Bvar & DSGE models

Stéphane Adjemian

Université du Maine, GAINS & CEPREMAP

stephane.adjemian@ens.fr

July 3, 2007

- The fit of a DSGE model is often evaluated by comparing its marginal density with the marginal density of a BVAR model, considered as more general.
- This comparison suffers from several limits:
  1. The VAR model is not really more general

     $$\text{DSGE} \in \text{VARMA} \notin \text{VAR}$$

  2. Some guidance is missing to choose the priors of the BVAR model.
  3. Comparison of marginal densities is uninformative about the directions where the DSGE model is successful (in terms of fit) or unsuccessful.
- Del Negro & Schorfheide (IER, 2004) answer to limits 2

and 3... They build the priors of a BVAR model from a DSGE model and evaluate the optimal weight of the DSGE prior.

$$\underset{1\times m}{y_t} = \sum_{k=1}^{p} y_{t-k} \underset{m\times m}{\mathbf{A}_k} + \underset{1\times q}{x_t} \underset{q\times m}{\mathbf{C}} + \varepsilon_t$$

with $\varepsilon_t \sim \mathcal{N}(0, \Sigma_\varepsilon)$. $(A_k)_{i,j}$ is the coefficient associated to the the variable i at lag k in equation j. Equivalently we have:

$$\underset{T\times m}{Y} = \underset{T\times (mp+q)}{Z} \mathcal{A} + E$$

with $\mathcal{A} = (\mathbf{A}_1', ..., \mathbf{A}_p', \mathbf{C}')'$, or:

$$\mathbf{y} = (I_m \otimes Z)\mathbf{a} + \mathbf{e}$$

where $\mathbf{y} = \mathrm{vec}Y$, $\mathbf{a} = \mathrm{vec}\mathcal{A}$, ...

- Our VAR is a gaussian linear model...

$$L(\mathcal{A},\Sigma;\mathcal{Y}) = (2\pi)^{-\frac{mT}{2}} \left|\Sigma \otimes I_T\right|^{-\frac{1}{2}} e^{-\frac{1}{2}(\mathbf{y}-(I_m\otimes Z)\mathbf{a})'\Sigma^{-1}\otimes I_T(\mathbf{y}-(I_m\otimes Z)\mathbf{a})}$$

- Or more compactly:

$$L(\mathcal{A},\Sigma;\mathcal{Y}) = (2\pi)^{-\frac{mT}{2}} \left|\Sigma\right|^{-\frac{T}{2}} e^{-\frac{1}{2}\mathrm{tr}\{(Y-Z\mathcal{A})\Sigma^{-1}(Y-Z\mathcal{A})'\}}$$

- The ML estimator (or OLS) is given by:

$$\widehat{\mathcal{A}} = (Z'Z)^{-1}Z'Y$$

$$\widehat{\Sigma} = T^{-1}(Y - Z\widehat{\mathcal{A}})'(Y - Z\widehat{\mathcal{A}})$$

- One can show that the likelihood may be written as:

$$L(\mathcal{A},\Sigma;\mathcal{Y}) = (2\pi)^{-\frac{mT}{2}} \times |\Sigma|^{-\frac{k}{2}}\, e^{-\frac{1}{2}\mathrm{tr}\left\{\Sigma^{-1}(\mathcal{A}-\widehat{\mathcal{A}})'Z'Z(\mathcal{A}-\widehat{\mathcal{A}})\right\}}$$
$$\times |\Sigma|^{-\frac{T-k}{2}}\, e^{-\frac{1}{2}\mathrm{tr}\left\{\Sigma^{-1}(Y-Z\widehat{\mathcal{A}})'(Y-Z\widehat{\mathcal{A}})\right\}}$$

