Dynamic Factor Models and Factor Augmented Vector Autoregressions

Lawrence J. Christiano

Dynamic Factor Models and Factor Augmented Vector Autoregressions

- Problem:
  - the time series dimension of data is relatively short.
  - the number of time series variables is huge.
- DFM's and FAVARs take the position:
  - there are many variables and, hence, shocks,
  - but, the principle driving force of all the variables may be just a small number of shocks.
- Factor view has a long-standing history in macro.
  - almost the *definition* of macroeconomics: a handfull of shocks - demand, supply, etc. - are the principle economic drivers.
  - Sargent and Sims: only two shocks can explain a large fraction of the variance of US macroeconomic data.
    - 1977, "Business Cycle Modeling Without Pretending to Have Too Much A-Priori Economic Theory," in *New Methods in Business Cycle Research*, ed. by C. Sims et al., Minneapolis: Federal Reserve Bank of Minneapolis.

Why Work with a Lot of Data?

- Estimates of impulse responses to, say, a monetary policy shock, may be distorted by not having enough data in the analysis (Bernanke, et. al. (QJE, 2005))
  - Price puzzle:
    - measures of inflation tend to show transitory rise to a monetary policy tightening shock in standard (small-sized) VARs.
    - One interpretation: Monetary authority responds to a signal about future inflation that is captured in data not included in a standard, small-sized VAR.
- May suppose that 'core inflation' is a factor that can only be deduced from a large number of different data.
- May want to know (as in Sargent and Sims), whether the data for one country or a collection of countries can be characterized as the dynamic response to a few factors.

Outline

- Describe Dynamic Factor Model
  - Identification problem and one possible solution.
- Derive the likelihood of the data and the factors.
- Describe priors, joint distribution of data, factors and parameters.
- Go for posterior distribution of parameters and factors.
  - Gibbs sampling, a type of MCMC algorithm.
  - Metropolis-Hastings could be used here, but would be very inefficient.
  - Gibbs exploits power of Kalman smoother algorithm and the type of fast 'direct sampling' done with BVARS.
- FAVAR

Dynamic Factor Model

- Let Y_t denote an $n \times 1$ vector of observed data
- Y_t related to $\kappa \ll n$ unobserved factors, f_t, by *measurement* (or, *observer*) equation:

$$y_{i,t} = a_i + \overbrace{\lambda_i'}^{\text{vector of } \kappa \text{ factor loadings}} f_t + \overbrace{\xi_{i,t}}^{\text{idiosyncratic component of } y_{i,t}}.$$

- Law of motion of factors:

$$f_t = \phi_{0,1} f_{t-1} + ... + \phi_{0,q} f_{t-q} + u_{0,t},\ u_{0,t} \sim \mathcal{N}(0, \Sigma_0).$$

- Idiosyncratic shock to $y_{i,t}$ ('measurement error'):

$$\xi_{i,t} = \phi_{i,1}\xi_{i,t-1} + ... + \phi_{i,p_i}\xi_{i,t-p_i} + u_{i,t},\ u_{i,t} \sim \mathcal{N}\left(0, \sigma_i^2\right).$$

- $u_{i,t}$, $i = 0, ..., n$, drawn independently from each other and over time.
- For convenience:

$$p_i = p, \text{ for all } i,\ q \leq p + 1.$$

Notation for Observer Equation

- Observer equation:

$$\begin{aligned} y_{i,t} &= a_i + \lambda_i' f_t + \xi_{i,t} \\ \xi_{i,t} &= \phi_{i,1}\xi_{i,t-1} + ... + \phi_{i,p_i}\xi_{i,t-p_i} + u_{i,t},\ u_{i,t} \sim \mathcal{N}\left(0, \sigma_i^2\right). \end{aligned}$$

- Let θ_i denote the parameters of the i^{th} observer equation:

$$\underbrace{\theta_i}_{(2+\kappa+p)\times 1} = \begin{bmatrix} \sigma_i^2 \\ a_i \\ \lambda_i \\ \phi_i \end{bmatrix},\ \phi_i = \begin{bmatrix} \phi_{i,1} \\ \vdots \\ \phi_{i,p} \end{bmatrix},\ i = 1, ..., n.$$

Notation for Law of Motion of Factors

- Factors:

$$f_t = \phi_{0,1} f_{t-1} + ... + \phi_{0,q} f_{t-q} + u_{0,t}, \;\; u_{0,t} \sim \mathcal{N}\left(0, \Sigma_0\right).$$

- Let θ_0 denote the parameters of factors:

$$\underbrace{\theta_0}_{\kappa(q+1)\times\kappa} = \begin{bmatrix} \Sigma_0 \\ \phi_0 \end{bmatrix}, \underbrace{\phi_0}_{\kappa q\times\kappa} = \begin{bmatrix} \phi_{0,1} \\ \vdots \\ \phi_{0,q} \end{bmatrix}$$

