Dynare workshop

Estimating DSGE models with Dynare

Michel Juillard[1]

SUFE, Shanghai, October 28, 2013

Disclaimer: The views expressed here are those exclusively from the author and not of Bank of France.

[1]Bank of France.

Outline

1. Bayesian paradigm
2. Estimation of DSGE model
3. Dynare implementation
4. Examples

(with thanks to Stéphane Adjemian for many of the slides)

Bayesian paradigm (motivations)

- Experience shows that it is quite difficult to estimate a DSGE model by maximum likelihood.
    1. Data are not informative enough... The likelihood is flat in some directions (identification issue). This suggests that (when possible) we should use other sources of information.
    2. DSGE models are misspecified. When a DSGE is estimated by ML or with a "non informative" Bayesian approach (uniform priors) the estimated parameters are often found to be incredible. Using prior informations we can shrink the estimates towards sensible values.
- A related motivation is the relative lack of precision of ML. Prior information reduces the uncertainty.
- Finally, the Bayesian approach allows easier comparison of (non nested) models.

Bayesian paradigm (basics)

- A model ($\mathcal{M}$) defines a joint probability distribution parametrized (by $\theta_{\mathcal{M}}$) function over a sample of variables (say $\mathcal{Y}_T$):
$$p(\mathcal{Y}_T|\theta_{\mathcal{M}},\mathcal{M}) \tag{1}$$
If all the variables are observed this is the likelihood:
$$p(\theta_{\mathcal{M}}|\mathcal{Y}_T,\mathcal{M}) \tag{2}$$
- We assume that our prior information about a set of parameters can be summarized by a joint probability density function. Let the prior density be $p_0(\theta_{\mathcal{M}}|\mathcal{M})$.
- The posterior distribution is given by (Bayes theorem):
$$p_1\left(\theta_{\mathcal{M}}|\mathcal{Y}_T,\mathcal{M}\right) = \frac{p_0\left(\theta_{\mathcal{M}}|\mathcal{M}\right)p(|\theta_{\mathcal{M}}|\mathcal{Y}_T,\mathcal{M})}{p(\mathcal{Y}_T|\mathcal{M})} \tag{3}$$

Bayesian paradigm (basics, cont'd)

- Where the denominator defined by

$$p\left(\mathcal{Y}_T|\mathcal{M}\right) = \int_\Theta p_0\left(\theta_\mathcal{M}|\mathcal{M}\right) p(\theta_\mathcal{M}|\mathcal{Y}_T, \mathcal{M})\mathrm{d}\theta_\mathcal{M} \qquad (4)$$

  is the marginal density of the sample (used for model comparison). A weighted mean of the sample conditional densities over the space of possible values for the parameters.
- The posterior density is proportional to the product of the prior density and the likelihood

$$p_1\left(\theta_\mathcal{M}|\mathcal{Y}_T\right) \propto p_0\left(\theta|\mathcal{M}\right) p\left(\mathcal{Y}_T|\theta_\mathcal{M}|\mathcal{Y}_T, \mathcal{M}\right)$$

- The prior density affects the shape of the posterior density!...

A simple example (I)

- Data Generating Process

$$y_t = \mu + \varepsilon_t$$

where $\varepsilon_t \sim \mathcal{N}(0,1)$ is a Gaussian white noise.

- Let $\mathcal{Y}_T \equiv (y_1, \dots, y_T)$. The likelihood is given by:

$$p(\mathcal{Y}_T | \mu) = (2\pi)^{-\frac{T}{2}} e^{-\frac{1}{2}\sum_{t=1}^{T}(y_t-\mu)^2}$$

- And the ML estimator of μ is:

$$\widehat{\mu}_{ML,T} = \frac{1}{T}\sum_{t=1}^{T} y_t \equiv \overline{y}$$

A simple example (II)

- Note that the variance of this estimator is a simple function of the sample size
$$\mathbb{V}[\widehat{\mu}_{ML,T}] = \frac{1}{T}$$
- Noting that:
$$\sum_{t=1}^{T}(y_t - \mu)^2 = \nu s^2 + T(\mu - \widehat{\mu})^2$$
with $\nu = T - 1$ and $s^2 = (T-1)^{-1}\sum_{t=1}^{T}(y_t - \widehat{\mu})^2$.
- The likelihood can be equivalently written as:
$$p(\mathcal{Y}_T|\mu) = (2\pi)^{-\frac{T}{2}} e^{-\frac{1}{2}\left(\nu s^2 + T(\mu - \widehat{\mu})^2\right)}$$
The two statistics s^2 and $\widehat{\mu}$ are summing up the sample information.

