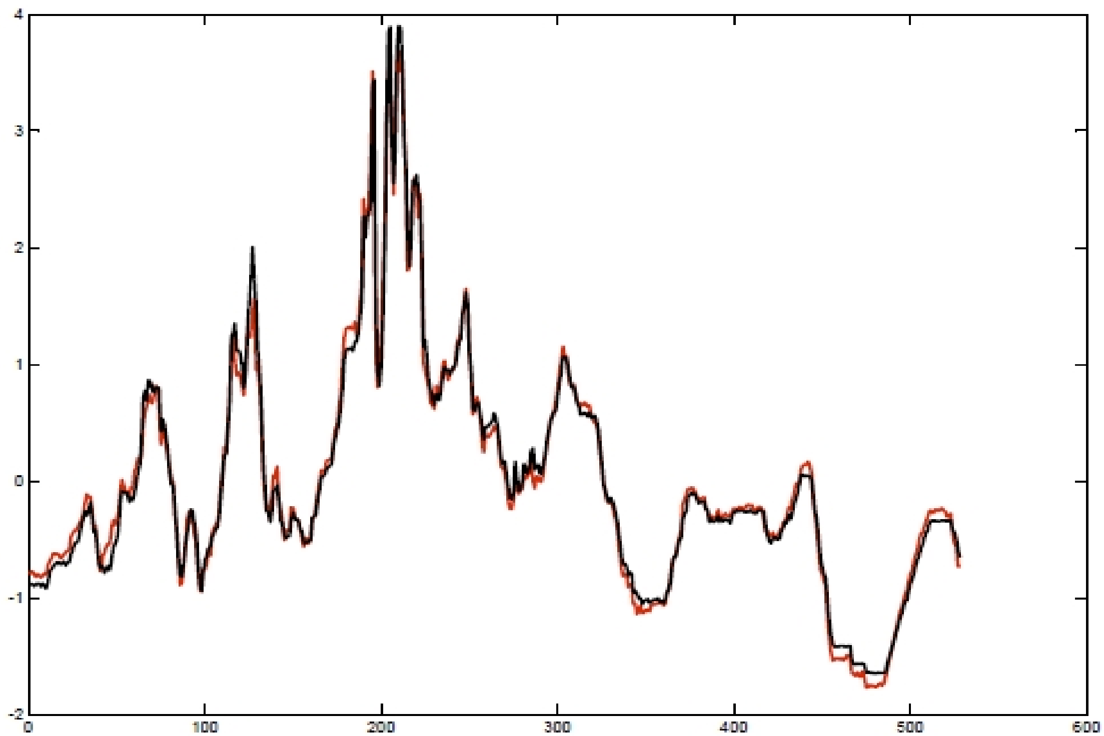


Time series of first three factors

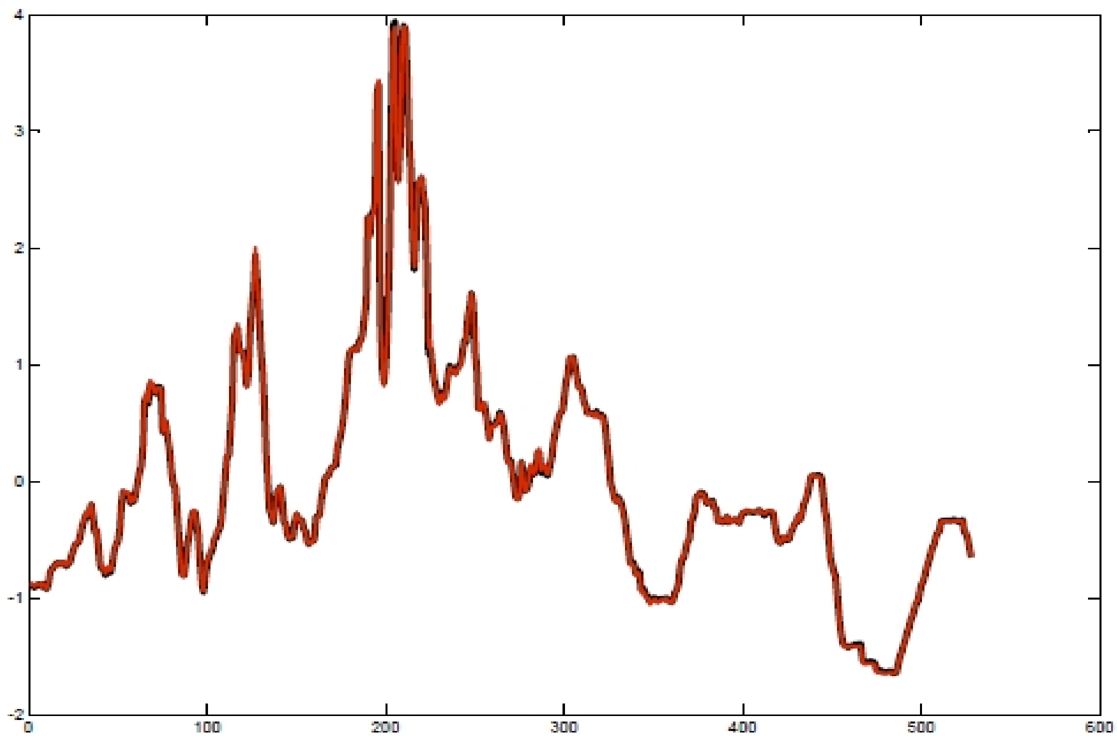
• 4) Check that indeed the factors explain the data. Plot $\tilde{x}_{1t} = W_3 \hat{F}_{3t}$ (the prediction of short FF Rate with up to 3 factors.)



Fit with one PC



Fit with two PC



Fit with three PC

• 5) Estimate the dynamics of the Factors $F_t = [f_{1t}, f_{2t}, f_{3t}]' = AF_{t-1} + E_t$.

