Econ. 556b Spring 1999 Chris Sims

Sampling from a VAR Likelihood*

Consider a reduced form VAR written as

$$\underset{m\times 1}{y(t)} = \sum_{s=1}^{k} B(s)y(t-s) + c + \varepsilon(t)\ ,\ \ t = 1, \dots , T\ . \tag{1}$$

We assume that ε is distributed as $N(0, \Sigma)$, i.i.d. across time t and with $\varepsilon(t)$ independent of $y(t-s)$ for all $s > 0$. For understanding the algebra of the likelihood, it is helpful to introduce

$$\underset{T\times m}{Y} = \begin{bmatrix} y(1) & \cdots & y(T) \end{bmatrix}' \tag{2}$$

$$\underset{T\times(mk+1)}{X} =$$

$$\begin{bmatrix} y_1(0) & \dots & y_m(0) & y_1(-1) & \dots & y_m(-1) & \dots & y_m(-k+1) & 1 \\ y_1(1) & \dots & y_m(1) & y_1(0) & \dots & y_m(0) & \dots & y_m(-k+2) & 1 \\ \vdots & & & & & & & & \vdots \\ y_1(T-1) & \dots & y_m(T-1) & y_1(T-2) & \dots & y_m(T-2) & \dots & y_m(T-k) & 1 \end{bmatrix} \tag{3}$$

$$\underset{T\times m}{\boldsymbol{\varepsilon}} = \begin{bmatrix} \varepsilon(1) & \cdots & \varepsilon(T) \end{bmatrix} \tag{4}$$

$$\mathbf{B} = \begin{bmatrix} B(1) & B(2) & \cdots & B(k) & c \end{bmatrix}'\ . \tag{5}$$

Using this notation, the model can be written as

$$Y = X\mathbf{B} + \boldsymbol{\varepsilon}\ . \tag{6}$$

This model has the likelihood function of a standard "seemingly unrelated" normal linear regression model. Its likelihood function is, therefore

$$|\Sigma|^{-T/2} \exp\left(-\tfrac{1}{2}\operatorname{trace}(\Sigma^{-1}u'u)\right)\ , \tag{7}$$

where

$$u = Y - X\mathbf{B}. \tag{8}$$

The algebra of seemingly unrelated regressions lets us conclude that, for a given Σ, the likelihood as a function of $\operatorname{vec}(\mathbf{B})$ is proportional to a $N(\operatorname{vec}(\hat{\mathbf{B}}_{OLS}), \Sigma \otimes (X'X)^{-1})$ pdf. (The operator $\operatorname{vec}(\cdot)$ converts a matrix to a column vector by stacking the matrix's columns on top of each other.) We can therefore easily integrate over $\mathbf{B}$ to arrive at a marginal likelihood for Σ, which is

$$|\Sigma|^{\frac{-T+mk+1}{2}} \left|X'X\right|^{\frac{-m}{2}} \exp\left(-\tfrac{1}{2}\operatorname{trace}(\Sigma^{-1}\hat{S})\right)\ , \tag{9}$$

where $\hat{S} = \hat{u}'\hat{u}$ is the cross product matrix of least squares residuals. As a function of Σ, (9) is proportional to an inverse-Wishart pdf with parameter S^{-1} and degrees of freedom $T - (m+1)k - 2$. It is common practice to use a Jeffreys prior on Σ, which is an improper pdf proportional to $|\Sigma|^{-(m+1)/2}$. With this prior, the degrees of freedom in the marginal posterior for Σ become $T - (mk+1)$, i.e. T less the number of regressors in each equation. Note that S here is the raw cross-product matrix of sample residuals, *not* the residual covariance matrix.

If W has a Wishart distribution with parameter V and ν degrees of freedom, W^{-1} has the inverse-Wishart distribution with the same degrees of freedom and parameter V. The sample cross-product of T independent draws from a $N(0, V)$ distribution is Wishart with T degrees of freedom and parameter V. So to draw from the marginal posterior for Σ derived above, we would generate $T - (mk+1)$ draws from a $N(0, \hat{S}^{-1})$ distribution, stack them into a $(T - (mk+1)) \times m$ matrix Z, and take the inverse of $Z'Z$.

To draw from the full joint posterior pdf on $\mathbf{B}$ and Σ, we first draw from the marginal distribution for Σ as just explained, then use the newly drawn Σ to draw $\mathbf{B}$ from its conditional $N(\text{vec}(\hat{\mathbf{B}}_{OLS}), \Sigma \otimes (X'X)^{-1})$ distribution.

Though this discussion has focused entirely on the case of a flat prior on $\mathbf{B}$ combined with a Jeffreys prior on Σ, the algebra of course works in exactly the same way if there is a prior and it is implemented with dummy observations (or, equivalently, the prior is conjugate to the likelihood).