A simple example (II, bis)

$$
\begin{aligned}
\sum_{t=1}^{T}(y_t-\mu)^2 &= \sum_{t=1}^{T}([y_t-\widehat{\mu}]-[\mu-\widehat{\mu}])^2 \\
&= \sum_{t=1}^{T}(y_t-\widehat{\mu})^2+\sum_{t=1}^{T}(\mu-\widehat{\mu})^2-\sum_{t=1}^{T}(y_t-\widehat{\mu})(\mu-\widehat{\mu}) \\
&= \nu s^2+T(\mu-\widehat{\mu})^2-\left(\sum_{t=1}^{T}y_t-T\widehat{\mu}\right)(\mu-\widehat{\mu}) \\
&= \nu s^2+T(\mu-\widehat{\mu})^2
\end{aligned}
$$

The last term cancels out by definition of the sample mean.

A simple example (III)

- Let our prior be a Gaussian distribution with expectation μ_0 and variance σ_μ^2.
- The posterior density is defined, up to a constant, by:

$$p\left(\mu|\mathcal{Y}_T\right) \propto \left(2\pi\sigma_\mu^2\right)^{-\frac{1}{2}} e^{-\frac{1}{2}\frac{(\mu-\mu_0)^2}{\sigma_\mu^2}} \times (2\pi)^{-\frac{T}{2}} e^{-\frac{1}{2}\left(\nu s^2+T(\mu-\widehat{\mu})^2\right)}$$

where the missing constant (denominator) is the marginal density (does not depend on μ).
- We also have:

$$p(\mu|\mathcal{Y}_T) \propto \exp\left\{-\frac{1}{2}\left(T(\mu-\widehat{\mu})^2+\frac{1}{\sigma_\mu^2}(\mu-\mu_0)^2\right)\right\}$$

A simple example (IV)

$$
\begin{aligned}
A(\mu) &= T(\mu - \widehat{\mu})^2 + \frac{1}{\sigma_\mu^2}(\mu - \mu_0)^2 \\
&= T\left(\mu^2 + \widehat{\mu}^2 - 2\mu\widehat{\mu}\right) + \frac{1}{\sigma_\mu^2}\left(\mu^2 + \mu_0^2 - 2\mu\mu_0\right) \\
&= \left(T + \frac{1}{\sigma_\mu^2}\right)\mu^2 - 2\mu\left(T\widehat{\mu} + \frac{1}{\sigma_\mu^2}\mu_0\right) + \left(T\widehat{\mu}^2 + \frac{1}{\sigma_\mu^2}\mu_0^2\right) \\
&= \left(T + \frac{1}{\sigma_\mu^2}\right)\left[\mu^2 - 2\mu\frac{T\widehat{\mu} + \frac{1}{\sigma_\mu^2}\mu_0}{T + \frac{1}{\sigma_\mu^2}}\right] + \left(T\widehat{\mu}^2 + \frac{1}{\sigma_\mu^2}\mu_0^2\right) \\
&= \left(T + \frac{1}{\sigma_\mu^2}\right)\left[\mu - \frac{T\widehat{\mu} + \frac{1}{\sigma_\mu^2}\mu_0}{T + \frac{1}{\sigma_\mu^2}}\right]^2 + \left(T\widehat{\mu}^2 + \frac{1}{\sigma_\mu^2}\mu_0^2\right) \\
&\qquad - \frac{\left(T\widehat{\mu} + \frac{1}{\sigma_\mu^2}\mu_0\right)^2}{T + \frac{1}{\sigma_\mu^2}}
\end{aligned}
$$

A simple example (V)

- Finally we have:

$$p(\mu|\mathcal{Y}_T) \propto \exp\left\{-\frac{1}{2}\left(T+\frac{1}{\sigma_\mu^2}\right)\left[\mu-\frac{T\widehat{\mu}+\frac{1}{\sigma_\mu^2}\mu_0}{T+\frac{1}{\sigma_\mu^2}}\right]^2\right\}$$

- Up to a constant, this is a Gaussian density with (posterior) expectation:

$$\mathbb{E}\left[\mu\right]=\frac{T\widehat{\mu}+\frac{1}{\sigma_\mu^2}\mu_0}{T+\frac{1}{\sigma_\mu^2}}$$

and (posterior) variance:

$$\mathbb{V}\left[\mu\right]=\frac{1}{T+\frac{1}{\sigma_\mu^2}}$$

A simple example (VI, The bridge)

- The posterior mean is a convex combination of the prior mean and the ML estimate.
  - If $\sigma_\mu^2 \rightarrow \infty$ (no prior information) then $\mathbb{E}[\mu] \rightarrow \widehat{\mu}$ (ML).
  - If $\sigma_\mu^2 \rightarrow 0$ (calibration) then $\mathbb{E}[\mu] \rightarrow \mu_0$.
- If $\sigma_\mu^2 < \infty$ then the variance of the ML estimator is greater than the posterior variance.
- Not so simple if the model is non linear in the estimated parameters...
  - Asymptotic approximation.
  - Simulation based approach.