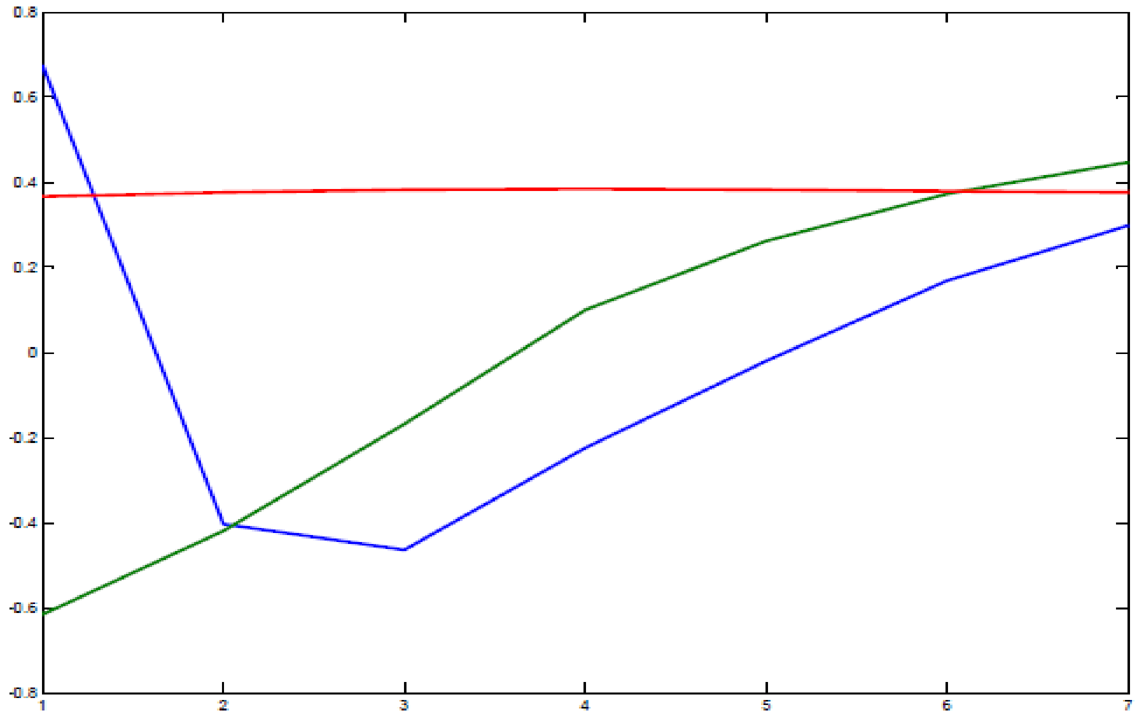
$\hat{A} = (F_{t-1}F_{t-1}'^{-1}(F_{t-1}F_t)$.

• 6) Identification of shocks ot factors $\Sigma_E = D\Sigma_U D'$.

• We have modeled comovements of 7 times series with three factors. Need to estimate A (3×3), W (7 parameters), Σ_E (6 parameters), Σ_ζ (7 parameters). Total 29 parameters.

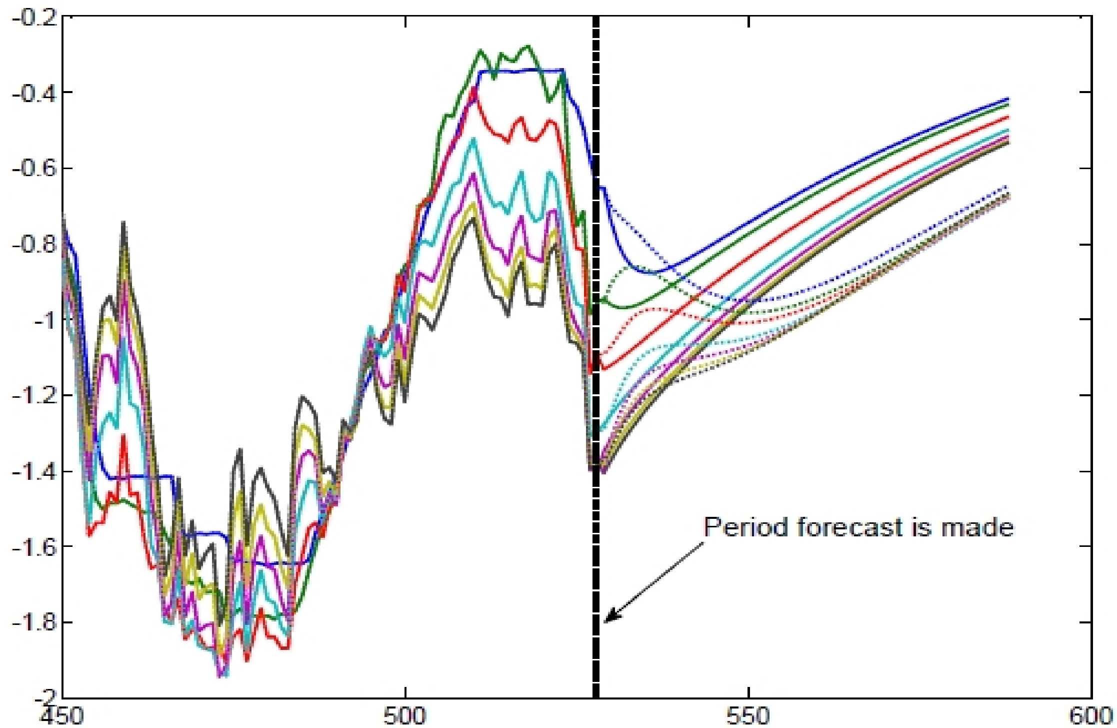
• In a VAR(1) with 7 variables we need to estimate A (7×7) and Σ (21 parameters). Total 70 parameters.

• 7) Interpreting the factors? Can not interpret them economically, but can give them a name. Plot factor loadings.



