Bayesian Estimation of VARs

Frank Schorfheide
Department of Economics, University of Pennsylvania

Preliminaries

- Suppose that the random matrix Φ has density

$$p(\Phi|\Sigma, X'X) \propto |\Sigma \otimes (X'X)^{-1}|^{-T/2} \exp\left\{-\frac{1}{2}tr[\Sigma^{-1}(\Phi - \hat{\Phi})'X'X(\Phi - \hat{\Phi})]\right\} \quad (1)$$

- Let $\beta = vec(\Phi)$ and $\hat{\beta} = vec(\hat{\Phi})$.

- Then

$$\beta|\Sigma, X'X \sim \mathcal{N}\left(\hat{\beta}, \Sigma \otimes (X'X)^{-1}\right). \quad (2)$$

- Note: to generate a draw Z from a multivariate $\mathcal{N}(\mu, \Sigma)$, decompose $\Sigma = CC'$, where C is the lower triangular Cholesky decomposition matrix. Then let $Z = \mu + C\mathcal{N}(0, \mathcal{I})$.

Preliminaries

- The multivariate version of the inverted Gamma distribution is called Wishart Distribution.

- Let Σ be a $n \times n$ positive definite random matrix. Σ has the inverted Wishart $IW(S, \nu)$ distribution if its density is of the form

$$p(\Sigma|S,\nu) \propto |S|^{\nu/2}|\Sigma|^{-(\nu+n+1)/2} \exp\left\{-\frac{1}{2}tr[\Sigma^{-1}S]\right\} \tag{3}$$

- To sample a Σ from an inverted Wishart $IW(S, \nu)$ distribution, draw $n \times 1$ vectors $Z_1, \ldots, Z_\nu$ from a multivariate normal $\mathcal{N}(0, S^{-1})$ and let

$$\Sigma = \left[\sum_{i=1}^{\nu} Z_i Z_i'\right]^{-1}$$

Preliminaries

- Recall:

$$p(Y|\Phi, \Sigma, Y_0) \propto |\Sigma|^{-T/2} \exp\left\{-\frac{1}{2} tr[\Sigma^{-1}S]\right\} \exp\left\{-\frac{1}{2} tr[\Sigma^{-1}(\Phi - \hat{\Phi})' X'X(\Phi - \hat{\Phi})]\right\}$$

- Let's interpret the likelihood as density:

$$\begin{aligned}
& p(\Phi, \Sigma|S, \hat{\Phi}, X'X) \\
& \quad \propto \; |\Sigma|^{-T/2} \exp\left\{-\frac{1}{2} tr[\Sigma^{-1}S]\right\} \exp\left\{-\frac{1}{2} tr[\Sigma^{-1}(\Phi - \hat{\Phi})' X'X(\Phi - \hat{\Phi})]\right\} \\
& \quad \propto \; |\Sigma|^{-T/2} |\Sigma \otimes (X'X)^{-1}|^{1/2} \exp\left\{-\frac{1}{2} tr[\Sigma^{-1}S]\right\} \\
& \qquad\quad (2\pi)^{-nk/2} |\Sigma \otimes (X'X)^{-1}|^{-1/2} \exp\left\{-\frac{1}{2} tr[\Sigma^{-1}(\Phi - \hat{\Phi})' X'X(\Phi - \hat{\Phi})]\right\}
\end{aligned}$$

Preliminaries

- We now integrate out Φ (Note: $|\Sigma \otimes (X'X)^{-1}|^{1/2} = |\Sigma|^{k/2}|X'X|^{-n/2}$):

$$p(\Sigma|S, \hat{\Phi}, X'X) \propto |\Sigma|^{-(T-k)/2}|X'X|^{-n/2} \exp\left\{-\frac{1}{2}tr[\Sigma^{-1}S]\right\}$$

- Hence,

$$\begin{aligned} \Sigma|S, \hat{\Phi}, X'X &\sim \mathcal{IW}(S, T-k-n-1), \\ \Phi|\Sigma, S, \hat{\Phi}, X'X &\sim \mathcal{N}\left(\hat{\Phi}, \Sigma \otimes (X'X)^{-1}\right) \end{aligned}$$

Dummy Observation Priors

- Suppose we have T^* dummy observations (Y^*, X^*).

