Dynamic sparse factor model[a]

Sylvia Kaufmann

Oesterreichische Nationalbank and University of Basel

Econometrics, Time Series Analysis,

and Systems Theory

Institute for Advanced Studies

Vienna, June 18-20 2009

[a]Jointly with Christian Schumacher, Deutsche Bundesbank

Motivation

Static and dynamic factor model setting with large N.

- Involves dimension reduction, extract $k \ll N$ common factors
- Well known: Extract k factors by principal or frequency components.
  Stock and Watson (JBES 2002), Forni et al., (RES 2000)
- Our approach: Extract k factors in a state space framework
  Otrok and Whiteman (mimeo 2002, 2007),
  Kaufmann (EcJ, 2000)
- Related to: Bayesian regression with a large number of predictors
  De Mol et al. (CEPR 2006)

Contribution

- Extract the factors in a state space framework.
- Achieve dimension reduction by estimating a sparse matrix of factor loadings.
  Aguilar and West (JBES 2000),West (Bayesian Statistics 2003), Carvalho et al. ISDS 2005)
- Estimate sparse factor loading matrix independently of the time series ordering $\longrightarrow$ estimate sparsity and identify the system each under different parametrizations
  Koop et al. (Bayesian Econometrics 2008)

Model specification

I.

$$\underset{N\times 1}{X_t} = \underset{(N\times k)(k\times 1)}{\lambda f_t} + \xi_t$$

$$\underset{\text{of order } p}{\Phi(L) f_t} = \eta_t, \quad \eta_t \sim N\left(0, \Sigma_\eta\right) \qquad \lambda'\lambda = I,\ \Sigma_\eta \text{ free}$$

$$\underset{\text{of order } q}{\Psi(L)\xi_t} = \varepsilon_t, \quad \varepsilon_t \sim N\left(0, \Sigma_\varepsilon\right) \qquad \Psi(L) \text{ diagonal}$$

II.

$$X_t = \lambda^* f_t^* + \Psi(L)\xi_t$$

$$\Phi^*(L) f_t^* = \eta_t^*, \quad \eta_t^* \sim N\left(0, \Sigma_{\eta^*}\right) \quad \lambda^* \text{ sparse},\ \Sigma_{\eta^*} = I$$

Relation between specification I. and II.

$$\lambda f_t = (\lambda H)\left(H^{-1} f_t\right) = \lambda^* f_t^*$$

Relation between specification I. and II.

$$\lambda f_t = (\lambda H)\left(H^{-1} f_t\right) = \lambda^* f_t^*$$

From I. to II.: conditional on f^T, $H = \Sigma_\eta^{1/2}$, $f_t^* = H^{-1} f_t$

$\Phi^*(L) = H^{-1}\Phi(L)H$, $\Sigma_{\eta^*} = H^{-1}\Sigma_\eta H^{-1\prime} = I$

$\longrightarrow$ estimate/simulate sparse λ^*

$\longrightarrow$ ensure $\left(\sum_{i=1}^N I_{\{\lambda_{ij}^*>0\}} / \sum_{i=1}^N I_{\{\lambda_{ij}^* \neq 0\}}\right) > 0.5 \ \forall j$

Relation between specification I. and II.

$$\lambda f_t = (\lambda H)\left(H^{-1} f_t\right) = \lambda^* f_t^*$$

From I. to II.: conditional on f^T, $H = \Sigma_\eta^{1/2}$, $f_t^* = H^{-1} f_t$

$\Phi^*(L) = H^{-1}\Phi(L)H$, $\Sigma_{\eta^*} = H^{-1}\Sigma_\eta H^{-1\prime} = I$

$\longrightarrow$ estimate/simulate sparse λ^*

$\longrightarrow$ ensure $\left(\sum_{i=1}^N I_{\{\lambda_{ij}^*>0\}} / \sum_{i=1}^N I_{\{\lambda_{ij}^*\neq 0\}}\right) > 0.5 \ \forall j$

From II. to I.: conditional on λ^*, $H = \left(\lambda^{*\prime}\lambda^*\right)^{1/2}$