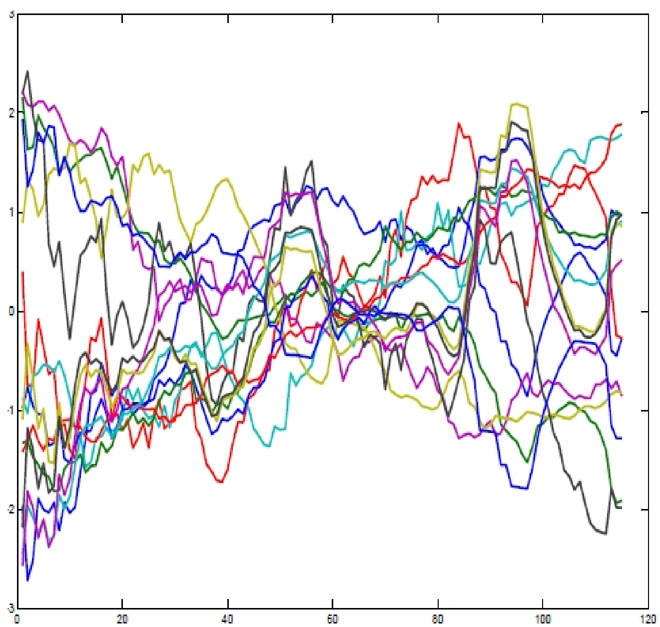
Factor loadings (level, slope, curvature)

- 8) Forecasting. h-step ahaed forecast $\hat{F}_{t,h} = \hat{A}^h F_t ; \hat{x}_{t+h} = W F_{t+h}$.

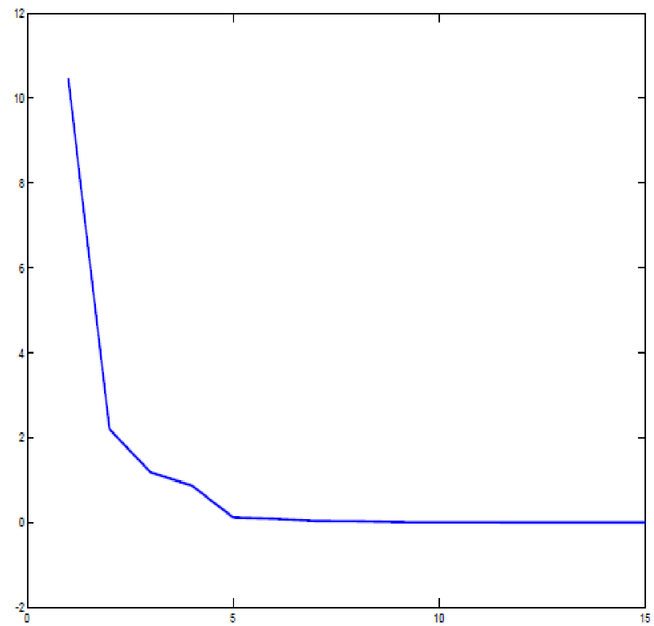


Factor forecast (solid) vs. VAR(1) forecast (dotted)

• Factor structure is obvious in bond yields data. Much less obvious in macro data.



15 Australian macro series



Scree plot

8 FAVAR models

i) VARs use only very sparse information. Potential problems:

• Estimated shocks are contaminated. Need long lags to make the residuals white noise. Long lags may not be enough if non-invertibility is present.

• Selection of variables is arbitrary (what is real activity? what is inflation?).

• Impulse responses can be computed only for the variables in the VAR.

ii) Forecasting ability maybe impaired. Difficult to decide which variable to include and which to exclude.

Example 6 *If*

$$\begin{bmatrix} \delta_{11}(\ell) & \delta_{12}(\ell) \\ \delta_{21}(\ell) & \delta_{22}(\ell) \end{bmatrix} \begin{bmatrix} y_{1t} \\ y_{2t} \end{bmatrix} = \begin{bmatrix} \epsilon_{1t} \\ \epsilon_{2t} \end{bmatrix}$$

and y_{2t} not used, the representation for y_{1t} is

$$\begin{aligned} [\delta_{11}(\ell) - \delta_{12}(\ell)\delta_{22}(\ell)^{-1}\delta_{21}(\ell)]y_{1t} &= \epsilon_{1t} - \delta_{12}(\ell)\delta_{22}(\ell)^{-1}\epsilon_{2t} && (35) \\ G(\ell)y_{1t} &= \upsilon_t && (36) \end{aligned}$$

• υ_t *linear combination of* ϵ_{1t} *and* ϵ_{2t}.

• *Need a very long $G(\ell)$ to be able to make the residuals white noise. May not be possible with finite number of data or if $\delta_{j2}(\ell), j = 1, 2$ has one root close to 1.*