- or equivalently:

$$L(\mathcal{A},\Sigma;\mathcal{Y}) = (2\pi)^{-\frac{mT}{2}} \times (2\pi)^{\frac{km}{2}} \left|Z'Z\right|^{-\frac{m}{2}} f_{MN_{k,m}}(\mathcal{A};\widehat{\mathcal{A}},(Z'Z)^{-1},\Sigma)$$
$$\times \frac{2^{\frac{\nu m}{2}}\pi^{\frac{m(m-1)}{4}} \prod_{i=1}^{m}\Gamma\left(\frac{\nu+m-i}{2}\right)}{|(Y-Z\widehat{\mathcal{A}})'(Y-Z\widehat{\mathcal{A}})|^{\frac{\nu}{2}}} \times f_{i\mathcal{W}_m}(\Sigma;(Y-Z\widehat{\mathcal{A}})'(Y-Z\widehat{\mathcal{A}}),\nu)$$

where $\nu = T-k-m-1$ is the degree of freedom.

- A multivariate random variable $\mathbf{X}$ is said to be distributed as a*matricvariate normal*, $\mathbf{X} \sim MN_{p,q}(\mathbf{M}, \mathbf{P}, \mathbf{Q})$ where $\mathbf{M}$, $\mathbf{Q}$ and $\mathbf{P}$ are $p \times q$, $q \times q$ and $p \times p$ matrices, with $\mathbf{P}$ and $\mathbf{Q}$ symmetric and positive definite, if

$$\text{vec}\mathbf{X} \sim \mathcal{N}_{pq}\left(\text{vec}\mathbf{M}, \mathbf{Q} \otimes \mathbf{P}\right)$$

- The density function is given by:

$$f_{MN_{p,q}}(X; \mathbf{M}, \mathbf{P}, \mathbf{Q}) = (2\pi)^{-\frac{pq}{2}} |\mathbf{Q}|^{-\frac{p}{2}} |\mathbf{P}|^{-\frac{q}{2}} e^{-\frac{1}{2}\mathrm{tr}\left\{\mathbf{Q}^{-1}(X-\mathbf{M})'\mathbf{P}^{-1}(X-\mathbf{M})\right\}}$$

- A multivariate random variable $\mathbf{X}$ is said to be distributed as an *inverted wishart*, $\mathbf{X} \sim iW_q(\mathbf{Q}, \nu)$ where $\mathbf{Q}$ is a $q \times q$ symmetric and positive definite matrix, if $\mathbf{X}^{-1} \sim \mathcal{W}_q(\mathbf{Q}^{-1}, \nu)$, a Wishart random variable ($\rightarrow$ multivariate chi squared distribution).

- The density function is defined as follows:

$$f_{i\mathcal{W}_q}(X; \mathbf{Q}, \nu) = \frac{|\mathbf{Q}|^{\frac{\nu}{2}} |X|^{-\frac{\nu+q+1}{2}}}{2^{\frac{\nu q}{2}} \pi^{\frac{q(q-1)}{4}} \prod_{i=1}^{q} \Gamma\left(\frac{\nu+q-i}{2}\right)} e^{-\frac{1}{2}\mathrm{tr}\left\{X^{-1}\mathbf{Q}\right\}}$$

- $\rightarrow$ The likelihood is proportional to the product of the density of an *inverted Wishart* and the density of a *matricvariate normal*.
- We have:

$$\begin{aligned} L(\mathcal{A},\Sigma;\mathcal{Y}) \propto\ & f_{MN_{k,m}}(\mathcal{A};\widehat{\mathcal{A}},(Z'Z)^{-1},\Sigma) \\ & \times f_{i\mathcal{W}_m}(\Sigma;(Y-Z\widehat{\mathcal{A}})'(Y-Z\widehat{\mathcal{A}}),\nu) \end{aligned}$$

- ... This property gives us some hints to carefully choose the shape of our priors $\rightarrow$ *conjugate priors*.