- All model parameters:

$$\theta = [\theta_0, \theta_1, ..., \theta_n]$$

Identification Problem in DFM

- DFM:

$$
\begin{aligned}
y_{i,t} &= a_i + \lambda_i' f_t + \xi_{i,t} \\
f_t &= \phi_{0,1} f_{t-1} + ... + \phi_{0,q} f_{t-q} + u_{0,t}, \ u_{0,t} \sim \mathcal{N}\left(0, \Sigma_0\right) \\
\xi_{i,t} &= \phi_{i,1} \xi_{i,t-1} + ... + \phi_{i,p} \xi_{i,t-p} + u_{i,t}.
\end{aligned}
$$

- Suppose H is an arbitrary invertible $\kappa \times \kappa$ matrix.
  - Above system is observationally equivalent to:

$$
\begin{aligned}
y_{i,t} &= a_i + \tilde{\lambda}_i' \tilde{f}_t + \xi_{i,t} \\
\tilde{f}_t &= \tilde{\phi}_{0,1} \tilde{f}_{t-1} + ... + \tilde{\phi}_{0,q} \tilde{f}_{t-q} + \tilde{u}_{0,t} \sim \mathcal{N}\left(0, \tilde{\Sigma}_0\right),
\end{aligned}
$$

  where

$$
\tilde{f}_t = H f_t, \ \tilde{\lambda}_i' = \lambda_i' H^{-1}, \ \tilde{\phi}_{0,j} = H \phi_{0,j} H^{-1}, \ \tilde{\Sigma}_0 = H \Sigma_0 H', \ .
$$

- Desirable to restrict model parameters so that there is no change of parameters that leaves the system observationally equivalent, yet has all different factors and parameter values.

Geweke-Zhou (1996) Identification

- Note for any model parameterization, can always choose an H so that $\Sigma_0 = I_\kappa$.
  - Find C such that $CC' = \Sigma_0$ (there is a continuum of these), set $H = C^{-1}$.
- Geweke-Zhou (1996) suggest the identifying assumption, $\Sigma_0 = I_\kappa$.
  - But, this is not enough to achieve identification.
  - Exists a continuum of orthonormal matrices with property, $CC' = I_\kappa$.
    - Simple example: for $\kappa = 2$, for each $\omega \in [-\pi, \pi]$,

$$C = \begin{bmatrix} \cos(\omega) & \sin(\omega) \\ -\sin(\omega) & \cos(\omega) \end{bmatrix}, \; 1 = \cos^2(\omega) + \sin^2(\omega)$$

  - For each C, set $H = C^{-1} = C'$. That produces an observationally equivalent alternative parameterization, while leaving intact the normalization, $\Sigma_0 = I_\kappa$, since $H\Sigma_0 H' = C'C = C^{-1}C = I_\kappa$.

Geweke-Zhou (1996) Identification

- Write:

$$\Lambda = \begin{bmatrix} \lambda_1 \\ \vdots \\ \lambda_\kappa \\ \lambda_{\kappa+1} \\ \vdots \\ \lambda_n \end{bmatrix} = \begin{bmatrix} \Lambda_{1,\kappa} \\ \Lambda_{2,\kappa} \end{bmatrix}, \ \Lambda_{1,\kappa} \sim \kappa \times \kappa$$

- Geweke-Zhou also require $\Lambda_{1,\kappa}$ is lower triangular.
  - then, in simple example, only orthonormal matrix C that preserves lower triangular $\Lambda_{1,\kappa}$ is lower triangular (i.e., $b = 0$, $a = \pm 1$).
- Geweke-Zhou resolve identification problem with last assumption: diagonal elements of $\Lambda_{1,\kappa}$ non-negative (i.e., $a = 1$ in example).

Geweke-Zhou (1996) Identification

- Identifying restrictions: $\Lambda_{1,\kappa}$ is lower triangular, $\Sigma_0 = I_\kappa$.
  - Only first factor, $f_{1,t}$, affects first variable, $y_{1,t}$.
  - Only $f_{1,t}$ and $f_{2,t}$ affect $y_{2,t}$, etc.

- Ordering of y_{it} affects the interpretation of the factors.

- Alternative identifications:
  - Σ_0 diagonal and diagonal elements of $\Lambda_{1,\kappa}$ equal to unity.
  - Σ_0 unrestricted (positive definite) and $\Lambda_{1,\kappa} = I_k$.

Next:

- Move In direction of using data to obtain posterior distribution of parameters and factors.

- Start by going after the likelihood.