$\lambda = \lambda^* H^{-1}$, $\Phi(L) = H\Phi^*(L)H^{-1}$, $\Sigma_\eta = HH'$

$\longrightarrow$ estimate/simulate f^T under $\lambda'\lambda = I$

$\longrightarrow$ sparsity of λ^* implies sparsity of λ
(still to derive)

Determination of the factor structure

Covariance structure of the data (N large)

$$Cov(X) = \lambda\Phi(L)^{-1}\Sigma_\eta\Phi(L)^{-1'}\lambda' + \Psi(L)^{-1}\Sigma_\varepsilon\Psi(L)^{-1'}$$

Maximum number k of factors given by

$$k^2(p + \frac{1}{2}) + k(N + \frac{1}{2}) - N(N - 1 - 2q)/2 \geq 0$$

For example: $N = 100,\ p = q = 2,\ k \leq 27$

Advantage of switching between specifications

Identification following Geweke and Zhou (1996), Aguilar and West (2000), Carvalho (2006)

$$\lambda^* = \begin{bmatrix} \lambda^*_{11} & 0 & \dots & \dots & 0 \\ \lambda^*_{21} & \lambda^*_{22} & 0 & \dots & 0 \\ \vdots & \vdots & & & \vdots \\ \lambda^*_{k1} & \lambda^*_{k2} & \lambda^*_{k3} & \dots & \lambda^*_{kk} \\ \lambda^*_{k+1,1} & \lambda^*_{k+1,2} & \lambda^*_{k+1,3} & \dots & \lambda^* k+1, k \\ \vdots & \vdots & & & \vdots \\ \lambda^*_{N1} & \dots & \dots & & \lambda^*_{Nk} \end{bmatrix}$$

$\longrightarrow$ Needs a procedure to order the variables

Advantage of switching between specifications

Here: λ^* is estimated freely, formal identification given by $\lambda'\lambda = I$ $(\lambda = \lambda^* H^{-1})$

$$\lambda^* = \begin{bmatrix} \lambda^*_{11} & 0 & \dots & \dots & \lambda^*_{1k} \\ 0 & \lambda^*_{22} & 0 & \dots & 0 \\ \dots & \dots & 0 & \lambda^*_{k-1,k-1} & \lambda^*_{k-1,k} \\ 0 & \dots & \dots & 0 & \lambda^*_{kk} \\ \vdots & \vdots & & \vdots & \vdots \\ 0 & 0 & \dots & & 0 \\ \lambda^*_{N1} & \dots & \dots & & \lambda^*_{Nk} \end{bmatrix}$$

$\longrightarrow$ May need further procedure to interpret the factors

Designing probabilistic sparsity

Designing probabilistic sparsity

The Bayesian setup assumes independent priors for the parameters

$$\pi\left(\lambda, \Phi, \Psi, \Sigma_{\eta}, \Sigma_{\varepsilon}\right) = \pi\left(\lambda\right)\pi\left(\Phi\right)\pi\left(\Psi\right)\pi\left(\Sigma_{\eta}\right)\pi\left(\Sigma_{\varepsilon}\right)$$

where $\pi\left(\lambda\right)$ is implied by a sparse prior on λ^*
$(i = 1, \ldots, N,\ j = 1, \ldots, k)$:

Designing probabilistic sparsity

The Bayesian setup assumes independent priors for the parameters

$$\pi(\lambda, \Phi, \Psi, \Sigma_\eta, \Sigma_\varepsilon) = \pi(\lambda)\,\pi(\Phi)\,\pi(\Psi)\,\pi(\Sigma_\eta)\,\pi(\Sigma_\varepsilon)$$

where $\pi(\lambda)$ is implied by a sparse prior on λ^*
$(i = 1, \ldots, N,\ j = 1, \ldots, k)$:

$$\begin{aligned}
\pi\left(\lambda_{ij}^*\right) &= (1-\beta_{ij})\delta_0\left(\lambda_{ij}^*\right) + \beta_{ij} N(0, \tau_j) \\
\pi(\beta_{ij}) &= (1-\rho_j)\delta_0(\beta_{ij}) + \rho_j B(a_j b_j, a_j(1-b_j)) \\
\pi(\rho_j) &= B(r_j s_j, r_j(1-s_j))
\end{aligned}$$