The Jeffrey's flat prior for our Bvar model is:

$$p_0\left(\mathcal{A},\Sigma\right) = |\Sigma|^{-\frac{m+1}{2}}$$

The posterior density is:

$$\begin{aligned} p\left(\mathcal{A},\Sigma|\mathcal{Y}_T^\star\right) \propto & (2\pi)^{-\frac{mT}{2}} \times (2\pi)^{\frac{km}{2}} \left|Z'Z\right|^{-\frac{m}{2}} \\ & \times f_{MN_{k,m}}(\mathcal{A};\widehat{\mathcal{A}},(Z'Z)^{-1},\Sigma) \\ & \times 2^{\frac{\nu m}{2}} \pi^{\frac{m(m-1)}{4}} |\widehat{S}|^{-\frac{\nu}{2}} \prod_{i=1}^{m} \Gamma\left(\frac{\nu+1-i}{2}\right) \\ & \times f_{i\mathcal{W}_m}(\Sigma;\widehat{S},\nu) \times |\Sigma|^{-\frac{m+1}{2}} \end{aligned}$$

We have:

$$
\begin{aligned}
p(\mathcal{A}, \Sigma; \mathcal{Y}_T^\star) \propto\ & f_{MN_{k,m}}(\mathcal{A}; \widehat{\mathcal{A}}, (Z'Z)^{-1}, \Sigma) \\
& \times f_{i\mathcal{W}_m}(\Sigma; \widehat{S}, \tilde{\nu})
\end{aligned} \tag{1}
$$

with $\tilde{\nu} = T - k$. So that the posterior density may be written as:

$$
\begin{aligned}
\mathcal{A}|\Sigma, \mathcal{Y}_T^\star &\sim MN_{k,m}\left(\widehat{\mathcal{A}}, \Sigma, (Z'Z)^{-1}\right) \\
\Sigma|\mathcal{Y}_T^\star &\sim i\mathcal{W}_m\left(\widehat{S}, \tilde{\nu}\right)
\end{aligned} \tag{2}
$$

$\looparrowright$ The posterior mean of $\mathcal{A}$ is the ML estimator of $\mathcal{A}$.

- Suppose our prior for $\mathcal{A}$ is:

$$p_0(\text{vec } \mathcal{A}) \sim \mathcal{N}(a_0, \Omega_0)$$

where Ω_0 is an $mp \times mp$ symmetric positive definite matrix.

- Suppose also that our prior for Σ is degenerate, $\Sigma = \widehat{\Sigma}$ with certainty.

- We can show that the posterior distribution of vec $\mathcal{A}$ is gaussian with mean a_1 and covariance matrix Ω_1:

$$\Omega_1 = \left(\Omega_0^{-1} + \Sigma^{-1} \otimes Z'Z\right)^{-1}$$

$$a_1 = \Omega_1 \left[\Omega_0^{-1} a_0 + \left(\Sigma^{-1} \otimes Z'Z\right) \text{vec}\widehat{\mathcal{A}}\right]$$

Proof

The posterior kernel is:

$$\mathcal{K}(\mathcal{A}|\mathcal{Y}_T^\star) = \exp\left\{-\frac{1}{2}\left[(\mathrm{vec}\mathcal{A} - a_0)'\Omega_0^{-1}(\mathrm{vec}\mathcal{A} - a_0) + \mathrm{tr}\left(\Sigma^{-1}(\mathcal{A} - \widehat{\mathcal{A}})'Z'Z(\mathcal{A} - \widehat{\mathcal{A}})\right)\right]\right\}$$
$$\times (2\pi)^{-\frac{km}{2}}|\Omega_0|^{-\frac{1}{2}}(2\pi)^{-\frac{mT}{2}}|\Sigma|^{-\frac{T}{2}}e^{-\frac{1}{2}\mathrm{tr}\Sigma^{-1}\widehat{S}}$$