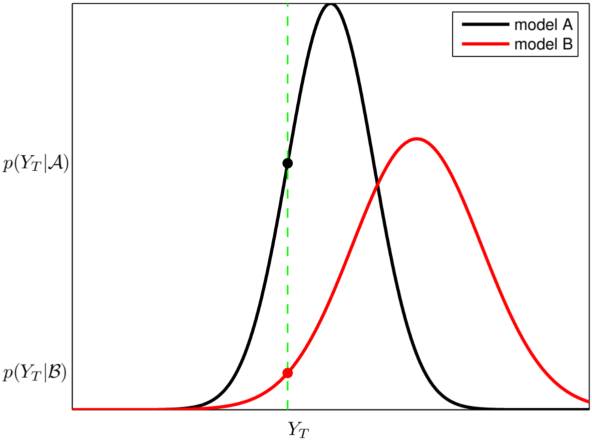
Bayesian paradigm (II, Model Comparison)

- Suppose we have two models $\mathcal{A}$ and $\mathcal{B}$ (with two associated vectors of deep parameters $\theta_{\mathcal{A}}$ and $\theta_{\mathcal{B}}$) estimated using the same sample $\mathcal{Y}_T$.
- For each model $\mathcal{I} = \mathcal{A}, \mathcal{B}$ we can evaluate, at least theoretically, the marginal density of the data conditional on the model:

$$p(\mathcal{Y}_T|\mathcal{I}) = \int_{\Theta_{\mathcal{I}}} p(\theta_{\mathcal{I}}|\mathcal{I}) \times p(\mathcal{Y}_T^{\star}|\theta_{\mathcal{I}}, \mathcal{I}) d\theta_{\mathcal{I}}$$

  by integrating out the deep parameters $\theta_{\mathcal{I}}$ from the posterior kernel.
- $p(\mathcal{Y}_T^{\star}|\mathcal{I})$ measures the fit of model $\mathcal{I}$.

Bayesian paradigm (II, Model Comparison)



Bayesian paradigm (II, Model Comparison)

- Suppose we have a prior distribution over models $\mathcal{A}$ and $\mathcal{B}$: $p(\mathcal{A})$ and $p(\mathcal{B})$.
- Again, using the Bayes theorem we can compute the posterior distribution over models:

$$p(\mathcal{I}|\mathcal{Y}_T^\star) = \frac{p(\mathcal{I})p(\mathcal{Y}_T^\star|\mathcal{I})}{\sum_{\mathcal{I}=\mathcal{A},\mathcal{B}} p(\mathcal{I})p(\mathcal{Y}_T^\star|\mathcal{I})}$$

- This formula may easily be generalized to a collection of N models.
- Posterior odds ratio:

$$\frac{p(\mathcal{A}|\mathcal{Y}_T^\star)}{p(\mathcal{B}|\mathcal{Y}_T^\star)} = \frac{p(\mathcal{A})}{p(\mathcal{B})}\frac{p(\mathcal{Y}_T^\star|\mathcal{A})}{p(\mathcal{Y}_T^\star|\mathcal{B})}$$

DSGE models (I, structural form)

- Our model is given by:

$$\mathbb{E}_t \left[f_\theta(y_{t+1}, y_t, y_{t-1}, u_t) \right] = 0 \tag{5}$$

  with $u_t \sim \mathrm{iid}\,(0, \Sigma)$ is a random vector $(r \times 1)$ of structural innovations, $y_t \in \Lambda \subseteq \mathbb{R}^n$ a vector of endogenous variables, $\mathcal{F}_\theta : \Lambda^3 \times \mathbb{R}^r \to \Lambda$ a real function in $\mathcal{C}^2$ parameterized by a real vector $\theta \in \Theta \subseteq \mathbb{R}^q$ gathering the deep parameters of the model.
- The model is stochastic, forward looking and non linear.
- We want to estimate (a subset of) θ. For any estimation approach (indirect inference, simulated moments, maximum likelihood,...) we need first to solve this model.

