Notes on Model Evaluation

Latest Revision: 04/02/04

Frank Schorfheide[1]

1 General Idea

In order to assess the overall fit of a DSGE model ($\mathcal{M}_1$) it is useful to compute posterior odds (or Bayes factors) of the DSGE model versus a more general reference model, such as a VAR ($\mathcal{M}_0$). The Bayes factor is given by

$$B(\mathcal{M}_1, \mathcal{M}_0) = \frac{p(Y|\mathcal{M}_1)}{p(Y|\mathcal{M}_0)} \tag{1}$$

The marginal data densities $p(Y|\mathcal{M}_i)$ are defined as

$$p(Y|\mathcal{M}_i) = \int p(Y|\theta_{(i)}, \mathcal{M}_i) p(\theta_{(i)}|\mathcal{M}_i) d\theta_{(i)}, \tag{2}$$

where $p(Y|\theta_{(i)}, \mathcal{M}_i)$ is the likelihood function for model $\mathcal{M}_i$ and $p(\theta_{(i)}|\mathcal{M}_i)$ is the prior density for the parameters of model $\mathcal{M}_i$. For the Bayes factor to be well defined, the priors have to be proper probability density functions that integrate to one.

2 Reference Model

Vector autoregressions (VAR) can serve as reference models for the evaluation of DSGE models. For instance, consider the following Gaussian bivariate VAR(2).

$$\begin{bmatrix} y_{1,t} \\ y_{2,t} \end{bmatrix} = \begin{bmatrix} \alpha_1 \\ \alpha_2 \end{bmatrix} + \begin{bmatrix} \beta_{11} & \beta_{12} \\ \beta_{21} & \beta_{22} \end{bmatrix} \begin{bmatrix} y_{1,t-1} \\ y_{2,t-1} \end{bmatrix} + \begin{bmatrix} \gamma_{11} & \gamma_{12} \\ \gamma_{21} & \gamma_{22} \end{bmatrix} \begin{bmatrix} y_{1,t-2} \\ y_{2,t-2} \end{bmatrix} + \begin{bmatrix} u_{1,t} \\ u_{2,t} \end{bmatrix} \tag{3}$$

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

$$\Phi = \begin{bmatrix} \beta_{11} & \beta_{21} \\ \beta_{12} & \beta_{22} \\ \gamma_{11} & \gamma_{21} \\ \gamma_{12} & \gamma_{22} \\ \alpha_1 & \alpha_2 \end{bmatrix}. \tag{4}$$

The VAR can be rewritten as follows

$$y'_t = x'_t\Phi + u'_t, \quad t = 1, \ldots, T, \quad u_t \sim iid\mathcal{N}(0, \Sigma_u) \tag{5}$$

[1]Department of Economics, 3718 Locust Walk, Room 525 McNeil Building, Telephone (215) 898 8486. Email: *schorf@ssc.upenn.edu*, URL: *http://www.econ.upenn.edu/˜schorf*.

or in matrix form

$$Y = X\Phi + U. \tag{6}$$

We denote the dimension of y_t by n.

2.1 Priors from Dummy Observations

Suppose we have T^* dummy observations (Y^*, X^*). The likelihood function for the dummy observations is of the form

$$\begin{aligned} &p(Y^*|\Phi,\Sigma_u) = \\ &\quad (2\pi)^{-nT^*/2}|\Sigma_u|^{-T^*/2}\exp\left\{-\frac{1}{2}tr[\Sigma_u^{-1}(Y^{*'}Y^* - \Phi' X^{*'}Y^* - Y^{*'}X^*\Phi + \Phi' X^{*'}X^*\Phi)]\right\}. \end{aligned} \tag{7}$$

If we combine the dummy observations with actual observations we obtain a likelihood function that can be factorized as follows

$$p(Y^*, Y|\Phi,\Sigma_u) = p(Y^*(\theta)|\Phi,\Sigma_u)p(Y|\Phi,\Sigma_u) \tag{8}$$

This factorization suggests that the term $p(Y^*|\Phi,\Sigma_u)$ can be interpreted as a prior density for Φ and Σ_u.