Let $a = \mathrm{vec}\ \mathcal{A}$, $\widehat{a} = \mathrm{vec}\ \widehat{\mathcal{A}}$ and

$$\mathcal{B}(a) = (\mathrm{vec}\mathcal{A} - a_0)'\Omega_0^{-1}(\mathrm{vec}\mathcal{A} - a_0) + \mathrm{tr}\left(\Sigma^{-1}(\mathcal{A} - \widehat{\mathcal{A}})'Z'Z(\mathcal{A} - \widehat{\mathcal{A}})\right)$$

we have:

$$\mathcal{B}(a) = (a - a_0)'\Omega_0^{-1}(a - a_0) + (a - \widehat{a})'\left(\Sigma^{-1} \otimes Z'Z\right)(a - \widehat{a})$$

$$\Leftrightarrow \mathcal{B}(a) = a'\Omega_0^{-1}a + a_0'\Omega_0^{-1}a_0 - 2a'\Omega_0^{-1}a_0$$
$$+ a'\left(\Sigma^{-1} \otimes Z'Z\right)a + \widehat{a}'\left(\Sigma^{-1} \otimes Z'Z\right)\widehat{a} - 2a'\left(\Sigma^{-1} \otimes Z'Z\right)\widehat{a}$$

$$\Leftrightarrow \mathcal{B}(a) = a'\left(\Omega_0^{-1} + \Sigma^{-1} \otimes Z'Z\right) a - 2a'\left(\Omega_0^{-1} a_0 + \left(\Sigma^{-1} \otimes Z'Z\right) \widehat{a}\right) + a_0'\Omega_0^{-1} a_0 + \widehat{a}'\left(\Sigma^{-1} \otimes Z'Z\right) \widehat{a}$$

$$\Leftrightarrow \mathcal{B}(a) = (a-a_1)'\Omega_1^{-1}(a-a_1) - a_1'\Omega_1^{-1}a_1 + a_0'\Omega_0^{-1}a_0 + \widehat{a}'\left(\Sigma^{-1} \otimes Z'Z\right)\widehat{a}$$

By substitution in the posterior kernel:

$$\mathcal{K}(\mathcal{A}|\mathcal{Y}_T^\star) = \exp\left\{-\frac{1}{2}(a-a_1)'\Omega_1^{-1}(a-a_1)\right\} \times \exp\left\{-\frac{1}{2}\left[a_0'\Omega_0^{-1}a_0 + \widehat{a}'\left(\Sigma^{-1} \otimes Z'Z\right)\widehat{a} - a_1'\Omega_1^{-1}a_1\right]\right\} \times (2\pi)^{-\frac{km}{2}}|\Omega_0|^{-\frac{1}{2}}(2\pi)^{-\frac{mT}{2}}|\Sigma|^{-\frac{T}{2}}e^{-\frac{1}{2}\mathrm{tr}\Sigma^{-1}\widehat{S}}$$

Only the first term depends on a. Thus the posterior distribution is gaussian. Integrating with respect to a we obtain the marginal density:

$$
\begin{aligned}
p(\mathcal{Y}_T^\star) &= \int \mathcal{K}(\mathcal{A}|\mathcal{Y}_T^\star)\mathrm{d}\mathcal{A} \\
&= (2\pi)^{\frac{km}{2}} |\Omega_1|^{\frac{1}{2}} \\
&\times \exp\left\{-\frac{1}{2}\left[a_0'\Omega_0^{-1}a_0 + \widehat{a}'\left(\Sigma^{-1}\otimes Z'Z\right)\widehat{a} - a_1'\Omega_1^{-1}a_1\right]\right\} \\
&\times (2\pi)^{-\frac{km}{2}}|\Omega_0|^{-\frac{1}{2}}(2\pi)^{-\frac{mT}{2}}|\Sigma|^{-\frac{T}{2}}e^{-\frac{1}{2}\mathrm{tr}\Sigma^{-1}\widehat{S}}
\end{aligned}
$$