DSGE models (II, reduced form)

- We assume that a unique, stable and invariant, solution exists. This solution is a non linear stochastic difference equation:

$$y_t = \mathcal{G}_\theta\left(y_{t-1}, \varepsilon_t\right) \tag{6}$$

  The endogenous variables are written as a function of their past levels and the contemporaneous structural shocks. $\mathcal{G}_\theta$ collects the policy rules and transition functions.
- Generally, it is not possible to get a closed form solution and we have to consider an approximation (local or global) of the true solution (6).
- Dynare uses a local approximation around the deterministic steady state. Global approximations are not yet implemented in DYNARE.

Local approximation of the reduced form (I)

- The deterministic steady state is defined by the following system of n equations:

$$f_\theta\left(\bar{y}(\theta), \bar{y}(\theta), \bar{y}(\theta), 0\right) = 0$$

- The steady state depends on the deep parameters θ. Even for medium scaled models, as in Smets and Wouters, it is often possible to obtain a closed form solution for the steady state $\Rightarrow$ Must be supplied to DYNARE.
- Obviously, function $\mathcal{G}_\theta$ must satisfy the following equality:

$$\bar{y}(\theta) = \mathcal{G}_\theta\left(\bar{y}(\theta), 0\right)$$

- Once the steady state is known, we can compute the Jacobian matrix associated to f_θ...

Local approximation of the reduced form (II)

- Finally the local dynamic is given by:

$$y_t = \bar{y}(\theta) + g_y(\theta)\left(y_{t-1} - \bar{y}(\theta)\right) + g_u(\theta)\varepsilon_t \tag{7}$$

  where $\bar{y}(\theta)$, $g_y(\theta)$ and $g_u(\theta)$ are nonlinear functions of the deep parameters.
- This result can be used to approximate the theoretical moments:

$$\begin{aligned}\mathbb{E}_\infty[y_t] &= \bar{y}(\theta)\\ \mathbb{V}_\infty[y_t] &= g_y(\theta)\mathbb{V}_\infty[y_t]g_y(\theta)' + g_u(\theta)\Sigma g_u(\theta)'\end{aligned}$$

- Equation (7) can also be used to approximate the likelihood...

Estimation (I, Likelihood)

- A direct estimation approach is to maximize the likelihood with respect to θ and $\mathrm{vech}\,(\Sigma)$.
- All the endogenous variables are not observed! Let $y_t^\star$ be a subset of y_t gathering all the observed variables.
- To bring the model to the data, we use a state-space representation:

$$y_t^\star = Zy_t + \eta_t \tag{8a}$$

$$y_t = \mathcal{G}_\theta\left(y_{t-1}, u_t\right) \tag{8b}$$

Equation (8b) is the reduced form of the DSGE model $\Rightarrow$ *state equation*. Equation (8a) selects a subset of the endogenous variables (Z is a $m \times n$ matrix) and a non structural error may be added $\Rightarrow$ *measurement equation*.

Estimation (II, Likelihood)

- Let $\mathcal{Y}_T^\star = \{y_1^\star, y_2^\star, \dots, y_T^\star\}$ be the sample.
- Let ψ be the vector of parameters to be estimated (θ,$\mathrm{vech}\left(\Sigma\right)$ and the covariance matrix of η).
- The likelihood, that is the density of $\mathcal{Y}_T^\star$ conditionally on the parameters, is given by:

$$\mathcal{L}(\psi;\mathcal{Y}_T^\star) = p\left(\mathcal{Y}_T^\star|\psi\right) = p\left(y_0^\star|\psi\right)\prod_{t=1}^{T} p\left(y_t^\star|\mathcal{Y}_{t-1}^\star,\psi\right) \tag{9}$$

- To evaluate the likelihood we need to specify the marginal density $p\left(y_0^\star|\psi\right)$ (or $p\left(y_0|\psi\right)$) and the conditional density $p\left(y_t^\star|\mathcal{Y}_{t-1}^\star,\psi\right)$.

Estimation (III, Likelihood)

- The state-space model (8), or the reduced form (6), describes the evolution of the endogenous variables' distribution.
- The distribution of the initial condition (y_0) is equal to the ergodic distribution of the stochastic difference equation (so that the distribution of y_t is time invariant).
- If the reduced form is linear (or linearized) and if the disturbances are Gaussian (say $\varepsilon \sim \mathcal{N}(0, \Sigma)$), then the initial (ergodic) distribution is Gaussian:

$$y_0 \sim \mathcal{N}\left(\mathbb{E}_\infty[y_t], \mathbb{V}_\infty[y_t]\right)$$

- Unit roots (diffuse Kalman filter).