- The likelihood function for the dummy observations is of the form

$$
\begin{aligned}
& p(Y^*|\Phi, \Sigma) = \qquad (4) \\
& \quad (2\pi)^{-nT^*/2} |\Sigma|^{-T^*/2} \exp \left\{ -\frac{1}{2} tr[\Sigma^{-1}(Y^{*'}Y^* - \Phi' X^{*'} Y^* - Y^{*'} X^* \Phi + \Phi' X^{*'} X^* \Phi)] \right\}.
\end{aligned}
$$

- Combining (4) with the improper prior $p(\Phi, \Sigma) \propto |\Sigma|^{-(n+1)/2}$ yields

$$
\begin{aligned}
& p(\Phi, \Sigma|Y^*) \qquad (5) \\
& \quad = c_*^{-1} |\Sigma|^{-\frac{T^*+n+1}{2}} \exp \left\{ -\frac{1}{2} tr[\Sigma^{-1}(Y^{*'}Y^* - \Phi' X^{*'} Y^* - Y^{*'} X^* \Phi + \Phi' X^{*'} X^* \Phi)] \right\},
\end{aligned}
$$

- which can be interpreted as a prior density for Φ and Σ.

Dummy Observation Priors

- Define

$$\begin{aligned} \Phi^* &= (X^{*'}X^*)^{-1}X^{*'}Y^* \\ S^* &= (Y^* - X^*\Phi^*)'(Y^* - X^*\Phi^*). \end{aligned}$$

- It can be verified that the prior $p(\Phi, \Sigma | Y^*)$ is of the Inverted Wishart-Normal $\mathcal{IW}-\mathcal{N}$ form

$$\Sigma \sim \mathcal{IW}\left(S^*, T^* - k\right) \quad (6)$$

$$\Phi|\Sigma \sim \mathcal{N}\left(\Phi^*, \Sigma \otimes (X^{*'}X^*)^{-1}\right). \quad (7)$$

Dummy Observation Priors

- The appropriate normalization constant for the prior density is given by

$$
\begin{aligned}
c_* \;=\; & (2\pi)^{\frac{nk}{2}} |X^{*'} X^*|^{-\frac{n}{2}} |S^*|^{-\frac{T^*-k}{2}} \\
& 2^{\frac{n(T^*-k)}{2}} \pi^{\frac{n(n-1)}{4}} \prod_{i=1}^{n} \Gamma[(T^* - k + 1 - i)/2],
\end{aligned} \tag{8}
$$

  k is the dimension of x_t and $\Gamma[\cdot]$ denotes the gamma function.

- The implementation of priors through dummy variables is often called mixed estimation and dates back to Theil and Goldberger (1961).

Dummy Observation Priors

- Now let's calculate the <u>posterior</u> ...

- Notice that

$$p(\Phi, \Sigma, Y) \propto p(Y|\Phi, \Sigma)p(Y^*|\Phi, \Sigma) \tag{9}$$

- Define:

$$\tilde{\Phi} = (X^{*'}X^* + X'X)^{-1}(X^{*'}Y^* + X'Y) \tag{10}$$

$$\tilde{S} = \left[Y^{*'}Y^* + Y'Y) - (X^{*'}Y^* + X'Y)'(X^{*'}X^* + X'X)^{-1}(X^{*'}Y^* + X'Y)\right]. \tag{11}$$

Dummy Observation Priors

- Since prior and likelihood function are conjugate, it is straightforward to show, that the posterior distribution of Φ and Σ is also of the Inverted Wishart – Normal form:

$$\Sigma|Y \sim \mathcal{IW}\Big(\tilde{S}, T^* + T - k\Big) \tag{12}$$

$$\Phi|\Sigma, Y \sim \mathcal{N}\Big(\tilde{\Phi}, \Sigma \otimes (X^{*'}X^* + X'X)^{-1}\Big). \tag{13}$$

- Draws $s = 1, \ldots, n_{sim}$ from the posterior can be generated as follows:

  (i) Draw $\Sigma^{(s)}$ from the $\mathcal{IW}$ distribution;

  (ii) draw $\Phi^{(s)}$ from the normal distribution of $\Phi|\Sigma^{(s)}, Y$.

Dummy Observation Priors

- Finally, we can compute the marginal data density ...