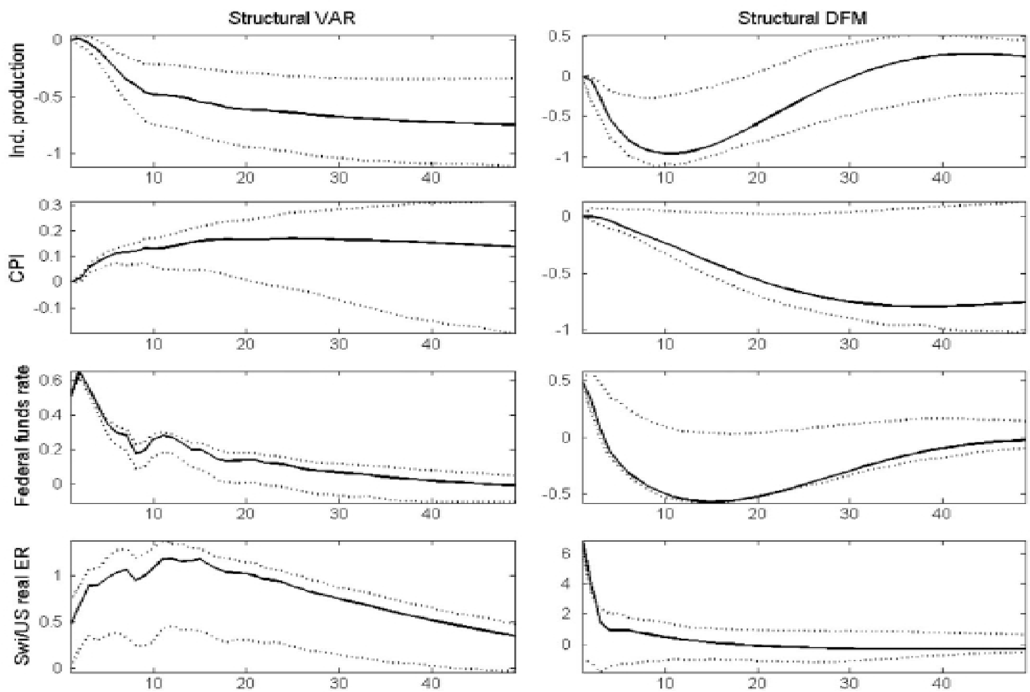
Example 7 *Sims'(1992) price puzzle: increase in prices after nominal interest rate increased. Grilli and Roubini (1996) exchange rate puzzle: real exchange rate (an asset price) reacts very slowly to nominal interest rate increases.*

- *Solution 1: add to the VAR a measure of commodity prices/ change identification scheme.*

- *Idea: Central bank cares about expected inflation. Index of commodity prices proxies for future inflation.*

- *Solution 2: Choleski restrictions are not a feature of dynamic models, use other restrictions.*

- *Solution 3: Forni and Gambetti (2010): add "factors" to the VAR.*



Structural VAR
Structural DFM
Ind. production
CPI
Federal funds rate
Swi/US real ER

Example 8 *Stock and Watson(2002), Bernanke and Boivin (2003)*

$$X_t = \Lambda F_t + e_t \tag{37}$$

$$w_{t+1} = \beta F_t + \epsilon_{t+1} \tag{38}$$

$$R_t = \gamma E_t w_{t+1} + u_t \tag{39}$$

X_t *is a* $n \times 1$ *vector, of fast moving variables,* n *large. Could contain monthly or quarterly data, regular and irregularly sparsed (starting in the middle of the sample) data.* w_t *are the variables we want to forecast (inflation, output, etc.),* R_t *policy instrument.*

- *Assume* $E(\epsilon_t) = 0$*, we can allow for some time series dependence in* e_t *see Stock and Watson (2002).*

- *Transform the data in* X_t *and* w_{t+1} *so that the system is stationary.*

- F_t *can contain both current and lagged values of the factors.*

- *How good is the setup relative to standard forecasting models for w_t?*

Table 1
Relative forecasting performance

	FM-VAR	FM-AR	VAR
CPI[a]			
Real-time	1.04	0.98	1.05
	0.97	0.96	0.95
Revised	1.08	1.00	1.05
	1.00	0.98	0.95
SW	0.83	0.82	1.05
	0.76	0.75	0.95
IP[b]			
Real-time	1.00	0.84	1.17
	1.07	0.92	1.12
Revised	1.06	0.86	1.17
	1.04	0.90	1.12
SW	0.69	0.63	1.17
	0.75	0.65	1.12
Unemployment[c]			
Real-time	0.90	0.86	1.06
	0.87	0.80	0.94
Revised	0.91	0.85	1.06
	0.85	0.78	0.94
SW	0.70	0.65	1.06
	0.90	0.55	0.94

Notes: The entries show the mean square error of forecast, relative to the autoregressive (AR) model, for the indicated forecasting method and conditioning data set. Methods are factor model plus univariate autoregressive terms (FM-AR); factor model plus vector autoregression in inflation, industrial production, unemployment, and the federal funds rate (FM-VAR); and a vector autoregression without factors, as above (VAR). Of the two numbers given in each entry, the first applies to forecasts at the 6-month horizon, the second to the 12-month horizon. CPI and IP are forecast as cumulative growth rates, and the unemployment rate in levels.

[a] AR RMSE: 1.3 (6-month), 2.6 (12-month).

[b] AR RMSEs: 4.1 (6-month), 5.8 (12-month).

[c] AR RMSEs: 0.74 (6-month), 1.17 (12-month).

General setup:

$$\begin{bmatrix} F_t \\ y_t \end{bmatrix} = \phi(\ell) \begin{bmatrix} F_{t-1} \\ y_{t-1} \end{bmatrix} + v_t \quad v_t \sim (0, Q) \tag{40}$$

where $\phi(L)$ is of order d, F_t unobservable, may include both current and lag values.

Recall (log-linear) solution of a DSGE is of the form:

$$\begin{aligned} y_{2t} &= PPy_{2t-1} + QQy_{3t} && (41) \\ y_{1t} &= RRy_{2t-1} + SSy_{3t} && (42) \end{aligned}$$

Same format as (40) where $F_t = y_{2t}$ are the unobservable states and $y_t = y_{1t}$ are the controls.

- What happens if we disregard F_t and estimate $y_t = G(\ell)y_{t-1} + u_t$? Then: $u_t = \phi_{21}F_{t-1} + v_{2t}$ and if F_{t-1} is correlated with y_{t-1}, $\hat{G}(\ell)$ biased(standard omitted variable problem).