- How should we choose the prior means and variances?
- A famous prior, known to be quite good in forecasting, is the **Minnesota prior**.
- According to this prior, $\{Y_t\}_{t\in\mathbb{N}}$ is generated by m uncorrelated random walks.
- For instance, if we choose the normal prior shape (considering Σ as diagonal and known) the prior mean is defined as:

$$\mathcal{A}_0 = (I_m, 0_{m,m(p-1)+q})'$$

- And the prior variance is as follows:
  - Ω_0 is diagonal (priors over different autoregressive parameters are independent)
  - For the autoregressive parameters we have:

$$\mathbb{V}\left[(A_k)_{i,j}\right] = \begin{cases} \frac{\gamma_1}{k^{\gamma_2}} & \text{if } i = j \\ \frac{\sigma_j^2}{\sigma_i^2} \frac{\gamma_3}{k^{\gamma_2}} & \text{otherwise.} \end{cases}$$

    for $k = 1, \dots, p$ and $(i, j) \in \{1, 2, \dots, m\}^2$.
  - ... and a nearly diffuse prior is assumed for the deterministic part ($\gamma_4 \to \infty$).

$$\mathbb{V}[C_{i,j}] = \gamma_4 \sigma_j^2$$

- A standard calibration for the hyper-parameters γ_i is:

γ_1	γ_2	γ_3	γ_4
5.0×10^{-2}	5.0×10^{-3}	1 or 2	1.0×10^{5}

Table 1: From Kadiyala & Karlsson (JAE, 1997)

- σ_i^2, for $i = 1, \ldots, m$, is set to s_i^2, the estimated residual variance of a p-lag univariate auto-regression for variable i.

- Compute the posterior mode.
- Compute the marginal density & Comparison with a DSGE model.
- Forecasts & Irfs with "error bands" acknowledging the uncertainty on the VAR model. This can be done
  - using the posterior densities obtained analytically previously, or
  - using Metropolis-Hastings or Gibbs sampling algorithm.
- The posterior mode may be obtained using an optimization package or a mixed estimation strategy (Theil & Goldberger, Sims).

- The likelihood of a VAR(p) is:

$$p\,(Y,\mathcal{A},\Sigma) \propto |\Sigma|^{-\frac{T}{2}} e^{-\frac{1}{2}\mathrm{tr}\left[\Sigma^{-1}(Y'Y - \mathcal{A}'Z'Y - Y'X\mathcal{A} + \mathcal{A}'Z'Z\mathcal{A})\right]}$$

- An alternative way of introducing priors is to augment the sample with artificial data generated by a model consistent with our prior.

- Let (Y^*, Z^*) be the artificial data, its likelihood is:

$$p\,(Y^*,\mathcal{A},\Sigma) \propto |\Sigma|^{-\frac{\lambda T}{2}} e^{-\frac{1}{2}\mathrm{tr}\left[\Sigma^{-1}\left(Y^{*\prime}Y^* - \mathcal{A}'Z^{*\prime}Y^* - Y^{*\prime}X^*\mathcal{A} + \mathcal{A}'Z^{*\prime}Z^*\mathcal{A}\right)\right]}$$

  where $\lambda \in \mathbb{R}_+$ gives the weight of our prior compared to the likelihood associated to real data ($T^* = \lambda T$ is the size of the artificial sample).

- For instance, we implement *minnesota* priors by building matrices (Y^*, Z^*) from the simulations of independent random walks.

- Intuition: By augmenting the sample with random walk artificial data, the ML (or OLS) estimator is shrunk towards the unit root. We would obtain the same result by mixing the likelihood (on real data) with a gaussian unit root prior.

- Del Negro & Schorfheide follow this strategy, but they use artificial data from a DSGE model...

- More formally, the joint density of artificial data and real data (conditional on model's parameter) is:

$$p\left(Y^*(\theta), Y|\mathcal{A}, \Sigma\right) = p\left(Y^*(\theta)|\mathcal{A}, \Sigma\right) \times p\left(Y|\mathcal{A}, \Sigma\right)$$

- θ is a vector of parameters defining the Data Generating Process of the artificial data (a unit root VAR model or a DSGE model).

- The first term on the RHS may be viewed as a prior density for $\mathcal{A}$ and Σ.