Estimation (IV, Likelihood)

- Evaluation of the density of $y_t^\star|\mathcal{Y}_{t-1}^\star$ is not direct, because $y_t^\star$ also depends on unobserved endogenous variables.
- The following identity can be used:

$$p\left(y_t^\star|\mathcal{Y}_{t-1}^\star,\psi\right) = \int_\Lambda p\left(y_t^\star|y_t,\psi\right) p(y_t|\mathcal{Y}_{t-1}^\star,\psi)\mathrm{d}y_t \qquad (10)$$

  The density of $y_t^\star|\mathcal{Y}_{t-1}^\star$ is the mean of the density of $y_t^\star|y_t$ weigthed by the density of $y_t|\mathcal{Y}_{t-1}^\star$.
- The first conditional density is given by the measurement equation (8a).
- A Kalman filter is used to evaluate the density of the latent variables (y_t) conditional on the sample up to time $t-1$ ($\mathcal{Y}_{t-1}^\star$) [$\Rightarrow$ *predictive density*].

Estimation (V, Likelihood & Kalman Filter)

- The Kalman filter can be seen as a Bayesian recursive estimation routine:

$$p\left(y_t|\mathcal{Y}^{\star}_{t-1},\psi\right) = \int_{\Lambda} p\left(y_t|y_{t-1},\psi\right) p\left(y_{t-1}|\mathcal{Y}^{\star}_{t-1},\psi\right) \mathrm{d}y_{t-1} \tag{11a}$$

$$p\left(y_t|\mathcal{Y}^{\star}_{t},\psi\right) = \frac{p\left(y^{\star}_t|y_t,\psi\right) p\left(y_t|\mathcal{Y}^{\star}_{t-1},\psi\right)}{\int_{\Lambda} p\left(y^{\star}_t|y_t,\psi\right) p\left(y_t|\mathcal{Y}^{\star}_{t-1},\psi\right) \mathrm{d}y_t} \tag{11b}$$

- Equation (11a) says that the predictive density of the latent variables is the mean of the density of $y_t|y_{t-1}$, given by the state equation (8b), weigthed by the density y_{t-1} conditional on $\mathcal{Y}^{\star}_{t-1}$ (given by (11b)).
- The update equation (11b) is an application of the Bayes theorem $\rightarrow$ how to update our knowledge about the latent variables when new information (data) becomes available.

Estimation (VI, Likelihood & Kalman Filter)

$$p\left(y_t|\mathcal{Y}_t^\star,\psi\right) = \frac{p\left(y_t^\star|y_t,\psi\right)p\left(y_t|\mathcal{Y}_{t-1}^\star,\psi\right)}{\int_\Lambda p\left(y_t^\star|y_t,\psi\right)p\left(y_t|\mathcal{Y}_{t-1}^\star,\psi\right)\mathrm{d}y_t}$$

- $p\left(y_t|\mathcal{Y}_{t-1}^\star,\psi\right)$ is the *a priori* density of the latent variables at time t.
- $p\left(y_t^\star|y_t,\psi\right)$ is the density of the observation at time t knowing the state and the parameters (this density is obtained from the measurement equation (8a)) $\Rightarrow$ the likelihood associated to $y_t^\star$.
- $\int_\Lambda p\left(y_t^*|y_t,\psi\right)p\left(y_t|\mathcal{Y}_{t-1}^\star,\psi\right)\mathrm{d}y_t$ is the marginal density of the new information.

Estimation (VII, Likelihood & Kalman Filter)

- The evaluation of the likelihood is a computationaly (very) intensive task... Except in some very simple cases. For instance: purely forward IS and Phillips curves with a simple Taylor rule (without lag on the interest rate).
- This comes from the multiple integrals we have to evaluate (to solve the model and to run the Kalman filter).
- But if the model is linear, or if we approximate the model around the deterministic steady state, and if the structural shocks are Gaussian, the recursive system of equations (7) collapses to the well known formulas of the (Gaussian–linear) Kalman filter.

