The Likelihood Function of a VAR

Frank Schorfheide
Department of Economics, University of Pennsylvania

Likelihood Function

- We will now derive the likelihood function for a Gaussian VAR(p), conditional on initial observations $y_0, \ldots, y_{-p+1}$:

$$y_t = \Phi_0 + \Phi_1 y_{t-1} + \ldots + \Phi_p y_{t-p} + u_t \tag{1}$$

- The likelihood function can be used for both Bayesian and frequentist inference.

- The density of y_t conditional on $y_{t-1}, y_{t-2}, \ldots$ and the coefficient matrices $\Phi_0, \Phi_1, \ldots, \Sigma$ is of the form

$$\begin{aligned} p(y_t|Y^{t-1}, \Phi_0, \ldots, \Sigma) \;\propto\; & |\Sigma|^{-1/2} \exp\left\{ -\frac{1}{2}(y_t - \Phi_0 - \Phi_1 y_{t-1} - \ldots - \Phi_p y_{t-p})' \right. \\ & \left. \times \Sigma^{-1}(y_t - \Phi_0 - \Phi_1 y_{t-1} - \ldots - \Phi_p y_{t-p}) \right\} \end{aligned} \tag{2}$$

Likelihood Function

- Define the $(np+1) \times 1$ vector x_t as

$$x_t = [1, y'_{t-1}, \ldots, y'_{t-p}]'$$

- Moreover, define the matrixes

$$Y = \begin{bmatrix} y'_1 \\ \vdots \\ y'_T \end{bmatrix}, \quad X = \begin{bmatrix} x'_1 \\ \vdots \\ x'_T \end{bmatrix}, \quad \Phi = [\Phi_0, \Phi_1, \ldots, \Phi_p]'$$

- The conditional density of y_t can be written in more compact notation as

$$p(y_t|Y^{t-1}, \Phi, \Sigma) \propto |\Sigma|^{-1/2} \exp\left\{ -\frac{1}{2}(y'_t - x'_t\Phi)\Sigma^{-1}(y'_t - x'_t\Phi)' \right\} \tag{3}$$

To manipulate the density we will use some matrix algebra facts.

Likelihood Function

- **Facts**:

  (i) Let a be a $n \times 1$ vector, B be a symmetric positive definite $n \times n$ matrix, and tr the trace operator that sums the diagonal elements of a matrix. Then

$$a'Ba = tr[Baa']$$

  (ii) Let A and B be two $n \times n$ matrices, then

$$tr[A + B] = tr[A] + tr[B]$$

Likelihood Function

- In a first step, we will replace the inner product in the expression for the conditional density by the trace of the outer product

$$p(y_t|Y^{t-1},\Phi,\Sigma) \propto |\Sigma|^{-1/2} \exp\left\{-\frac{1}{2} tr[\Sigma^{-1}(y_t'-x_t'\Phi)'(y_t'-x_t'\Phi)]\right\} \tag{4}$$

- In the second step, we will take the product of the conditional densities of $y_1,\ldots,y_T$ to obtain the joint density. Let Y_0 be a vector with initial observations

$$\begin{aligned} p(Y|\Phi,\Sigma,Y_0) &= \prod_{t=1}^{T} p(y_t|Y^{t-1},Y_0,\Phi,\Sigma) \\ &\propto |\Sigma|^{-T/2} \exp\left\{-\frac{1}{2}\sum_{t=1}^{T} tr[\Sigma^{-1}(y_t'-x_t'\Phi)'(y_t'-x_t'\Phi)]\right\} \\ &\propto |\Sigma|^{-T/2} \exp\left\{-\frac{1}{2} tr\left[\Sigma^{-1}\sum_{t=1}^{T}(y_t'-x_t'\Phi)'(y_t'-x_t'\Phi)\right]\right\} \\ &\propto |\Sigma|^{-T/2} \exp\left\{-\frac{1}{2} tr[\Sigma^{-1}(Y-X\Phi)'(Y-X\Phi)]\right\} \end{aligned} \tag{5}$$

Likelihood Function

- Define the "OLS" estimator

$$\hat{\Phi} = (X'X)^{-1}X'Y \tag{6}$$

and the sum of squared OLS residual matrix

$$S = (Y - X\hat{\Phi})'(Y - X\hat{\Phi}) \tag{7}$$

- It can be verified that

$$(Y - X\Phi)'(Y - X\Phi) = S + (\Phi - \hat{\Phi})'X'X(\Phi - \hat{\Phi}) \tag{8}$$

- This leads to the following representation of the likelihood function

$$p(Y|\Phi,\Sigma,Y_0) \propto |\Sigma|^{-T/2} \exp\left\{-\frac{1}{2}tr[\Sigma^{-1}S]\right\} \exp\left\{-\frac{1}{2}tr[\Sigma^{-1}(\Phi - \hat{\Phi})'X'X(\Phi - \hat{\Phi})]\right\} \tag{9}$$

Likelihood Function

- Let $\beta = vec(\Phi)$ and $\hat{\beta} = vec(\hat{\Phi})$. It can be verified that

$$tr[\Sigma^{-1}(\Phi - \hat{\Phi})' X' X (\Phi - \hat{\Phi})] = (\beta - \hat{\beta})'[\Sigma \otimes (X'X)^{-1}]^{-1}(\beta - \hat{\beta}) \tag{10}$$

- and the likelihood function has the alternative representation

$$p(Y|\Phi, \Sigma, Y_0) \propto |\Sigma|^{-T/2} \exp\left\{-\frac{1}{2} tr[\Sigma^{-1} S]\right\} \exp\left\{-\frac{1}{2}(\beta - \hat{\beta})'[\Sigma \otimes (X'X)^{-1}]^{-1}(\beta - \hat{\beta})\right\} \tag{11}$$