- Uses a dummy variable approach.
- Priors close to Minnesota.
- All computations (posterior distribution and posterior marginal density) may be done analytically (Normal–Wishart prior shape).
- Implemented in DYNARE : **dyn_bvar** and **dyn_bvar2**.
- The likelihood associated to artificial data is combined with a diffuse (Jeffrey's) prior... The prior is given by:

$$p\left(Y^*,\mathcal{A},\Sigma\right) = (2\pi)^{-\frac{m\lambda T}{2}}\left|\Sigma\right|^{-\frac{\lambda T}{2}}\left|\Sigma\right|^{-\frac{n+1}{2}} e^{-\frac{1}{2}\operatorname{tr}\left[\Sigma^{-1}\left(Y^{*\prime}Y^* - \mathcal{A}'Z^{*\prime}Y^* - Y^{*\prime}X^*\mathcal{A} + \mathcal{A}'Z^{*\prime}Z^*\mathcal{A}\right)\right]}$$

- m “unit root” dummies, weighted by the hyper-parameter μ [**2**]. If strongly weighted, the posterior mode will be close to the unit root.

- a “co-integration” dummy, weighted by the hyper-parameter λ [**5**]. Forces the appearance of at least one common stochastic trend.

- mp “minnesota” dummies, weighted by the hyper-parameter ζ [**3**].

- VAR(2) model: $\underset{1\times 2}{y_t} = c + y_{t-1}\phi_1 + y_{t-2}\phi_2 + u_t$.
- "minnesota" dummies for ϕ_1:

$$\begin{pmatrix} \zeta\sigma_1 & 0 \\ 0 & \zeta\sigma_2 \end{pmatrix} = \begin{pmatrix} \zeta\sigma_1 & 0 & 0 & 0 & 0 \\ 0 & \zeta\sigma_2 & 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} \phi_1 \\ \phi_2 \\ c \end{pmatrix} + u_t$$

From the first dummy observation we have:

$$\phi_{1,11} \sim \mathcal{N}\left(1, \frac{\Sigma_{u,11}}{\zeta^2\sigma_1^2}\right) \text{ and } \phi_{1,12} \sim \mathcal{N}\left(0, \frac{\Sigma_{u,22}}{\zeta^2\sigma_1^2}\right)$$

- "minnesota" dummies for ϕ_2:

$$\begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix} = \begin{pmatrix} 0 & 0 & \zeta\sigma_1 2^d & 0 & 0 \\ 0 & 0 & 0 & \zeta\sigma_2 2^d & 0 \end{pmatrix} \begin{pmatrix} \phi_1 \\ \phi_2 \\ c \end{pmatrix} + u_t$$

From the first dummy observation we have:

$$\phi_{2,11} \sim \mathcal{N}\left(0, \frac{\Sigma_{u,11}}{\zeta^2\sigma_1^2 2^{2d}}\right) \text{ and } \phi_{2,12} \sim \mathcal{N}\left(0, \frac{\Sigma_{u,22}}{\zeta^2\sigma_1^2 2^{2d}}\right)$$

- The default value for d is 0.5.

- "unit root" dummies:

$$\begin{pmatrix} \mu\bar{y}_1 & 0 \\ 0 & \mu\bar{y}_2 \end{pmatrix} = \begin{pmatrix} \mu\bar{y}_1 & 0 & \mu\bar{y}_1 & 0 & 0 \\ 0 & \mu\bar{y}_2 & 0 & \mu\bar{y}_2 & 0 \end{pmatrix} \begin{pmatrix} \phi_1 \\ \phi_2 \\ c \end{pmatrix} + u_t$$

From the first dummy observation we have:

$$1 - \phi_{1,11} - \phi_{2,11} \sim \mathcal{N}\left(0, \frac{\Sigma_{u,11}}{\mu^2 \bar{y}_1^2}\right)$$

and

$$\phi_{1,12} + \phi_{2,12} \sim \mathcal{N}\left(0, \frac{\Sigma_{u,22}}{\mu^2 \bar{y}_1^2}\right)$$

- How should we choose the hyper-parameters ?
  - `dyn_bvar` implements Sims' hyper-parameters by default. σ_i is the standard deviation of variable i estimated over the initial conditions.
  - `dyn_bvar2` implements Smets & Wouter's hyper-parameters by default. $\{\sigma_i\}$ are the residual standard deviation from the estimation of a VAR(p) model on a pre-sample.
- Experience shows that the results, with respect to the marginal density estimation, are quite sensitive to the way the user choose the hyper-parameters.