Estimation (VIII, Likelihood & Kalman Filter)

The linear–Gaussian Kalman filter recursion is given by:

$$
\begin{aligned}
v_t &= y_t^\star - \bar{y}(\theta)^* - Z\widehat{y}_t \\
F_t &= ZP_tZ' + \mathbb{V}\left[\eta\right] \\
K_t &= g_y(\theta)P_t g_y(\theta)' F_t^{-1} \\
\widehat{y}_{t+1} &= g_y(\theta)\widehat{y}_t + K_t v_t \\
P_{t+1} &= g_y(\theta)P_t(g_y(\theta) - K_t Z)' + g_u(\theta)\Sigma g_u(\theta)'
\end{aligned}
$$

for $t = 1, \dots, T$, with $\widehat{y}_0$ and P_0 given.
Finally the (log)-likelihood is:

$$
\ln L\left(\psi|\mathcal{Y}_T^\star\right) = -\frac{Tk}{2}\ln(2\pi) - \frac{1}{2}\sum_{t=1}^{T}|F_t| - \frac{1}{2}v_t' F_t^{-1} v_t
$$

Prior Elicitation

- The results may depend heavily on our choice for the prior density **or** the parametrization of the model.
- How to choose the prior ?
  - Subjective choice (data driven or theoretical), **example: the Calvo parameter for the Phillips curve**.
  - Objective choice, **examples: the (optimized) Minnesota prior for VAR (Phillips, 1996).**
- Robustness of the results must be evaluated:
  - Try different parametrization.
  - Use more general prior densities.
  - Uninformative priors.

Bayes' asymptotics (Approximation of the posterior)

- Asymptotically, when the size of the sample (T) grows, the choice of the prior doesn't matter.
- Under general conditions, the posterior distribution is asymptotically Gaussian.

Bayes' asymptotics (Approximation of the posterior) (cont'd)

- Let ψ^* be the posterior mode obtained by maximizing the posterior kernel $\mathcal{K}(\psi) \equiv \mathcal{K}(\psi, \mathcal{Y}_T^\star)$. With a second order Taylor expansion around ψ^*, we have:

$$\log \mathcal{K}(\psi) = \log \mathcal{K}(\psi^*) + (\psi - \psi^*)' \left.\frac{\partial \log \mathcal{K}(\psi)}{\partial \psi}\right|_{\psi=\psi^*} + \frac{1}{2}(\psi - \psi^*)' \left.\frac{\partial^2 \log \mathcal{K}(\psi)}{\partial \psi \partial \psi'}\right|_{\psi=\psi^*} (\psi - \psi^*) + \dots$$

or equivalently:

$$\log \mathcal{K}(\psi) = \log \mathcal{K}(\psi^*) - \frac{1}{2}(\psi - \psi^*)'[\mathcal{H}(\psi^*)]^{-1}(\psi - \psi^*) + \mathcal{O}(||\psi - \psi^*||^3)$$

where $\mathcal{H}(\psi^*)$ is minus the inverse of the Hessian matrix evaluated at the posterior mode.

Bayes' asymptotics (Approximation of the posterior) (cont'd)

- The posterior kernel can be approximated by:

$$\mathcal{K}(\psi) \doteq \mathcal{K}(\psi^*) e^{-\frac{1}{2}(\psi-\psi^*)'[\mathcal{H}(\psi^*)]^{-1}(\psi-\psi^*)}$$

- Up to a constant

$$c = \mathcal{K}(\psi^*)(2\pi)^{\frac{k}{2}} |\mathcal{H}(\theta^*)|^{\frac{1}{2}}$$

  we recognize the density of a multivariate normal distribution.
- Completing for constant of integration we obtain an approximation of the posterior density:

$$p_1(\psi) \doteq (2\pi)^{-\frac{k}{2}} |\mathcal{H}(\psi^*)|^{-\frac{1}{2}} e^{-\frac{1}{2}(\psi-\psi^*)'[\mathcal{H}(\psi^*)]^{-1}(\psi-\psi^*)} \tag{12}$$

- If the model is stationnary the Hessian matrix is of order $\mathcal{O}(T)$, as T tends to infinity the posterior distribution concentrates around the posterior mode.

Bayes' asymptotics (Approximation of the posterior) (cont'd)

- This asymptotic result, allows us to approximate any posterior moment. For instance:

$$\mathbb{E}\left[\varphi(\psi)\right] = \frac{\int_\Psi \varphi(\psi) p(\mathcal{Y}_T^\star|\psi) p_0(\psi)\mathrm{d}\psi}{\int_\Psi p(\mathcal{Y}_T^\star|\psi) p_0(\psi)\mathrm{d}\psi}$$

  Tierney and Kadane (1986) show that if we approximate at order two the numerator around the mode of $\varphi(\psi)p(\mathcal{Y}_T^\star|\psi)p_0(\psi)$ and the denominator around the mode of $p(\mathcal{Y}_T^\star|\psi)p_0(\psi)$ (the posterior mode), then the approximation error is of order $\mathcal{O}(T^{-2})$.
- Except for the marginal density (the constant of integration c) this approach is not yet implemented in DYNARE.
- The asymptotic approximation is reliable iff the true posterior distribution is not too far from the Gaussian distribution.