$\delta_0(\cdot)$	: Dirac delta function at zero
$B((ab, a(1-b))$	: Beta distribution with mean b, variance $b(1-b)/(1+a)$
$1-\rho_j b_j$	: $E(1-\beta_{ij})$, prob. of zero loading on factor j

Bayesian setup

Bayesian setup

Define $X^t = (X_t, X_{t-1}, \ldots, X_1)$, $\theta = (\lambda, \Phi, \Psi, \Sigma_\eta, \Sigma_\varepsilon)$
Complete data likelihood:

$$L\left(X^T | f^t, \theta\right) = \prod_{t=1}^{T} \pi\left(X_t | f^t, \theta\right)$$

with

$$\pi\left(X_t | f^t, \theta\right) = \frac{1}{(2\pi)^{N/2} |\Sigma_\varepsilon|^{1/2}} \exp\left\{-1/2\left(\Psi(L)(X_t - \lambda f_t)\right)' \times \Sigma_\varepsilon^{-1}\left(\Psi(L)(X_t - \lambda f_t)\right)\right\}$$

Bayesian setup

Define $X^t = (X_t, X_{t-1}, \ldots, X_1)$, $\theta = (\lambda, \Phi, \Psi, \Sigma_\eta, \Sigma_\varepsilon)$
Complete data likelihood:

$$L\left(X^T | f^t, \theta\right) = \prod_{t=1}^{T} \pi\left(X_t | f^t, \theta\right)$$

with

$$\pi\left(X_t | f^t, \theta\right) = \frac{1}{(2\pi)^{N/2} |\Sigma_\varepsilon|^{1/2}} \exp\left\{-1/2\left(\Psi(L)(X_t - \lambda f_t)\right)' \times \Sigma_\varepsilon^{-1}\left(\Psi(L)(X_t - \lambda f_t)\right)\right\}$$

Prior density of unobserved factors

$$\pi\left(f^T | \theta\right) = \prod_{t=p+1}^{T} \pi\left(f_t | f^{t-1}, \theta\right) \pi\left(f^p | \theta\right)$$

Sampler for $\vartheta = \left(f^T, \theta\right)$

$$\pi\left(\vartheta | X^T\right) \propto L\left(X^T | f^T, \theta\right) \pi\left(f^T | \theta\right) \pi\left(\theta\right)$$

Sampler for $\vartheta = \left(f^T, \theta\right)$

$$\pi\left(\vartheta | X^T\right) \propto L\left(X^T | f^T, \theta\right) \pi\left(f^T | \theta\right) \pi\left(\theta\right)$$

Sampling steps

Sampler for $\vartheta = \left(f^T, \theta\right)$

$$\pi\left(\vartheta | X^T\right) \propto L\left(X^T | f^T, \theta\right) \pi\left(f^T | \theta\right) \pi(\theta)$$

Sampling steps

(i) Simulate f^T from $\pi\left(f^T | X^T, \theta\right)$ under $\lambda' \lambda = I$.

(ii) Simulate from $\pi\left(\theta_{-\lambda} | f^T, X^T, \lambda\right)$

Sampler for $\vartheta = \left(f^T, \theta\right)$

$$\pi\left(\vartheta | X^T\right) \propto L\left(X^T | f^T, \theta\right) \pi\left(f^T | \theta\right) \pi\left(\theta\right)$$

Sampling steps

(i) Simulate f^T from $\pi\left(f^T | X^T, \theta\right)$ under $\lambda'\lambda = I$.

(ii) Simulate from $\pi\left(\theta_{-\lambda} | f^T, X^T, \lambda\right)$

(iii) Transform the system to $f_t^* = H^{-1} f_t$ with $H = \Sigma_\eta^{1/2}$ and $\Phi^*(L) = H^{-1}\Phi(L)H$, $\Sigma_{\eta^*} = I$.
Simulate from $\pi\left(\lambda^* | f^{*T}, X^T, \Psi(L), \Sigma_\varepsilon\right)$ under a sparse prior.
Transform to $\lambda = \lambda^* H^{-1}$ with $H = \left(\lambda^{*\prime}\lambda^*\right)^{1/2}$,
$\Sigma_\eta = HH'$, $\Phi(L) = H\Phi^*(L)H^{-1}$.