Likelihood of Data and Factors

- System, $i = 1, ..., n$:

$$\begin{aligned} y_{i,t} &= a_i + \lambda_i' f_t + \xi_{i,t} \\ f_t &= \phi_{0,1} f_{t-1} + ... + \phi_{0,q} f_{t-q} + u_{0,t}, \; u_{0,t} \sim \mathcal{N}(0, \Sigma_0) \\ \xi_{i,t} &= \phi_{i,1} \xi_{i,t-1} + ... + \phi_{i,p} \xi_{i,t-p} + u_{i,t}. \end{aligned}$$

- Define:

$$\phi_i(L) = \phi_{i,1} + ... + \phi_{i,p} L^{p-1}, \; Lx_t \equiv x_{t-1}.$$

- Then, the *quasi-differenced observer equation* is:

$$\begin{aligned} [1 - \phi_i(L) L] y_{i,t} &= [1 - \phi_i(1)] a_i + \lambda_i' [1 - \phi_i(L) L] f_t \\ &\quad + \overbrace{[1 - \phi_i(L) L] \xi_{i,t}}^{u_{i,t}} \end{aligned}$$

Likelihood of Data and of Factors

- Quasi-differenced observer equation:

$$y_{i,t} = \phi_i(L)\, y_{i,t-1} + [1 - \phi_i(1)]\, a_i + \lambda_i' [1 - \phi_i(L)\, L] f_t + u_{i,t}$$

- Consider the MATLAB notation:

$$x_{t_1:t_2} \equiv x_{t_t}, ..., x_{t_2}.$$

- Note: $y_{i,t}$, conditional on $y_{i,t-p:t-1}, f_{t-p:t}, \theta_i$, is Normal:

$$\begin{aligned} & p\left(y_{i,t} | y_{i,t-p:t-1}, f_{t-p:t}, \theta_i\right) \\ \sim \; & \mathcal{N}\left(\phi_i(L)\, y_{i,t-1} + [1 - \phi_i(1)]\, a_i + \lambda_i' [1 - \phi_i(L)\, L] f_t, \sigma_i^2\right) \end{aligned}$$

Likelihood of Data and of Factors

- Independence of $u_{i,t}$'s implies the conditional density of $Y_t = \left[\begin{array}{ccc} y_{1,t} & \cdots & y_{n,t} \end{array}\right]'$:

$$\prod_{i=1}^{n} p\left(y_{i,t} | y_{i,t-p:t-1}, f_{t-p:t}, \theta_i\right).$$

- Density of f_t conditional on $f_{t-q:t-1}$:

$$p\left(f_t | f_{t-q:t-1}, \theta_0\right).$$

- Conditional joint density of Y_t, f_t :

$$\prod_{i=1}^{n} p\left(y_{i,t} | y_{i,t-p:t-1}, f_{t-p:t}, \theta_i\right) p\left(f_t | f_{t-q:t-1}, \theta_0\right).$$

Likelihood of Data and of Factors

- Likelihood of $Y_{p+1:T}, f_{p+1:T}$, conditional on initial conditions:

$$
\begin{aligned}
& p\left(Y_{p+1:T}, f_{p+1:T} | Y_{1:p}, f_{p-q:p}, \theta\right) \\
= & \prod_{t=p+1}^{T}\left[\prod_{i=1}^{n} p\left(y_{i,t} | y_{i,t-p:t-1}, f_{t-p:t}, \theta_i\right) p\left(f_t | f_{t-q:t-1}, \theta_0\right)\right]
\end{aligned}
$$

- Likelihood of initial conditions:

$$
\begin{aligned}
& p\left(Y_{1:p}, f_{p-q+1:p} | \theta\right) \\
= & p\left(Y_{1:p} | f_{p-q+1:p}, \theta\right) p\left(f_{p-q+1:p} | \theta_0\right)
\end{aligned}
$$

- Likelihood of $Y_{1:T}, f_{p-q:T}$ conditional on parameters only, θ :

$$
\begin{aligned}
& \prod_{t=p+1}^{T}\left[\prod_{i=1}^{n} p\left(y_{i,t} | y_{i,t-p:t-1}, f_{t-p:t}, \theta_i\right) p\left(f_t | f_{t-q:t-1}, \theta_0\right)\right] \\
& \times p\left(Y_{1:p} | f_{p-q+1:p}, \theta_i,\ i = 1, .., n\right) p\left(f_{p-q+1:p} | \theta_0\right)
\end{aligned}
$$

Joint Density of Data, Factors and Parameters

- Parameter priors: $p(\theta_i)$, $i = 0, ..., n$.
- Joint density of $Y_{1:T}$, $f_{p-q:T}$, θ :

$$\prod_{t=p+1}^{T} p\left(f_t | f_{t-q:t-1}, \theta_0\right) \prod_{i=1}^{n} p\left(y_{i,t} | y_{i,t-p:t-1}, f_{t-p:t}, \theta_i\right)$$

$$\times \left[\prod_{i=1}^{n} p\left(y_{i,1:p} | f_{p-q+1:p}, \theta\right) p\left(\theta_i\right)\right] p\left(f_{p-q+1:p} | \theta_0\right) p\left(\theta_0\right)$$

- From here on, drop the density of initial observations.
  - if T is not too small, then has no effect on results.
  - BVAR lecture notes describe an example of how to *not* ignore initial conditions; for general discussion, see Del Negro and Otrok (forthcoming, RESTAT, "Dynamic Factor Models with Time-Varying Parameters: Measuring Changes in International Business Cycles").