- Suppose that we are using a prior constructed from dummy observations. Then the marginal data density is given by

$$p(Y|Y^*) = \frac{\int p(Y, Y^*|\Phi, \Sigma)|\Sigma|^{-(n+1)/2} d\Phi d\Sigma}{\int p(Y^*|\Phi, \Sigma)|\Sigma|^{-(n+1)/2} d\Phi d\Sigma} \tag{14}$$

- The integrals in the numerator and denominator are given by the appropriate modification of c_* defined above:

$$\int p(Y|\Phi, \Sigma)|\Sigma|^{-(n+1)/2} d\Phi d\Sigma = \pi^{-\frac{n(T-k)}{2}} |X'X|^{-\frac{n}{2}} |S|^{-\frac{T-k}{2}} \pi^{\frac{n(n-1)}{4}} \prod_{i=1}^{n} \Gamma[(T-k+1-i)/2], \tag{15}$$

where

$$
\begin{aligned}
\hat{\Phi} &= (X'X)^{-1}X'Y \\
S &= (Y - X\hat{\Phi})'(Y - X\hat{\Phi}).
\end{aligned}
$$

Dummy Observation Priors – Examples

- Minnesota Prior
- Training Sample Prior
- DSGE Model Prior: DSGE-VAR

Minnesota Prior

- Reference: Doan, Litterman, and Sims (1984). The version below is described in the Appendix of Lubik and Schorfheide (Macro Annual, 2005).

- Consider the following Gaussian bivariate VAR(2).

$$\begin{bmatrix} y_{1,t} \\ y_{2,t} \end{bmatrix} = \begin{bmatrix} \alpha_1 \\ \alpha_2 \end{bmatrix} + \begin{bmatrix} \beta_{11} & \beta_{12} \\ \beta_{21} & \beta_{22} \end{bmatrix} \begin{bmatrix} y_{1,t-1} \\ y_{2,t-1} \end{bmatrix} + \begin{bmatrix} \gamma_{11} & \gamma_{12} \\ \gamma_{21} & \gamma_{22} \end{bmatrix} \begin{bmatrix} y_{1,t-2} \\ y_{2,t-2} \end{bmatrix} + \begin{bmatrix} u_{1,t} \\ u_{2,t} \end{bmatrix} \qquad (16)$$

- Define $y_t = [y_{1,t}, y_{2,t}]'$, $x_t = [y'_{t-1}, y'_{t-2}, 1]'$, and $u_t = [u_{1,t}, u_{2,t}]'$ and

$$\Phi = \begin{bmatrix} \beta_{11} & \beta_{21} \\ \beta_{12} & \beta_{22} \\ \gamma_{11} & \gamma_{21} \\ \gamma_{12} & \gamma_{22} \\ \alpha_1 & \alpha_2 \end{bmatrix}. \qquad (17)$$

Minnesota Prior

- The VAR can be rewritten as follows

$$y_t' = x_t'\Phi + u_t', \quad t = 1, \ldots, T, \quad u_t \sim iid\mathcal{N}(0, \Sigma) \tag{18}$$

or in matrix form

$$Y = X\Phi + U. \tag{19}$$

- Based on a short pre-sample Y_0 (typically the observations used to initialized the lags of the VAR) one calculates: $s = std(Y_0)$ and $\bar{y} = mean(Y_0)$.

Minnesota Prior

- In addition there are a number of tuning parameters for the prior
  - τ is the overall tightness of the prior. Large values imply a small prior covariance matrix.
  - d: the variance for the coefficients of lag h is scaled down by the factor l^{-2d}.
  - w: determines the weight for the prior on Σ. Suppose that $Z_i = \mathcal{N}(0, \sigma^2)$. Then an estimator for σ^2 is $\hat{\sigma}^2 = \frac{1}{w}\sum_{i=1}^{w} Z_i^2$. The larger w, the more informative the estimator, and in the context of the VAR, the tighter the prior.
  - λ and μ: additional tuning parameters.