Mapping DSGE models into factor models:

Example 9

$$Y_t = E_t Y_{t+1} - \frac{1}{\varphi}(r_t - E_t \pi_{t+1}) + g_t \quad (43)$$

$$\pi_t = \beta E_t \pi_{t+1} + \kappa(Y_t - Y_t^p) + s_t \quad (44)$$

$$r_t = \phi_r r_{t-1} + (1 - \phi_r)(\phi_\pi \pi_t + \phi_y Y_t) + e_t \quad (45)$$

$$Y_t^p = \rho_1 Y_{t-1}^p + \eta_t \quad (46)$$

Assume g_t, s_t, e_t *are iid. Endogenous variables* (Y_t, π_t, r_t, Y_t^p)*. Solution will be of the form (41)-(42). What is* F_t*?*

i) If all variables are observable: the solution is a VAR. No F_t*.*

ii) If (Y_t^P) *is not observable, set* $F_t = (Y_t^P)$*. Econometrician needs to use* X_t *to capture the effects of* (Y_t^P) *(monetary authority knows the model).*

iii) What if also r_t is a function of X_t (rather than (π_t, Y_t) as in the Taylor rule). For the econometrician, the system is still a FAVAR.

iv) Only r_t is observable; (Y_t, π_t, Y_t^P) are unobservables (because they are either noisy or not available). The system is now a FAVAR for both the monetary authority and for econometrician.

• How do you estimate (40) when F_t is unknown?

• Let X_t be informational variables and assume that

$$X_t = \Lambda^f F_t + \Lambda^y Y_t + e_t \tag{47}$$

where Λ^f is of dimension $N \times k$ and Λ^f is of dimension $N \times m$ and $E(e_t) = 0$, Y_t are observable variables.

• Assume either normality and serial uncorrelation of the e_t or if N is large can allow for some form of dependence in the e_t.

• Conditional on Y_t, X_t is a noisy measure of F_t. So use the PCs of X to get estimates of F_t.

Estimation: 2 step approach

- Find the space spanned by the factors. This is estimated as the first $k+M$ principal components of X_t(denoted by $\hat{C}(F_t, Y_t)$).

- If N is large and if the number of estimated PC is at least as large as the number of true factors, PC are consistent estimators of the space spanned by (F_t, Y_t) even if Y_t is not used in the estimation.

- Compute the part of $\hat{C}$ not spanned by Y_t, i.e. regress $\hat{C}(F_t, Y_t) = bC^*(F_t) + b_y Y_t + e_t$ where $C^*(F_t)$ is obtained using variables not in Y_t. Use $\hat{F}_t = \hat{C}(F_t, Y_t) - b_y Y_t$.

- Substitute $\hat{F}_t$ for F_t in the VAR. Careful: generated regressor problem. Standard error produced by standard packages are wrong. However, when $N >>> T$ uncertainty in the factors can be safely disregarded.

- Alternative ML (one step) approach: Jointly estimate (40) and (47)). Difficult to do it with classical methods.

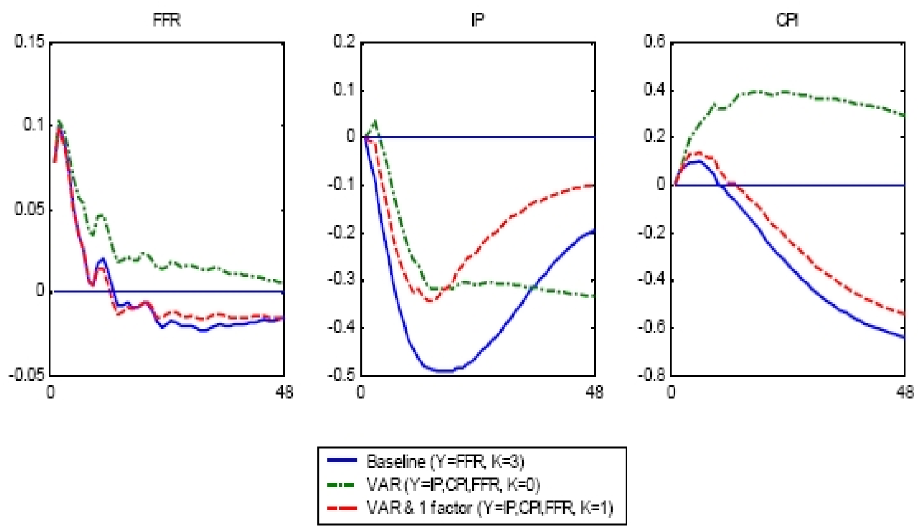
- Ahmadi and Uhlig (2015) show how to do system-wide estimation with Bayesian methods.