Sampler for $\vartheta = \left(f^T, \theta\right)$

$$\pi\left(\vartheta | X^T\right) \propto L\left(X^T | f^T, \theta\right) \pi\left(f^T | \theta\right) \pi\left(\theta\right)$$

Sampling steps

(i) Simulate f^T from $\pi\left(f^T | X^T, \theta\right)$ under $\lambda'\lambda = I$.

(ii) Simulate from $\pi\left(\theta_{-\lambda} | f^T, X^T, \lambda\right)$

(iii) Transform the system to $f_t^* = H^{-1} f_t$ with $H = \Sigma_\eta^{1/2}$ and $\Phi^*(L) = H^{-1}\Phi(L)H$, $\Sigma_{\eta^*} = I$.
Simulate from $\pi\left(\lambda^* | f^{*T}, X^T, \Psi(L), \Sigma_\varepsilon\right)$ under a sparse prior.
Transform to $\lambda = \lambda^* H^{-1}$ with $H = \left(\lambda^{*\prime}\lambda^*\right)^{1/2}$, $\Sigma_\eta = HH'$, $\Phi(L) = H\Phi^*(L)H^{-1}$.

(iv) Update the hyperparameters of the sparse prior (τ_j, β_{ij}, ρ_j).

Application

Simulated data, Bai and Ng (2006), $N_1 = N_2 = 20$, $p = 1$, $q = 0$

$$\begin{aligned} f^*_{jt} &= 0.5 f^*_{j,t-1} + \eta^*_{jt}, \quad \eta^*_{jt} \sim N\left(0,1\right), \; j = 1,2 \\ N_1: \; X_{it} &= 0.8 f^*_{1t} + \varepsilon_{it} \quad \varepsilon_{it} \sim N\left(0, 0.35\right) \\ N_2: \; X_{it} &= 0.4 f^*_{1t} + 0.4 f^*_{2t} + \varepsilon_{it} \quad \varepsilon_{it} \sim N\left(0, 0.68\right) \end{aligned}$$

Estimates:

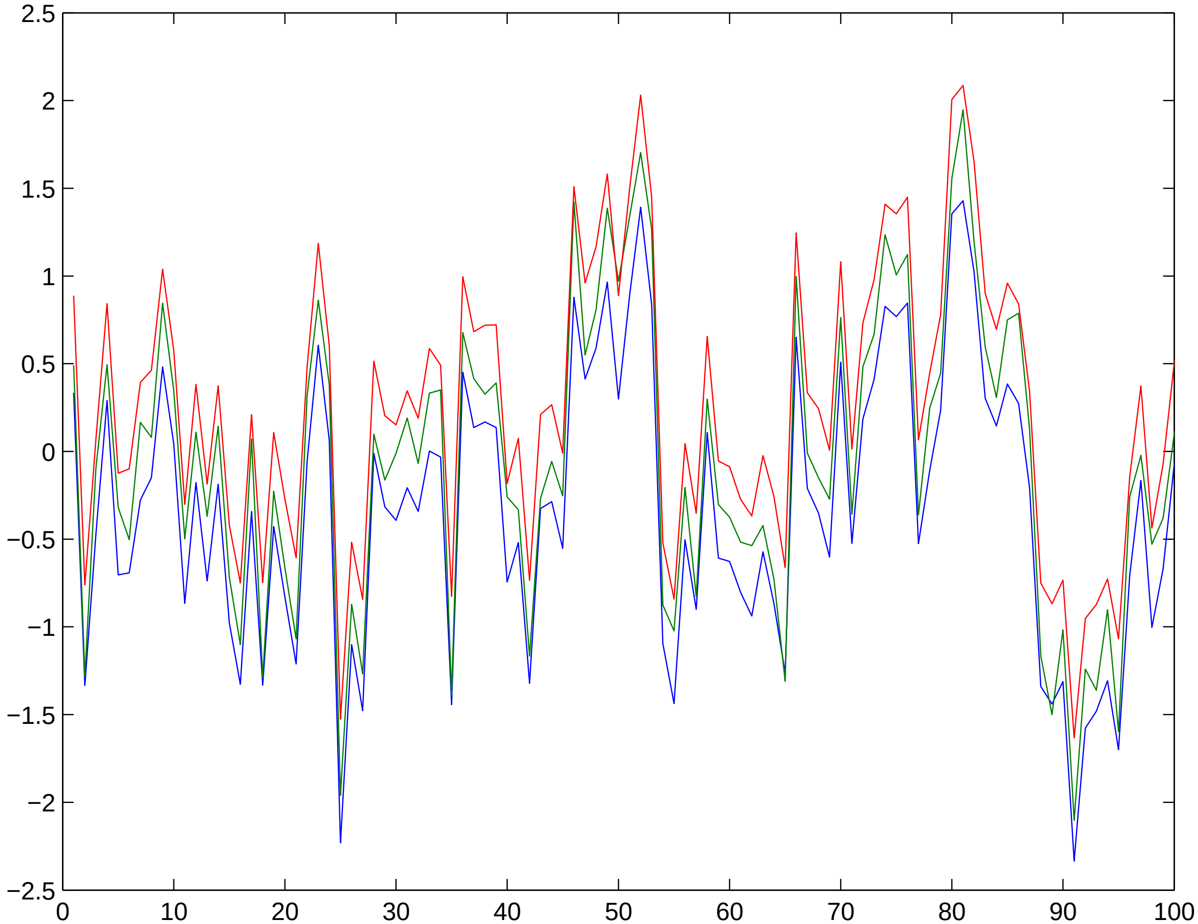
$$f_t^* = \begin{bmatrix} \underset{(0.05)}{0.49} & \underset{(0.03)}{0.01} \\ \underset{(0.08)}{0.11} & \underset{(0.06)}{0.53} \end{bmatrix} f_{t-1}^* + \eta_t^*$$

$$N_1: \; X_{it} = \underset{(0.10)}{0.53} \; f_{1t}^* + \underset{(0.10)}{0.50} \; f_{2t}^* + \varepsilon_{it} \quad \varepsilon_{it} \sim N\left(0, \; \underset{(0.06)}{0.37}\right)$$

$$N_2: \; X_{it} = \underset{(0.03)}{0.56} \; f_{1t}^* + \underset{(0.02)}{0.00} \; f_{2t}^* + \varepsilon_{it} \quad \varepsilon_{it} \sim N\left(0, \; \underset{(0.04)}{0.70}\right)$$

$$\begin{bmatrix} 0.8 & 0 \\ 0.4 & 0.4 \end{bmatrix} = \begin{bmatrix} 0.53 & 0.50 \\ 0.56 & 0.00 \end{bmatrix} \begin{bmatrix} 0.71 & 0.71 \\ 0.84 & -0.76 \end{bmatrix}$$

Estimated common component of X_1,
(green: simulated,
red/blue: +/- 2 standard deviations in estimated component)



Conclusion

Conclusion

- Suggestion to estimate sparse dynamic factor model in the state space framework
- Identification: $\lambda'\lambda = I$, Σ_η free

  Sparse estimation λ^* free, $\Sigma_{\eta^*} = I$

  $\longrightarrow \lambda = \lambda^* H^{-1}$, $H = (\lambda^{*\prime}\lambda^*)^{1/2}$

Conclusion

- Suggestion to estimate sparse dynamic factor model in the state space framework

- Identification: $\lambda'\lambda = I$, Σ_η free

  Sparse estimation λ^* free, $\Sigma_{\eta^*} = I$

  $\longrightarrow \lambda = \lambda^* H^{-1}$, $H = (\lambda^{*\prime}\lambda^*)^{1/2}$

- Way of circumventing the problem raised in Boivin and Ng (2006)? Improve in forecasting with many data $\rightarrow$ still to do

- What is needed: Posterior transformation according to a structural interpretation