Minnesota Prior

The dummy observations can be classified as follows:

- Dummies for the β coefficients:

$$\begin{aligned} Y^* &= X^*\Phi + U \\ \begin{bmatrix} \tau s_1 & 0 \\ 0 & \tau s_2 \end{bmatrix} &= \begin{bmatrix} \tau s_1 & 0 & 0 & 0 & 0 \\ 0 & \tau s_2 & 0 & 0 & 0 \end{bmatrix} \Phi + \begin{bmatrix} u_{11} & u_{12} \\ u_{21} & u_{22} \end{bmatrix} \end{aligned}$$

The first observation implies, for instance, that

$$\begin{aligned} \tau s_1 &= \tau s_1 \beta_{11} + u_{11} \quad \Longrightarrow \quad \beta_{11} = 1 - \frac{u_{11}}{\tau s_1} \quad \Longrightarrow \quad \beta_{11} \sim \mathcal{N}\left(1, \frac{\Sigma_{11}}{\tau^2 s_1^2}\right) \\ 0 &= \tau s_1 \beta_{21} + u_{12} \quad \Longrightarrow \quad \beta_{21} = -\frac{u_{12}}{\tau s_1} \quad \Longrightarrow \quad \beta_{21} \sim \mathcal{N}\left(0, \frac{\Sigma_{22}}{\tau^2 s_1^2}\right) \end{aligned}$$

Minnesota Prior

The dummy observations can be classified as follows (continued...):

- Dummies for the γ coefficients:

$$\begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix} = \begin{bmatrix} 0 & 0 & \tau s_1 2^d & 0 & 0 \\ 0 & 0 & 0 & \tau s_2 2^d & 0 \end{bmatrix} \Phi + U$$

- The prior for the covariance matrix is implemented by

$$\begin{bmatrix} s_1 & 0 \\ 0 & s_2 \end{bmatrix} = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix} \Phi + U$$

Minnesota Prior

The dummy observations can be classified as follows (continued...):

- Co-persistence prior dummy observations, reflecting the belief that when data on all y's are stable at their initial levels, thy will tend to persist at that level:

$$\begin{bmatrix} \lambda\bar{y}_1 & \lambda\bar{y}_2 \end{bmatrix} = \begin{bmatrix} \lambda\bar{y}_1 & \lambda\bar{y}_2 & \lambda\bar{y}_1 & \lambda\bar{y}_2 & \lambda \end{bmatrix} \Phi + U$$

- Own-persistence prior dummy observations, reflecting the belief that when y_i has been stable at its initial level, it will tend to persist at that level, regardless of the value of other variables:

$$\begin{bmatrix} \mu\bar{y}_1 & 0 \\ 0 & \mu\bar{y}_2 \end{bmatrix} = \begin{bmatrix} \mu\bar{y}_1 & 0 & \mu\bar{y}_1 & 0 & 0 \\ 0 & \mu\bar{y}_2 & 0 & \mu\bar{y}_2 & 0 \end{bmatrix} \Phi + U$$

Dummy Observation Priors – Examples

- Training Sample Prior: replace dummy observations by actual observations from a pre- or training sample.

- DSGE Model Prior: use artificial observations generated by a DSGE model. Details will follow later.

Implementation of Structural VAR Analysis

- Consider a simple VAR of the form $y_t = \Phi_1 y_{t-1} + u_t$, $u_t = A\Omega(\varphi)\epsilon_t$, $\Phi = \Phi_1'$. For $s = 1, \ldots, n_{sim}$:

  1. Generate a draw from the posterior distribution of (Φ, Σ), e.g., using sampling techniques for the $\mathcal{IW} - \mathcal{N}$ distribution. Let $A = chol(\Sigma)$.
  2. Compute moving average representation $y_t = \sum_{j=0} C_j(\Phi) u_t$.
  3. Short-run and long-run identification schemes: determine φ as function of A and the $C_j(\Phi)$'s.

     Sign Restrictions: conditional on Φ and A assign a prior distribution to the set of φ's for which the sign restrictions are satisfied. Generate a draw φ from this prior. Note: the sample has no information about φ given Φ, A. Hence prior equals posterior. [NOTE: SAMPLING IS MORE DELICATE DUE TO VARYING

NORMALIZATION OF $p(\varphi|\Phi, \Sigma)$

4. Once φ is determined, compute impulse responses and variance decompositions.

- This algorithm leaves you with n_{sim} draws from the posterior of the impulse responses and variance decompositions. You can now compute summary statistics for this posterior, such as means, medians, standard deviations, and (pointwise) confidence sets.