Outline

- Describe Dynamic Factor Model (done!)
  - Identification problem and one possible solution.
- Derive the likelihood of the data and the factors. (done!)
- Describe priors, joint distribution of data, factors and parameters. (done!)
- Go for posterior distribution of parameters and factors.
  - Gibbs sampling, a type of MCMC algorithm.
  - Metropolis-Hastings could be used here, but would be very inefficient.
  - Gibbs exploits power of Kalman smoother algorithm and the type of fast 'direct sampling' done with BVARS.
- FAVAR

Gibbs Sampling

- Idea is similar to what we did with the Metropolis-Hastings algorithm.

Gibbs Sampling versus Metropolis-Hastings

- Metropolis-Hastings: we needed to compute the posterior distribution of parameters, θ, conditional on the data.
  - output of Metropolis-Hastings algorithm: sequence of values of θ whose distribution corresponds to the posterior distribution of θ given the data:

$$\mathcal{P} = \left[\begin{array}{ccc} \theta^{(1)} & \cdots & \theta^{(M)} \end{array} \right]$$

- Gibbs sampling algorithm: sequence of values of DFM model parameters, θ, and unobserved factors, f, whose distribution corresponds to the posterior distribution conditional on the data:

$$\mathcal{P} = \left[\begin{array}{ccc} \theta^{(1)} & \cdots & \theta^{(M)} \\ f^{(1)} & \cdots & f^{(M)} \end{array} \right].$$

  Histogram of elements in individual rows of $\mathcal{P}$ represent marginal distribution of corresponding parameter or factor.

Gibbs Sampling Algorithm

- Computes sequence:

$$\mathcal{P} = \begin{bmatrix} \theta^{(1)} & \cdots & \theta^{(M)} \\ f^{(1)} & \cdots & f^{(M)} \end{bmatrix} = \begin{bmatrix} \mathcal{P}_1 & \cdots & \mathcal{P}_M \end{bmatrix}.$$

- Given $\mathcal{P}_{s-1}$ compute $\mathcal{P}_s$ in two steps.
  - Step 1: draw $\theta^{(s)}$ given $\mathcal{P}_{s-1}$ (direct sampling, using approach for BVAR)
  - Step 2: draw $f^{(s)}$ given $\theta^{(s)}$ (direct sampling, based on information from Kalman smoother).

Step 1: Drawing Model Parameters

- Parameters, θ

$$\begin{array}{rl} \text{observer equation:} & a_i,\ \lambda_i \\ \text{measurement error:} & \sigma_i^2,\ \phi_i \\ \text{law of motion of factors:} & \phi_0. \end{array}$$

where the identification, $\Sigma_0 = I$, is imposed.

  – Algorithm must be adjusted if some other identification is used.

- For each i :
  – Draw $a_i, \lambda_i, \sigma_i^2$ from Normal-Inverse Wishart, conditional on the $\phi_i^{(s-1)}$'s.
  – Draw ϕ_i from Normal, given a_i, λ_i, σ_i^2.

Drawing Observer Equation Parameters and Measurement Error Variance

- The joint density of $Y_{1:T}, f_{p-q:T}, \theta$:

$$\prod_{t=p+1}^{T}\left[p\left(f_t|f_{t-q:t-1},\theta_0\right)\prod_{i=1}^{n}p\left(y_{i,t}|y_{i,t-p:t-1},f_{t-p:t},\theta_i\right)\right]$$
$$\times p\left(\theta_0\right)\prod_{i=1}^{n}p\left(\theta_i\right),$$

  was derived earlier (but we have now dropped the densities associated with the initial conditions).

- Recall,

$$p\left(A|B\right)=\frac{p\left(A,B\right)}{p\left(B\right)}=\frac{p\left(A,B\right)}{\int_A p\left(A,B\right)dA}$$

Drawing Observer Equation Parameters and Measurement Error Variance

- Conditional density of θ_i obtained by dividing joint density by itself, after integrating out θ_i :

$$p\left(\theta_i | Y_{1:T}, f_{p-q:T}, \{\theta_j\}_{j \neq i}\right) = \frac{p\left(Y_{1:T}, f_{p-q:T}, \theta\right)}{\int_{\theta_i} p\left(Y_{1:T}, f_{p-q:T}, \theta\right) d\theta_i}$$

$$\propto p\left(\theta_i\right) \prod_{t=p+1}^{T} p\left(y_{i,t} | y_{i,t-p:t-1}, f_{t-p:t}, \theta_i\right)$$

here, we have taken into account that the numerator and denominator have many common terms.

- We want to draw $\theta_i^{(s)}$ from this posterior distribution for θ_i. Gibbs sampling procedure:
  - first, draw a_i, λ_i, σ_i^2 taking the other elements of θ_i from $\theta_i^{(s-1)}$.
  - then, draw other elements of θ_i taking a_i, λ_i, σ_i^2 as given.

