

Bayesian Methods for Macroeconometrics

Frank Schorfheide

Department of Economics, University of Pennsylvania

Preliminaries

- Suppose that the random matrix Φ has density

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

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

- 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

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

- To sample a Σ from an inverted Wishart $IW(S, \nu)$ distribution, draw $n \times 1$ vectors Z_1, \dots, Z_ν 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:

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

- Let's interpret the likelihood as density:

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

Preliminaries

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

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

- 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

$$p(Y^*|\Phi, \Sigma) = \tag{4}$$

$$(2\pi)^{-nT^*/2} |\Sigma|^{-T^*/2} \exp \left\{ -\frac{1}{2} \text{tr} [\Sigma^{-1} (Y^{*'} Y^* - \Phi' X^{*'} Y^* - Y^{*'} X^* \Phi + \Phi' X^{*'} X^* \Phi)] \right\}.$$

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

$$p(\Phi, \Sigma|Y^*) \tag{5}$$

$$= c_*^{-1} |\Sigma|^{-\frac{T^*+n+1}{2}} \left\{ -\frac{1}{2} \text{tr} [\Sigma^{-1} (Y^{*'} Y^* - \Phi' X^{*'} Y^* - Y^{*'} X^* \Phi + \Phi' X^{*'} X^* \Phi)] \right\},$$

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

Dummy Observation Priors

- Define

$$\hat{\Phi}^* = (X^{*'} X^*)^{-1} X^{*'} Y^*$$

$$S^* = (Y^* - X^* \hat{\Phi}^*)' (Y^* - X^* \hat{\Phi}^*).$$

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

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

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

Dummy Observation Priors

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

$$c_* = (2\pi)^{\frac{nk}{2}} |X^{*'} X^*|^{-\frac{n}{2}} |S^*|^{-\frac{T^*-k}{2}} \quad (8)$$

$$2^{\frac{n(T^*-k)}{2}} \pi^{\frac{n(n-1)}{4}} \prod_{i=1}^n \Gamma[(T^* - k + 1 - i)/2],$$

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 posterior ...
- Notice that

$$p(\Phi, \Sigma_u | Y) \propto p(Y | \Phi, \Sigma_u) p(Y^* | \Phi, \Sigma_u) \quad (9)$$

- Define:

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

$$\tilde{\Sigma}_u = \frac{1}{T^* + T} \left[Y^{*'} Y^* + Y' Y \right. \\ \left. - (X^{*'} Y^* + X' Y)' (X^{*'} X^* + X' X)^{-1} (X^{*'} Y^* + X' Y) \right]. \quad (11)$$

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:

$$\Sigma|Y \sim \mathcal{IW}\left((T^* + T)\tilde{\Sigma}_u, T^* + T - k\right) \quad (12)$$

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

- Draws $s = 1, \dots, n_{sim}$ 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

$$p(Y|Y^*) = \frac{\int p(Y, Y^*|\Phi, \Sigma)d\Phi d\Sigma}{\int p(Y^*|\Phi, \Sigma)d\Phi d\Sigma} \quad (14)$$

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

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

where

$$\hat{\Phi} = (X'X)^{-1}X'Y$$

$$S = (Y - X\hat{\Phi})'(Y - X\hat{\Phi}).$$

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).

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

- 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

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

Minnesota Prior

- The VAR can be rewritten as follows

$$y'_t = x'_t \Phi + u'_t, \quad t = 1, \dots, T, \quad u_t \sim iid \mathcal{N}(0, \Sigma_u) \quad (18)$$

or in matrix form

$$Y = X\Phi + U. \quad (19)$$

- Based on a short pre-sample Y_0 (typically the observations used to initialize 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_u . 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:

$$\begin{bmatrix} \tau s_1 & 0 \\ 0 & \tau s_2 \end{bmatrix} = \begin{bmatrix} \tau s_1 & 0 & 0 & 0 & 0 \\ 0 & \tau s_2 & 0 & 0 & 0 \end{bmatrix} \Phi + u'$$

The first observation implies, for instance, that

$$\begin{aligned} \tau s_1 &= \tau s_1 \beta_{11} + u_1 &\implies & \beta_{11} = 1 - \frac{u_1}{\tau s_1} &\implies & \beta_{11} \sim \mathcal{N}\left(1, \frac{\Sigma_{u,11}}{\tau^2 s_1^2}\right) \\ 0 &= \tau s_1 \beta_{21} + u_2 &\implies & \beta_{21} = -\frac{u_2}{\tau s_1} &\implies & \beta_{21} \sim \mathcal{N}\left(0, \frac{\Sigma_{u,22}}{\tau^2 s_1^2}\right) \end{aligned}$$

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, they 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.