Simulations for exact posterior analysis

- We need a simulation approach if we want to obtain exact results (*ie* not relying on asymptotic approximation).
- Noting that:

$$\mathbb{E}\left[\varphi(\psi)\right] = \int_{\Psi} \varphi(\psi) p_1(\psi|\mathcal{Y}_T^{\star})\mathrm{d}\psi$$

  we can use the empirical mean of $\left(\varphi(\psi^{(1)}), \varphi(\psi^{(2)}), \dots, \varphi(\psi^{(n)})\right)$, where $\psi^{(i)}$ are draws from the posterior distribution to evaluate the expectation of $\varphi(\psi)$. The approxomation error goes to zero when $n \to \infty$.
- We need to simulate draws from the posterior distribution $\Rightarrow$ Metropolis-Hastings.

Metropolis-Hastings (I)

1. Choose a starting point Ψ^0 & run a loop over 2-3-4.
2. Draw a *proposal* $\Psi^\star$ from a *jumping* distribution

$$J(\Psi^\star|\Psi^{t-1}) = \mathcal{N}(\Psi^{t-1}, c \times \Omega_m)$$

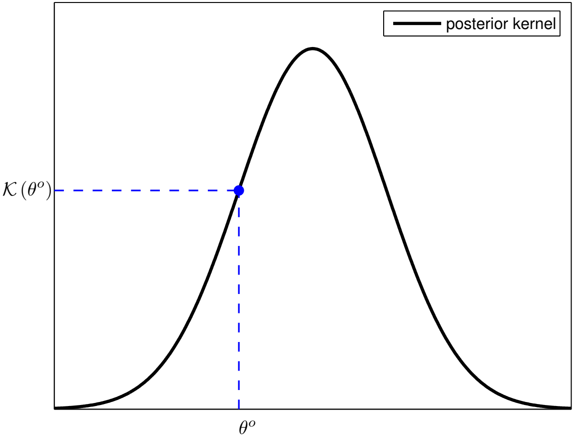
3. Compute the acceptance ratio

$$r = \frac{p_1(\Psi^\star|\mathcal{Y}_T^\star)}{p(\Psi^{t-1}|\mathcal{Y}_T^\star)} = \frac{\mathcal{K}(\Psi^\star|\mathcal{Y}_T^\star)}{\mathcal{K}(\Psi^{t-1}|\mathcal{Y}_T^\star)}$$

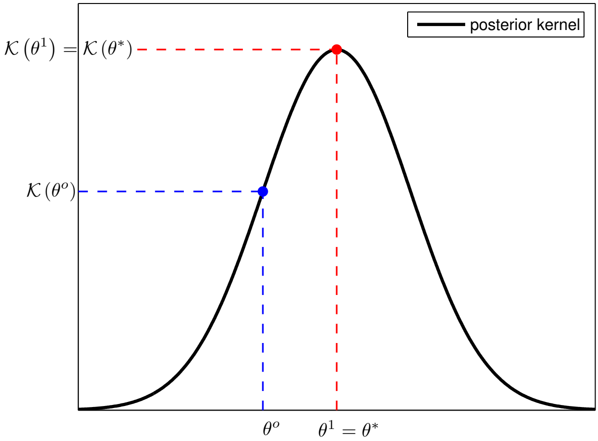
4. Finally

$$\Psi^t = \begin{cases} \Psi^\star & \text{with probability } \min(r, 1) \\ \Psi^{t-1} & \text{otherwise.} \end{cases}$$

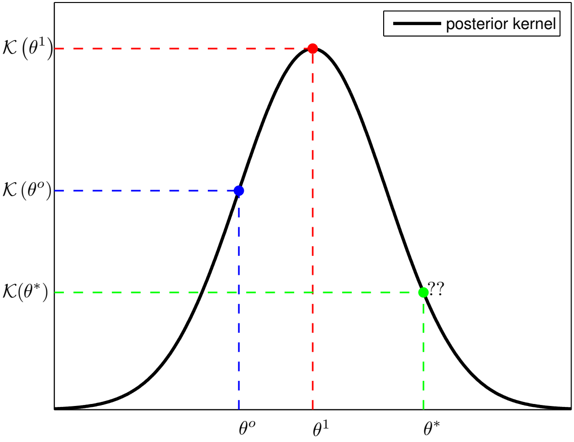
Metropolis-Hastings (II)



Simulations (Metropolis-Hastings)



Simulations (Metropolis-Hastings)



Simulations (Metropolis-Hastings)

- How should we choose the scale factor c (variance of the jumping distribution) ?
- The acceptance ratio should be strictly positive and not too important.