Drawing Observer Equation Parameters and Measurement Error Variance

- The quasi-differenced observer equation:

$$\overbrace{y_{i,t} - \phi_i(L)\, y_{i,t-1}}^{\tilde{y}_{i,t}} = (1 - \phi_i(1))\, a_i + \lambda_i' \overbrace{[1 - \phi_i(L)\, L] f_t}^{\tilde{f}_{i,t}} + u_{i,t},$$

or,

$$\tilde{y}_{i,t} = [1 - \phi_i(1)]\, a_i + \lambda_i' \tilde{f}_{i,t} + u_{i,t}.$$

- Let

$$A_i = \begin{bmatrix} a_i \\ \lambda_i \end{bmatrix},\ x_{i,t} = \begin{bmatrix} (1 - \phi_i(1)) \\ \tilde{f}_{i,t} \end{bmatrix},$$

so

$$\tilde{y}_{i,t} = A_i' x_{i,t} + u_{i,t},$$

where $\tilde{y}_{i,t}$ and x_t are known, conditional on $\phi_i^{(s-1)}$.

Drawing Observer Equation Parameters and Measurement Error Variance

- From the Normality of the observer equation error:

$$
\begin{aligned}
& p\left(y_{i,t} | y_{i,t-p:t-1}, f_{t-p:t}, \theta_i\right) \\
\propto \quad & \frac{1}{\sigma_i} \exp\left\{-\frac{1}{2} \frac{\left(y_{i,t} - \left[\phi_i\left(L\right) y_{i,t-1} + A_i' x_{i,t}\right]\right)^2}{\sigma_i^2}\right\} \\
= \quad & \frac{1}{\sigma_i} \exp\left\{-\frac{1}{2} \frac{\left(\tilde{y}_{i,t} - A_i' x_{i,t}\right)^2}{\sigma_i^2}\right\}
\end{aligned}
$$

- Then,

$$
\begin{aligned}
& \prod_{t=p+1}^{T} p\left(y_{i,t} | y_{i,t-p:t-1}, f_{t-p:t}, \theta_i\right) \\
\propto \quad & \frac{1}{\sigma_i^{T-p}} \exp\left\{-\frac{1}{2} \sum_{t=p+1}^{T} \frac{\left(\tilde{y}_{i,t} - A_i' x_{i,t}\right)^2}{\sigma_i^2}\right\}.
\end{aligned}
$$

Drawing Observer Equation Parameters and Measurement Error Variance

- As in the BVAR analysis, express in matrix terms:

$$\begin{aligned} p\left(y_i | y_{i,1:p}, f_{1:T}, \theta_i\right) &\propto \frac{1}{\sigma_i^{T-p}} \exp\left\{-\frac{1}{2}\sum_{t=p+1}^{T}\frac{\left(\tilde{y}_{i,t} - A_i' x_t\right)^2}{\sigma_i^2}\right\} \\ &= \frac{1}{\sigma_i^{T-p}} \exp\left\{-\frac{1}{2}\frac{\left[y_i - X_i A_i\right]'\left[y_i - X_i A_i\right]}{\sigma_i^2}\right\} \end{aligned}$$

  where

$$f_{p+1:T} = f^{(s-1)},\ y_i = \begin{bmatrix} \tilde{y}_{i,p+1} \\ \vdots \\ \tilde{y}_{i,T} \end{bmatrix},\ X_i = \begin{bmatrix} x'_{i,p+1} \\ \vdots \\ x'_{i,T} \end{bmatrix},$$

  where $f_{q-p:p}$ fixed (could set to unconditional mean of zero).

- Note: calculations are conditional on factors, $f^{(s-1)}$, from previous Gibbs sampling iteration.

Including Dummy Observations

- As in the BVAR analysis, $\bar{T}$ dummy equations are one way to represent priors, $p(\theta_i)$:

$$p(\theta_i) \prod_{t=p+1}^{T} p\left(y_{i,t} | y_{i,t-p:t-1}, f_{t-p:t}, \theta_i\right)$$

- Dummy observations (can include restriction that $\Lambda_{1,\kappa}$ is lower triangular by suitable construction of dummies)

$$\begin{aligned} \bar{y}_i &= \bar{X}_i A_i + \bar{U}_i, \\ \bar{U}_i &= \begin{bmatrix} u_{i,1} \\ \vdots \\ u_{i,\bar{T}} \end{bmatrix}. \end{aligned}$$

- Stack the dummies with the actual data:

$$\underbrace{\underline{y}_i}_{(T-p+\bar{T})\times 1} = \begin{bmatrix} y_i \\ \bar{y}_i \end{bmatrix}, \quad \underbrace{\underline{X}_i}_{(T-p+\bar{T})\times(1+\kappa)} = \begin{bmatrix} X_i \\ \bar{X}_i \end{bmatrix}.$$