Identification and practical issues

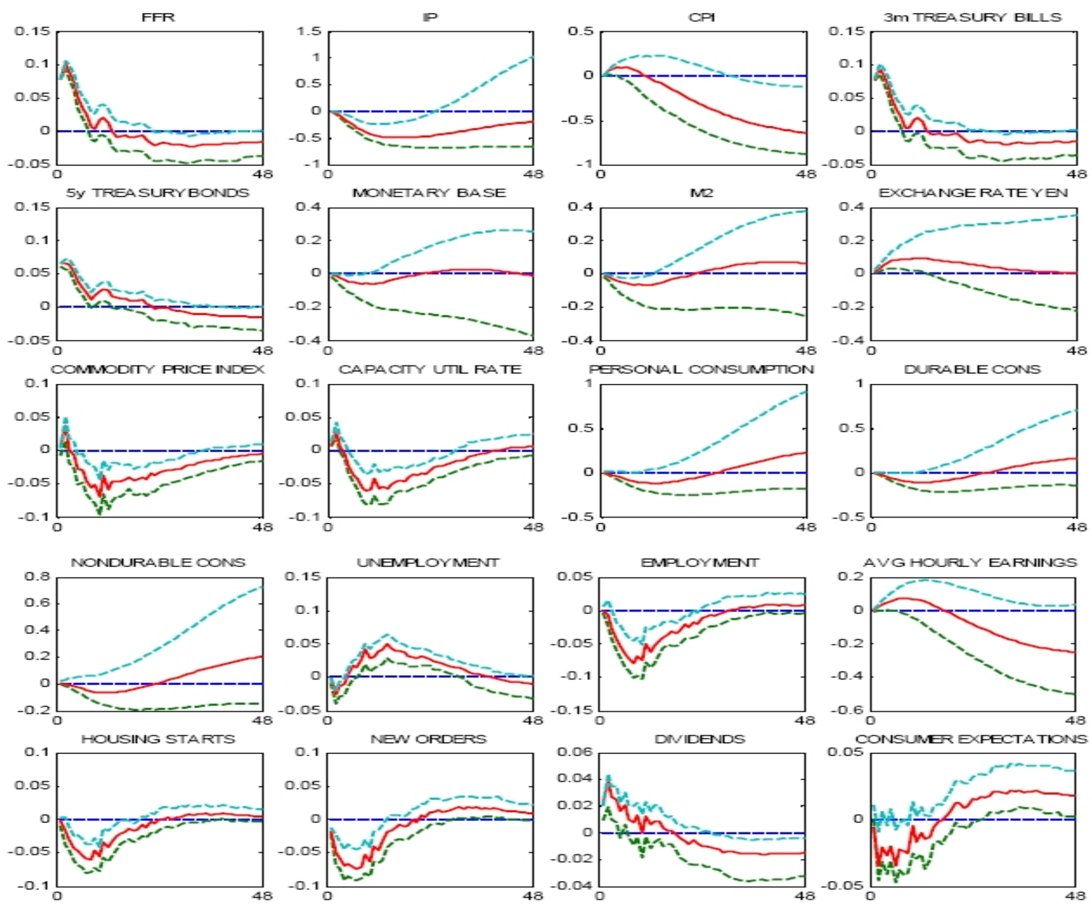
• Since neither Λ^f nor F are observable $\Lambda^f H$ $H'F$ are equivalent for any orthogonal H. Need normalization. As in PC use $\frac{C'C}{T} = I$ where $C = [C_1(F_1, y_1) \ldots C_1(F_T, y_T)]$ and set $\hat{C} = T^{0.5}\hat{Z}_t$ where $\hat{Z}_t$ are the eigenvectors of the K largest eigenvalues of XX', sorted in ascending order.

• Shock identification (standard approach). Note: if want to "name" shocks moving the factors need to give an economic content to the factors.

• How do you compute responses of VAR variables? Standard. How do you compute responses of the variables in X_t? Compute responses of the factor and then use (47) to compute the responses of the variables in X_t.



Estimated impulse responses to an identified policy shock for alternative FAVAR specifications, based on the two-step principal component.


FFR
IP
CPI
3m TREASURY BILLS
5y TREASURY BONDS
MONETARY BASE
M2
EXCHANGE RATE YEN
COMMODITY PRICE INDEX
CAPACITY UTIL RATE
PERSONAL CONSUMPTION
DURABLE CONS
NONDURABLE CONS
UNEMPLOYMENT
EMPLOYMENT
AVG HOURLY EARNINGS
HOUSING STARTS
NEW ORDERS
DIVIDENDS
CONSUMER EXPECTATIONS

Example 10 *FAVAR for measuring the effects of expenditure policy (fiscal multiplier)*

$$\begin{bmatrix} F_t \\ G_t \end{bmatrix} = \mathbf{\Phi} \begin{bmatrix} F_{t-1} \\ G_{t-1} \end{bmatrix} + u_t \tag{48}$$

$\mathbf{\Phi}$ companion matrix.

• 1) Regress $Y_t = \beta G_t + v_t$, find the factors in $\hat{v}_t$ (want factors to capture information orthogonal to R_t) via eigenvalue decomposition of $\mathbf{\Sigma}_v$.

• 2) Estimate (48) by OLS, treating $\hat{F}_t$ as if it was the true F_t (Why is it possible?)

• 3) Identify a monetary policy shock, e.g. with a Cholesky decomposition, i.e. recover A_0, A_1 in

$$A_0 \begin{bmatrix} F_t \\ G_t \end{bmatrix} = A_1 \begin{bmatrix} F_{t-1} \\ G_{t-1} \end{bmatrix} + e_t \quad e_t \sim (0, I) \tag{49}$$

• Responses of G_{t+j} to expenditure shocks are the (2,2) elements of $A_0 A_1^j$

• Responses of y_{t+j} to expenditure shocks are W times the (1,2) element of $A_0 A_1^j$, where W is the vector of loadings plus β times the (2,2) element of $A_0 A_1^j$.

9 Bayesian FAVAR

• Often we have available several indicators for the same variable (e.g. CPI, PCE, GDP deflator, etc.)

• We have also many variables that could potentially enter a VAR. Which ones do we choose? Curse of dimensionality problem.

• FAVAR is a good solution. Construct a few factors from a number of indicators. Run a VAR with observables variables and factors. Construct responses of variables not in the VAR using the factors.

• Can we run Bayesian FAVAR? Yes. What is the justification (no mode dimensionality reduction because a FAVAR already does that)?

• No new issues for naive treatment. Treat factors as observable variables. Use conjugate priors for the β and the Σ coefficients. Derive posterior distribution for the coefficients of variables not in the VAR using posterior dynamics of the factor.

• Coefficients of the variables in the VAR and outside the VAR is asymmetric (the former have a prior, the latter do not).

• How do we set up a consistent prior for the FAVAR? Need to setup a proper unobservable factor model and use a Gibbs sampler (see notes about state space models).

10 Large scale VARs, panel VARs, and factor models

• Large scale VARs or Panel VARs with some flexible restrictions on the coefficients generate observable factor models (Canova and Ciccarelli (2009)). Panel VAR:

$$y_{it} = D_i(L)Y_{t-1} + F_i\,(L)\,W_{t-1} + e_{it} \tag{50}$$

$i = 1, ..., N$ countries, y_{it} is $G \times 1$, W_t are the exogenous variables, $Y_t = \left(y'_{1t}, \ldots y'_{Nt}\right)'$.