- The high sensitivity of the marginal density estimates is problematic if we use BVAR models to evaluate the fit of DSGE model (as in SW)
- Conclusions might be easily reversed by changing the hyper-parameters.
- A comparison exercise would be meaningless if the DSGE model beats a "*ill specified*" BVAR model.
- As the VAR's parameter are not (directly) economically interpretable we may use a data-driven choice of the BVAR's hyper-parameters → we choose the hyper-parameters in order to get the best forecasts from the BVAR model.

- Minnesota BVAR with normal prior as defined earlier.
- Phillips (Econometrica, 1996) shows that the minimization of the following function with respect to the hyper-parameters

$$\mathrm{PIC} = \log|\widehat{\Sigma}| + \frac{1}{T}\log\frac{|\Omega_0^{-1} + \Sigma^{-1}\otimes Z'Z|}{|\Omega_0^{-1} + \Sigma^{-1}\otimes Z_0'Z_0|}$$

  leads to the best BVAR model, among the Minnesota BVARs, in terms of predictions.
- It would be more appropriate to compare our DSGE models to this type of data determined model.

- Another idea is to use a structural model as a prior for the BVAR model instead of an a-theoretical Minnesota prior.
- This is quite simple to implement trough the use of artificial data from a DSGE model.
- The question is: **What is the optimal weight (T^*) of the dsge prior in the bvar model ?**
- If we find that T^* (artificial sample size) is important relative to T, it means that the DSGE model imposes useful restrictions to improve the forecasts abilities of the BVAR model.

- They use the theoretical counterparts of the moments instead of artificial data moments in the likelihood associated to the dummy variables.

- For instance, they replace $Y^{*'}(\theta)Y^*(\theta)$ by

$$\lambda T\mathbb{E}\left[y^{*'}(\theta)y^*(\theta)\right] = \lambda T\Gamma^*_{yy}(\theta)$$

  where $\Gamma^*_{yy}(\theta)$ is the theoretical covariance matrix of the observed variables implied by the DSGE model.

- The DSGE prior of the BVAR model, *ie* the likelihood associated to the artificial data, is (adding a Jeffrey's prior $\Rightarrow$ Normal–Wishart prior):

$$
\begin{aligned}
p(\mathcal{A},\Sigma|\theta) = & c(\theta)^{-1}|\Sigma|^{-\frac{\lambda T+m+1}{2}} \\
& \times e^{-\frac{1}{2}\mathrm{tr}\left[T^*\Sigma^{-1}\left(\Gamma^*_{yy}(\theta)-\mathcal{A}'\Gamma^*_{zy}(\theta)-\Gamma^*_{yz}(\theta)\mathcal{A}+\mathcal{A}'\Gamma^*_{zz}(\theta)\mathcal{A}\right)\right]}
\end{aligned}
$$

  where $c(\theta)$ is a constant of integration.

- Let
  - $\mathcal{A}^*(\theta) = \left[\Gamma^*_{zz}(\theta)\right]^{-1}\Gamma^*_{zy}(\theta)$
  - $\Sigma^*(\theta) = \Gamma^*_{yy}(\theta) - \Gamma^*_{yz}(\theta)\left[\Gamma^*_{zz}(\theta)\right]^{-1}\Gamma^*_{zy}(\theta)$

- Conditionally on θ (the deep parameters of the DSGE model), we have a Normal-Wishart prior:

$$\begin{cases} \mathcal{A}|\Sigma,\theta & \sim MN_{k,m}\left(\mathcal{A}^*(\theta),\Sigma,[\lambda T\Gamma^*_{zz}(\theta)]^{-1}\right) \\ \Sigma|\theta & \sim i\mathcal{W}_m\left(\lambda T\Sigma^*(\theta),\lambda T-k-m\right) \end{cases}$$

- To complete our priors we need to specify a prior distribution over the deep parameters (θ). Finally the BVAR-DSGE model has the following prior:

$$p_0(\mathcal{A},\Sigma,\theta) = p_0(\mathcal{A},\Sigma|\theta)\times p_0(\theta)$$

for an implicit value of λ.