Including Dummy Observations

- As in BVAR:

$$p\left(y_i | y_{i,1:p}, f_{1:T}, \theta_i\right) p\left(\lambda_i, a_i | \sigma_i^2\right)$$

$$\propto \frac{1}{\sigma_i^{T+\bar{T}-p}} \exp\left\{ -\frac{1}{2} \frac{\left[\underline{y}_i - \underline{X}_i A_i\right]' \left[\underline{y}_i - \underline{X}_i A_i\right]}{\sigma_i^2} \right\}$$

$$= \frac{1}{\sigma_i^{T+\bar{T}-p}} \exp\left\{ -\frac{1}{2} \frac{\underline{S} + \left(A_i - \underline{A}_i\right)' \underline{X}_i' \underline{X}_i \left(\underline{A}_i - A_i\right)}{\sigma_i^2} \right\}$$

$$= \frac{1}{\sigma_i^{T+\bar{T}-p}} \exp\left\{ -\frac{1}{2} \frac{\underline{S}}{\sigma_i^2} \right\} \exp\left\{ -\frac{1}{2} \frac{\left(A_i - \underline{A}_i\right)' \underline{X}_i' \underline{X}_i \left(\underline{A}_i - A_i\right)}{\sigma_i^2} \right\}$$

where

$$\underline{S} = \left[\underline{y}_i - \underline{X}_i \underline{A}_i\right]' \left[\underline{y}_i - \underline{X}_i \underline{A}_i\right], \; \underline{A}_i = \left(\underline{X}_i' \underline{X}_i\right)^{-1} \underline{X}_i' \underline{y}_i.$$

Inverse Wishart Distribution

- Scalar version of Inverse Wishart distribution with (i.e., $m = 1$ in BVAR discussion) :

$$p\left(\sigma_i^2\right) = \frac{\left|S^*\right|^{\nu/2}}{2^\nu \Gamma\left[\frac{\nu}{2}\right]} \left|\sigma_i^2\right|^{-\frac{\nu+2}{2}} \exp\left\{-\frac{S^*}{2\sigma_i^2}\right\},$$

  degrees of freedom, ν, and shape, S^* (Γ denotes the Gamma function).

- Easy to verify (after collecting terms), that

$$\begin{aligned} & p\left(y_i | y_{i,1:p}, f_{1:T}, \theta_i\right) p\left(\lambda_i, a_i | \sigma_i^2\right) p\left(\sigma_i^2\right) \\ = \; & \mathcal{N}\left(\underline{A}_i, \sigma_i^2 \left(\underline{X}_i' \underline{X}\right)^{-1}\right) \\ & \times \mathcal{IW}\left(\nu + T - p + \bar{T} - (\kappa + 1), \underline{S} + S^*\right). \end{aligned}$$

- Direct sampling from posterior of distribution:
  - draw σ_i^2 from $\mathcal{IW}$. Then, draw A_i from $\mathcal{N}$, given σ_i^2

Draw Distributed Lag Coefficients in Measurement Error Law of Motion

- Given $\lambda_i, a_i, \sigma_i^2$, draw ϕ_i.
- Observer equation and measurement error process:

$$\begin{aligned} y_{i,t} &= a_i + \lambda_i' f_t + \xi_{i,t} \\ \xi_{i,t} &= \phi_{i,1}\xi_{i,t-1} + ... + \phi_{i,p}\xi_{i,t-p} + u_{i,t}. \end{aligned}$$

- Conditional on a_i, λ_i and the factors, $\xi_{i,t}$ can be computed from

$$\xi_{i,t} = y_{i,t} - a_i - \lambda_i' f_t,$$

so the measurement error law of motion can be written,

$$\xi_{i,t} = A_i' x_{i,t} + u_{i,t},\ A_i = \phi_i = \begin{bmatrix} \phi_{i,1} \\ \vdots \\ \phi_{i,p} \end{bmatrix},\ x_{i,t} = \begin{bmatrix} \xi_{i,t-1} \\ \vdots \\ \xi_{i,t-p} \end{bmatrix}$$

Draw Distributed Lag Coefficients in Measurement Error Law of Motion

- The likelihood of $\xi_{i,t}$ conditional on $x_{i,t}$ is

$$
\begin{aligned}
p\left(\xi_{i,t}|x_{i,t},\phi_i,\sigma_i^2\right) &= \mathcal{N}\left(A_i' x_{i,t},\sigma_i^2\right) \\
&= \frac{1}{\sigma_i}\exp\left\{-\frac{1}{2}\frac{\left(\xi_{i,t}-A_i' x_{i,t}\right)^2}{\sigma_i^2}\right\},
\end{aligned}
$$

  where σ_i^2, drawn previously, is for present purposes treated as known.