• Parameter specific to each cross sectional unit.

• Allow for general form lagged interdependencies.

• Impossible to estimate this model with classical unrestricted methods: each equation has $k = NGp + Mq$ coefficients, and $r = NG$ equations: T will be smaller than $k \times r$. Short cuts available in the literature.

• Assume homogenous dynamics ($D_i(L) = D(L), F_i(L) = F(L)$); pool the data.

• Assume no dynamic interdependences ($D_i(L) = diag(D_{ii}(L))$.

• Assume restricted dynamics interdependences with a spatial structure: $D_{ij}(L) = \rho^{i-j} D_i(L)$, where ρ is a measure of distance.

• Assume that dynamics interdependences can be captured with one factor, i.e. $\Sigma_j D_{ij}(L) Y_{t-1} = a\bar{Y}_t$ where $\bar{Y}_t$ is computed using fixed weights (e.g. trade weights).(GVAR approach)

• Parsimonious representation:

$$Y_t = X_t\delta + E_t \qquad E_t \sim N(0, \Omega) \tag{51}$$
$$\delta = \Xi\lambda + u \qquad u \sim N(0, \Omega \otimes V) \tag{52}$$

where $\lambda = [\lambda_1, \lambda_2, \lambda_3, \ldots]'$.

• Idea: factorize the coefficient vector δ into components: λ is $s \times 1$ vector, $s << k * r$, Ξ_j are matrices with elements equal to zero or one.

• (52) can be intepreted as a shrinkage prior.

Example:

- λ_1 captures movements in δ common to all countries and variables (a 1×1 vector).

- λ_2 captures movements in δ common to all the variables of a country (a $N \times 1$ vector).

- λ_3 captures movements in δ in a variables across all countries ($G \times 1$ vector).

- λ_4 captures movements in δ specific to the exogenous variables (1×1 vector).

- etc.

- u captures unmodelled features of the coefficients vector.

Observable Index model

Using (52) into (51) we have

$$Y_t = \mathcal{Z}_{1t}\lambda_1 + \mathcal{Z}_{2t}\lambda_2 + \mathcal{Z}_{3t}\lambda_3 + \mathcal{Z}_{4t}\lambda_4 + v_t = \mathcal{Z}_t\lambda + v_t \tag{53}$$

where $\mathcal{Z}_{1t} = X_t\Xi_1, \mathcal{Z}_{2t} = X_t\Xi_2, \mathcal{Z}_{3t} = X_t\Xi_3, \mathcal{Z}_{4t} = X_t\Xi_4$, and $v_t = E_t + X_t u$.

- Regressors of (53) are different averages of lags of the VAR variables. Dynamically span lagged interdependencies between variables and countries.

- $\lambda_i, i = 1, 2, \dots$ are the factor loadings.

- $\mathcal{Z}_{it}$ easy to construct (they observable and correlated).

• Interpretation: $\mathcal{Z}_{1t}\lambda_1$ is a leading indicator of the common cycle,$\mathcal{Z}_{2t}\lambda_2$ is a leading indicator of country specific cycles.

• Indicators emphasize low frequency movements, since they are average of lags of VAR variables. Good for medium term forecasting.

• Analysis feasible with small T and small N and when degrees of freedom in Panel VAR small. Estimate loadings λ not VAR coefficients δ.

Example 11 $G = 2$ *variables,* $N = 2$ *countries,* 1 *lag, no exogenous:* δ_t *is a vector* 16×1. *Then*

$$\delta = \Xi_1 \lambda_1 + \Xi_2 \lambda_2 + \Xi_3 \lambda_3 + u$$

λ_1 *is scalar,* λ_2 *is* 2×1, λ_3 *is* 2×1, *and the VAR can be rewritten as*

$$\begin{bmatrix} y_t^1 \\ x_t^1 \\ y_t^2 \\ x_t^2 \end{bmatrix} = \begin{bmatrix} \mathcal{Z}_{1t} \\ \mathcal{Z}_{1t} \\ \mathcal{Z}_{1t} \\ \mathcal{Z}_{1t} \end{bmatrix} \lambda_1 + \begin{bmatrix} \mathcal{Z}_{2,1,t} & 0 \\ \mathcal{Z}_{2,1,t} & 0 \\ 0 & \mathcal{Z}_{2,2,t} \\ 0 & \mathcal{Z}_{2,2,t} \end{bmatrix} \lambda_2 + \begin{bmatrix} \mathcal{Z}_{3,1,t} & 0 \\ 0 & \mathcal{Z}_{3,2,t} \\ \mathcal{Z}_{3,1,t} & 0 \\ 0 & \mathcal{Z}_{3,2,t} \end{bmatrix} \lambda_3 + v_t$$

where $\mathcal{Z}_{1t} = y_{t-1}^1 + x_{t-1}^1 + y_{t-1}^2 + x_{t-1}^2$ *is the common information,* $\mathcal{Z}_{2,1,t} = y_{t-1}^1 + x_{t-1}^1$ *is country 1 information (across variables),* $\mathcal{Z}_{3,1,t} = y_{t-1}^1 + y_{t-1}^2$ *is variable* y *(across countries).*

- if λ_1 large relative to λ_2, y_t^1 and x_t^1 comove with y_t^2 and x_t^2.

- if $\lambda_1 = 0$: y_t^1 and x_t^1 may drift apart from y_t^2 and x_t^2.

- Here a leading indicator for Y_t based on the common information is $CLI_t = \mathcal{Z}_{1t}\lambda_1$; a leading indicators based on common and unit specific information is $CULI_t = \mathcal{Z}_{1t}\lambda_1 + \mathcal{Z}_{2t}\lambda_2$, etc.

How do you estimate the model?

• Form $\mathcal{Z}_{it}$ (these are simply linear combinations of right hand side variables).