- The moments of the posterior density, which is also Normal-Wishart, are obtained by considering the ML estimate (with real and artificial data):

$$\widetilde{\mathcal{A}}(\theta) = V(\theta)^{-1}\left(\lambda T \Gamma^*_{zz}(\theta)\mathcal{A}^*(\theta) + Z'Z\widehat{\mathcal{A}}\right)$$

$$\begin{aligned}\widetilde{\Sigma}(\theta) =& \frac{1}{(1+\lambda)T}\Big[\left(\lambda T \Gamma^*_{yy}(\theta) + Y'Y\right) \\ & - \left(\lambda T \Gamma^*_{yz}(\theta) + Y'Z\right) V(\theta)^{-1}\left(\lambda T \Gamma^*_{zy}(\theta) + Z'Y\right)\Big]\end{aligned}$$

With $V(\theta) = (\lambda T \Gamma^*_{zz}(\theta) + X'X)$

- Finally:

$$\begin{cases} \mathcal{A}|\Sigma,\theta,\mathcal{Y} & \sim MN_{k,m}\left(\widetilde{\mathcal{A}}(\theta),\Sigma,V(\theta)^{-1}\right) \\ \Sigma|\theta,\mathcal{Y} & \sim i\mathcal{W}_m\left((\lambda+1)T\widetilde{\Sigma}(\theta),(\lambda+1)T-k-n\right) \end{cases}$$

- How should we choose λ ?...

- ... Del Negro and Schorfheide choose the value of λ that maximises the marginal density. They estimate, say, 10 BVAR-DSGE models with different values of λ. For each model they also estimate the marginal density. In the end they select the model (*ie* the value of λ) with the highest marginal density.

- This is not the way we implement BVAR-DSGE in DYNARE.
- Instead of doing a loop over values of λ (each time estimating the model and its marginal density), we estimate λ as another parameter.
- We may have a prior on λ (on the ability of the DSGE model to fit the data)... DYNARE compute the posterior distribution of this ability.
- Our joint prior over $\mathcal{A}$, Σ, λ and θ is given by:

$$p_0\left(\mathcal{A}, \Sigma, \lambda, \theta\right) = p_0(\mathcal{A}, \Sigma|\theta, \lambda)p_0(\theta)p_0(\lambda)$$

a priori we have $\theta \perp \lambda$.

- ... But it is quite hard to find the posterior mode of the BVAR-DSGE model with a standard optimization routine.
- Doesn't seem to work with `mode_compute=1,...,5`.
- I have added a new optimization routine: `mode_compute =6` that I use to initialize the Metropolis-Hastings algorithm.

- To estimate a BVAR-DSGE model with DYNARE
  - you have to declare the parameter `dsge_prior_weight` in the preamble of the `*.mod` file.
  - you may give a value to this parameter, as for any parameter of the DSGE model (necessary if you want to estimate the BVAR-DSGE model calibrating the DSGE prior weight as Del Negro & Schorfheide).
  - you have to specify a prior distribution for `dsge_prior_weight` in the `estimated_params` block.

l

- Add measurement errors in the measurement equation...
- ... These errors may be modeled as a VAR model.
- We can then evaluate the fit of the DSGE model by comparing the (second order) moments of the observed variables with the moments of the measurement errors.
- The share of the variance of observed inflation unexplained by the measurement errors, is the share explained by the DSGE model (measurement errors and structural shocks are orthogonal).