- Then, the likelihood of $\xi_{i,p+1},...,\xi_{i,T}$ is

$$
\begin{aligned}
& p\left(\xi_{i,p+1:T}|x_{i,p+1},\phi_i,\sigma_i^2\right) \\
\propto\ & \frac{1}{\left(\sigma_i\right)^{T-p}}\exp\left\{-\frac{1}{2}\sum_{t=p+1}^{T}\frac{\left(\xi_{i,t}-A_i' x_{i,t}\right)^2}{\sigma_i^2}\right\}
\end{aligned}
$$

Draw Distributed Lag Coefficients in Measurement Error Law of Motion

$$\sum_{t=p+1}^{T} \left(\xi_{i,t} - A_i' x_{i,t}\right)^2 = \left[y_i - X_i A_i\right]' \left[y_i - X_i A_i\right],$$

where

$$y_i = \begin{bmatrix} \xi_{i,p+1} \\ \vdots \\ \xi_{i,T} \end{bmatrix},\ X_i = \begin{bmatrix} x_{i,p+1}' \\ \vdots \\ x_{i,T}' \end{bmatrix},$$

Draw Distributed Lag Coefficients in Measurement Error Law of Motion

- If we impose priors by dummies, then

$$
\begin{aligned}
& p\left(\xi_{i,p+1:T} | x_{i,p+1}, \phi_i, \sigma_i^2\right) p\left(\phi_i\right) \\
\propto \; & \frac{1}{\left(\sigma_i\right)^{T-p}} \exp\left\{ -\frac{1}{2} \frac{\left[\underline{y}_i - \underline{X}_i A_i\right]' \left[\underline{y}_i - \underline{X}_i A_i\right]}{\sigma_i^2} \right\},
\end{aligned}
$$

  where $\underline{y}_i$ and $\underline{X}_i$ represents the stacked data that includes dummies.

- By Bayes' rule,

$$
p\left(\phi_i | \xi_{i,p+1:T}, x_{i,p+1}, \phi_i, \sigma_i^2\right) = \mathcal{N}\left(\underline{A}_i, \sigma_i^2 \left(\underline{X}_i' \underline{X}\right)^{-1}\right).
$$

  So, we draw ϕ_i from $\mathcal{N}\left(\underline{A}_i, \sigma_i^2 \left(\underline{X}_i' \underline{X}\right)^{-1}\right)$.

Draw Parameters in Law of Motion for Factors

- Law of motion of factors:

$$f_t = \phi_{0,1} f_{t-1} + ... + \phi_{0,q} f_{t-q} + u_{0,t},\ u_{0,t} \sim \mathcal{N}\left(0, \Sigma_0\right)$$

- The factors, $f_{p+1:T}$, are treated as known, and they correspond to $f^{(s-1)}$, the factors in the $s-1$ iteration of Gibbs sampling.
- By Bayes' rule:

$$p\left(\phi_0 | f_{p+1:T}\right) \propto p\left(f_{p+1:T} | \phi_0\right) p\left(\phi_0\right).$$

- The priors can be implemented by dummy variables.
  - direct application of the methods developed for inference about the parameters of BVARs.
- Draw ϕ_0 from $\mathcal{N}$.

This Completes Step 1 of Gibbs Sampling

- Gibbs sampling computes sequence:

$$\mathcal{P} = \begin{bmatrix} \theta^{(1)} & \cdots & \theta^{(M)} \\ f^{(1)} & \cdots & f^{(M)} \end{bmatrix} = \begin{bmatrix} \mathcal{P}_1 & \cdots & \mathcal{P}_M \end{bmatrix}.$$

- Given $\mathcal{P}_{s-1}$ compute $\mathcal{P}_s$ in two steps.
  - Step 1: draw $\theta^{(s)}$ given $\mathcal{P}_{s-1}$ (direct sampling)
  - Step 2: draw $f^{(s)}$ given θ^{s} (Kalman smoother).

- We now have $\theta^{(s)}$, and must now draw factors.
  - This is done using the Kalman smoother.

Drawing the Factors

- For this, we will put the DFM in the state-space form used to study Kalman filtering and smoothing.
  - In that previous state space form, the measurement error was assumed to be iid.
  - We will make use of the fact that we have all model parameters.
- The DFM:

$$
\begin{aligned}
y_{i,t} &= a_i + \lambda_i' f_t + \xi_{i,t} \\
f_t &= \phi_{0,1} f_{t-1} + \ldots + \phi_{0,q} f_{t-q} + u_{0,t},\ u_{0,t} \sim \mathcal{N}(0, \Sigma_0) \\
\xi_{i,t} &= \phi_{i,1} \xi_{i,t-1} + \ldots + \phi_{i,p} \xi_{i,t-p} + u_{i,t}.
\end{aligned}
$$

- This can be put into our state space form (in which the errors in the observation equation are iid) by quasi-differencing the observer equation.

Observer Equation

- Quasi differencing:

$$\overbrace{\left[1-\phi_i\left(L\right)L\right]y_{i,t}}^{\tilde{y}_{i,t}} = \overbrace{\left[1-\phi_i\left(1\right)\right]a_i}^{\text{constant}} + \lambda_i'\left[1-\phi_i\left(L\right)L\right]f_t + u_{i,t}$$