Classical estimation:

• If $u = 0$ can use OLS to estimate the θ.

• If $u \neq 0$ need to do an heteroschedasticity correction since v_t depends on X_t

Bayesian estimation. Add a prior for λ.

• If g(λ) is characterized by fixed parameters and is conjugate: estimation is easy, see BVAR notes.

• If g(λ) is characterized by random parameters and but is conjugate, and if we assume that $var(u) \propto var(e_t)$, can use Gibbs sampler. Model:

$$\begin{aligned} Y_t &= \mathcal{Z}_t \lambda + v_t & & (54) \\ \lambda &= \lambda_0 + \eta & \eta \sim N(0, B) & \quad (55) \end{aligned}$$

Hierarchical model

$$Y_t = X_t\delta + E_t \qquad E_t \sim N(0, \Omega) \tag{56}$$

$$\delta = \Xi\lambda + u \qquad u \sim N(0, \Omega \otimes V) \tag{57}$$

$$\lambda = \lambda_0 + \eta \qquad \eta \sim N(0, B) \tag{58}$$

- E_t, u, η uncorrelated, $V = \sigma^2 I_k$, $B = diag(B_1, B_2, B_3, B_4)$.

- Here get posterior distribution of $(\Omega,\ \delta, \lambda, \sigma^2)$, given λ_0, B.

- Need prior densities for (Ω, σ^2) (choose them proper but loose).

Hierarchical TV coefficients panel VAR

$$
\begin{aligned}
Y_t &= X_t\delta_t + E_t \qquad E_t \sim N(0, \Omega) &(59)\\
\delta_t &= \Xi\lambda_t + u_t \qquad u_t \sim N(0, \Omega \otimes V) &(60)\\
\lambda_t &= \lambda_{t-1} + \eta_t \qquad \eta_t \sim N(0, B_t) &(61)
\end{aligned}
$$

- E_t, u_t, η_t uncorrelated, $V = \sigma^2 I_k$, B_t could be time-varying, e.g. $B_t = \gamma_1 B_{t-1} + \gamma_2 B_0$, with $B_0 = diag(B_{01}, B_{02}, B_{03}, B_{04})$.

- Same logic. Now get posterior distribution of $(\Omega, \{\delta_t\}_{t=1}^T, \{\lambda_t\}_{t=1}^T, \sigma^2)$.

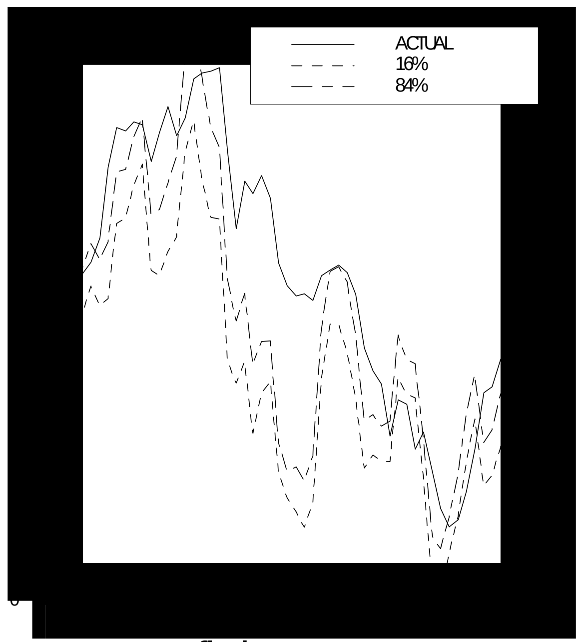
- Need prior densities for (Ω, B_0, λ_0) (choose them proper but loose).

- Treat the vector of time varying parameters $\{\delta_t\}_{t=1}^T$, $\{\lambda_t\}_{t=1}^T$as a new vector of parameters whose conditional posterior needs to be found.

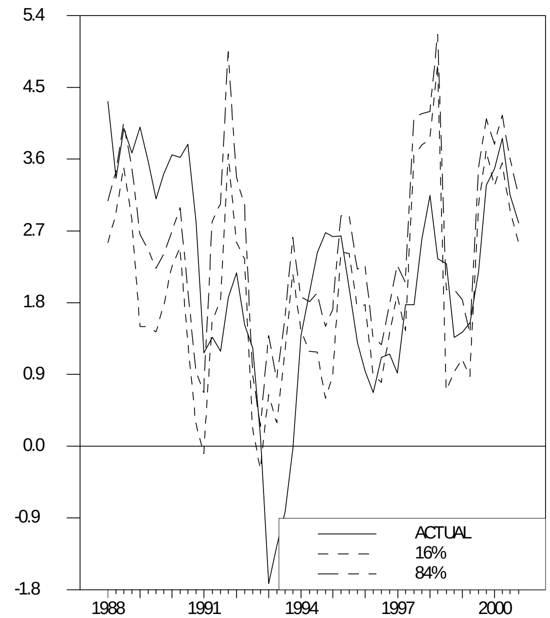
Example 12 *Use VAR model for G-7 countries with GDP growth, inflation, employment growth and the real exchange rate for each country. Specify: a* 2×1 *vector of common factors - (one EU and one non-EU), a* 7×1 *vector of country specific factors and a* 4×1 *vector of variables specific factors.*

Assume time variations in all factors, no exchangeable prior and non-informative priors on the hyperparameters. Posterior distributions one year in advance constructed recursively at each t. *Figure leading indicator 68% bands for EU GDP growth and inflation (with actual values).*

Leading indicator $=$ *sum of the three estimated components. Model predicts the ups and downs of both series well using one year ahead info. Theil-U for 1996:1-2000:4 and 1991:1-1995:4 are 0.87 and 0.66, much lower than single country BVAR (0.96, 0.94) or univariate AR(0.98,0.96).*



Inflation



GDP growth