Combining (7) with the improper prior $p(\Phi,\Sigma_u) \propto |\Sigma_u|^{-(n+1)/2}$ yields

$$p(\Phi,\Sigma_u) = c_*^{-1}|\Sigma_u|^{-\frac{T^*+n+1}{2}}\left\{-\frac{1}{2}tr[\Sigma_u^{-1}(Y^{*'}Y^* - \Phi' X^{*'}Y^* - Y^{*'}X^*\Phi + \Phi' X^{*'}X^*\Phi)]\right\}, \tag{9}$$

where

$$\begin{aligned} c_* &= (2\pi)^{\frac{nk}{2}}|X^{*'}X^*|^{-\frac{n}{2}}|S^*|^{-\frac{T^*-k}{2}} \\ &\quad 2^{\frac{n(T^*-k)}{2}}\pi^{\frac{n(n-1)}{4}}\prod_{i=1}^{n}\Gamma[(T^*-k+1-i)/2], \end{aligned} \tag{10}$$

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

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

Details of this calculation can be found in Zellner (1971). The implementation of priors through dummy variables is often called mixed estimation and dates back to Theil and Goldberger (1961).

2.2 Minnesota Prior and Dummy Observations

The *Matlab* program *varprior.m*, written by Chris Sims implements a version of the Minnesota Prior (Doan, Litterman, and Sims, 1984). A brief description follows.

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

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

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

- 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

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

Note: I think the code is only valid for $w = 1$. In general one needs w of these observations.

- 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:

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

2.3 Training Sample Priors

In the same way we constructed a prior from dummy observations, we can also construct a prior from a training sample. Suppose we split the actual sample $Y = [Y^-, Y^+]$, where Y^- is interpreted as training sample, then

$$p(\Phi, \Sigma_u) = c_-^{-1} |\Sigma_u|^{-\frac{T^- + n + 1}{2}} \left\{ -\frac{1}{2} tr[\Sigma_u^{-1}(Y^{-'} Y^- - \Phi' X^{-'} Y^- - Y^{-'} X^- \Phi + \Phi' X^{-'} X^- \Phi)] \right\}, \tag{11}$$

Of course one can also combine the dummy observations and training sample to construct a prior distribution.

2.4 Marginal Data Density

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

$$p(Y^+ | Y^-, Y^*, \mathcal{M}_0) = \frac{\int p(Y^+, Y^-, Y^* | \Phi, \Sigma_u) d\Phi d\Sigma_u}{\int p(Y^-, Y^* | \Phi, \Sigma_u) d\Phi d\Sigma_u} \tag{12}$$

where the integrals in the numerator and denominator are given by the appropriate modification of c_* defined above. This calculation is implemented by the procedure *mgnldnsty.m*. More specifically:

$$\int p(Y|\Phi, \Sigma_u) d\Phi d\Sigma_u = \pi^{-\frac{T-k}{2}} |X'X|^{-\frac{n}{2}} |S|^{-\frac{T-k}{2}} \pi^{\frac{n(n-1)}{4}} \prod_{i=1}^{n} \Gamma[(T - k + 1 - i)/2], \tag{13}$$

where

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

2.5 Comparison with DSGE Model

If a DSGE model is compared to a VAR with training sample prior, then the marginal data density for the DSGE model should be adjusted to reflect the use of the training sample. Note that,

$$p(Y^+ | Y^-, \mathcal{M}_1) = \frac{p(Y^+, Y^- | \mathcal{M}_1)}{p(Y^{-1} | \mathcal{M}_1)}, \tag{14}$$

which amounts to running the marginal data density procedure for DSGE models in DYNARE twice: once for the training sample Y^- only (requires separate calculation of posterior mode, hessian, and a separate run of the Metropolis algorithm) and, second, for the combined sample $Y = [Y^-, Y^+]$.

3 Recommendations for Dynare

- Incorporate the VAR procedures provided by Chris Sims. Check whether the *vprior.w* option is correctly implemented. Let users choose the extent of training sample and the tuning parameters for the Minnesota prior.
- If user chooses to use a training sample prior, then automatically adjust the calculation of the marginal data density for the DSGE model to account for training sample.

References

Doan, Thomas, Robert Litterman, and Christopher Sims (1984): "Forecasting and Conditional Projections Using Realistic Prior Distributions," *Econometric Reviews*, **3**, 1-100.

Sims, Christopher (2003): "Matlab Procedures to Compute Marginal Data Densities for VARs with Minnesota and Training Sample Priors," Department of Economics, Princeton University.

Theil, Henry and Arthur S. Goldberger (1961): "On Pure and Mixed Estimation in Economics," *International Economic Review*, **2**, 65-78.

Zellner, Arnold (1971): *"Introduction to Bayesian Inference in Econometrics,"* John Wiley & Sons, New York.