Then,

$$\begin{aligned}
a &= \begin{bmatrix} \left[1-\phi_i\left(1\right)\right]a_i \\ \vdots \\ \left[1-\phi_i\left(1\right)\right]a_i \end{bmatrix}, \tilde{y}_t = \begin{pmatrix} \tilde{y}_{1,t} \\ \vdots \\ \tilde{y}_{n,t} \end{pmatrix},\ F_t = \begin{pmatrix} f_t \\ \vdots \\ f_{t-p} \end{pmatrix} \\
H &= \begin{bmatrix} \lambda_1' & -\lambda_1'\phi_{1,1} & \cdots & -\lambda_1'\phi_{1,p} \\ \vdots & \vdots & \ddots & \vdots \\ \lambda_n' & -\lambda_n'\phi_{n,1} & \cdots & -\lambda_n'\phi_{n,p} \end{bmatrix},\ u_t = \begin{pmatrix} u_{1,t} \\ \vdots \\ u_{n,t} \end{pmatrix} \\
\tilde{y}_t &= a + HF_t + u_t
\end{aligned}$$

Law of Motion of the State

- Here, the state is denoted by F_t.
- Law of motion:

$$\begin{pmatrix} f_t \\ f_{t-1} \\ f_{t-2} \\ \vdots \\ f_{t-p} \end{pmatrix} = \begin{bmatrix} \phi_{0,1} & \phi_{0,2} & \cdots & \phi_{0,q} & 0_{\kappa\times(p+1-q)} \\ I_\kappa & 0_\kappa & \cdots & 0_\kappa & 0_{\kappa\times(p+1-q)} \\ 0 & I_\kappa & \cdots & 0_\kappa & 0_{\kappa\times(p+1-q)} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & I_\kappa & 0_{\kappa\times(p+1-q)} \end{bmatrix} \begin{pmatrix} f_{t-1} \\ f_{t-2} \\ f_{t-3} \\ \vdots \\ f_{t-1-p} \end{pmatrix} + \begin{pmatrix} u_{0,t} \\ 0_{\kappa\times 1} \\ 0_{\kappa\times 1} \\ \vdots \\ 0_{\kappa\times 1} \end{pmatrix}$$

- LoM:

$$F_t = \Phi F_{t-1} + u_t,\ u_t \sim N\left(0_{\kappa(p+1)\times 1}, V_{(p+1)\kappa\times(p+1)\kappa}\right).$$

State Space Representation of the Factors

- Observer equation:

$$\tilde{y}_t = a + HF_t + u_t.$$

- Law of motion of state:

$$F_t = \Phi F_{t-1} + u_t.$$

- Kalman smoother provides:

$$P\left[F_j | \tilde{y}_1, ..., \tilde{y}_T\right],\ j = 1, ..., T,$$

  together with appropriate second moments.

- Use this information to directly sample $f^{(s)}$ from the Kalman-smoother-provided Normal distribution, completing step 2 of the Gibbs sampler.

Factor Augmented VARs (FAVAR)

- Favar's are DFM's which more closely resemble macro models.
  - There are observables that act like 'factors', hitting all variables directly
  - Examples: the interest rate in the monetary policy rule, government spending, taxes, price of housing, world trade, international price of oil, uncertainty, etc.
- The measurement equation:

$$y_{i,t} = a_i + \gamma_i y_{0,t} + \lambda_i f_t + \xi_{i,t},\ i = 1, ..., n,\ t = 1, ..., T,$$

where $y_{0,t}$ and γ_i are $m \times 1$ and $1 \times m$ vectors, respectively.

- The vectors, $y_{0,t}$ and f_t follow a VAR:

$$\begin{aligned}
\begin{bmatrix} f_t \\ y_{0,t} \end{bmatrix} &= \Phi_{0,1} \begin{bmatrix} f_{t-1} \\ y_{0,t-1} \end{bmatrix} + ... + \Phi_{0,q} \begin{bmatrix} f_{t-q} \\ y_{0,t-q} \end{bmatrix} + u_{0,t}, \\
u_{0,t} &\sim \mathcal{N}\left(0, \Sigma_0\right)
\end{aligned}$$

Literature on FAVARs is Large

- Initial paper: Bernanke and Boivin (2005QJE), "Measuring the Effects of Monetary Policy: A Factor-Augmented Vector Autoregressive (FAVAR) Approach."
- Intention was to correct problems with conventional VAR-based estimates of the effects of monetary policy shocks.
- Include a large number of variables:
  - better capture the actual policy rule of monetary authorities, which look at lots of data in making their decisions.
  - include a lot of variables so that the FAVAR can be used to obtain a comprehensive picture of the effects of a monetary policy shock on the whole economy.
  - Bernanke, et al, include 119 variables in their analysis.

Literature on FAVARs is Large

- Literature is growing: "Large Bayesian Vector Autoregressions," Banbura, Giannone, Reichlin (2010Journal of Applied Economicts), studies importance of including sectoral data to get better estimates of impulse response functions to policy shocks and a better estimate of their impact.
- DFM have been taken in interesting directions, more suitable for multicountry settings, see, e.g., Canova and Ciccarelli (2013,ECB WP1507)
- Time varying FAVARs: Eickmeier, Lemke, Marcellino, "Classical time-varying FAVAR models - estimation, forecasting and structural analysis," (2011Bundesbank Discussion Paper, no. 04/2011). Argue that by allowing parameters to change over time, get better forecasts and characterize how the economy is changing.