- How many draws ?
- Convergence has to be assessed...
- Parallel Markov chains $\rightarrow$ **Pooled moments** have to be close to **Within moments**.

Marginal density

- The marginal density of the sample may be written as:

$$p(\mathcal{Y}_T^\star|\mathcal{A}) = \int_{\Psi_\mathcal{A}} p(\mathcal{Y}_T^\star, \psi_\mathcal{A}|\mathcal{A})\mathrm{d}\psi_\mathcal{A}$$

- ... or equivalently:

$$p(\mathcal{Y}_T^\star|\mathcal{A}) = \int_{\Psi_\mathcal{A}} \underbrace{p(\mathcal{Y}_T^\star|\Psi_\mathcal{A}, \mathcal{A})}_{\text{likelihood}} \underbrace{p(\psi_\mathcal{A}|\mathcal{A})}_{\text{prior}} \mathrm{d}\psi_\mathcal{A}$$

- We face an integration problem.

Marginal density (Asymptotic approximation)

- For DSGE models we are unable to compute this integral analytically or with standard numerical tools (curse of dimensionality).
- We assume that the posterior distribution is not too far from a Gaussian distribution. In this case we can approximate the marginal density of the sample.
- We have:

$$p(\mathcal{Y}_T^\star|\mathcal{A}) \approx (2\pi)^{\frac{n}{2}}|\mathcal{H}(\psi^*)|^{\frac{1}{2}}p(\mathcal{Y}_T^\star|\psi_{\mathcal{A}}^*,\mathcal{A})p(\psi_{\mathcal{A}}^*|\mathcal{A})$$

- This approach gives accurate estimation of the marginal density if the posterior distribution is uni-modal.

Marginal density (A first simulation based method)

- We can estimate the marginal density using a Monte Carlo

$$\widehat{p}(\mathcal{Y}_T^\star|\mathcal{A}) = \frac{1}{B}\sum_{b=1}^{B} p(\mathcal{Y}_T^\star|\psi_{\mathcal{A}}^{(b)}, \mathcal{A})$$

  where $\psi_{\mathcal{A}}^{(b)}$ is simulated from the prior distribution.
- $\widehat{p}(\mathcal{Y}_T^\star|\mathcal{A}) \underset{B\rightarrow\infty}{\longrightarrow} p(\mathcal{Y}_T^\star|\mathcal{A})$.
- But this method is highly inefficient, because:
  - $\widehat{p}(\mathcal{Y}_T^\star|\mathcal{A})$ has a huge variance.
  - We are not using simulations already done to obtain the posterior distributions (*ie* Metropolis-Hastings).

marginal density (Harmonic mean)

- Note that

$$\mathbb{E}\left[\frac{f(\psi_{\mathcal{A}})}{p(\psi_{\mathcal{A}}|\mathcal{A})p(\mathcal{Y}_T^\star|\psi_{\mathcal{A}},\mathcal{A})}\middle|\psi_{\mathcal{A}},\mathcal{A}\right] = \int_{\Psi_{\mathcal{A}}} \frac{f(\psi_{\mathcal{A}})p(\psi_{\mathcal{A}}|\mathcal{Y}_T^\star,\mathcal{A})}{p(\psi_{\mathcal{A}}|\mathcal{A})p(\mathcal{Y}_T^\star|\psi_{\mathcal{A}},\mathcal{A})}\mathrm{d}\psi_{\mathcal{A}}$$

  where f is any density function.

- The right member of the equality, using the definition of the posterior density, may be rewritten as

$$\int_{\Psi_{\mathcal{A}}} \frac{f(\psi_{\mathcal{A}})}{p(\psi_{\mathcal{A}}|\mathcal{A})p(\mathcal{Y}_T^\star|\psi_{\mathcal{A}},\mathcal{A})} \frac{p(\psi_{\mathcal{A}}|\mathcal{A})p(\mathcal{Y}_T^\star|\psi_{\mathcal{A}},\mathcal{A})}{\int_{\Psi_{\mathcal{A}}} p(\psi_{\mathcal{A}}|\mathcal{A})p(\mathcal{Y}_T^\star|\psi_{\mathcal{A}},\mathcal{A})\mathrm{d}\psi_{\mathcal{A}}}\mathrm{d}\psi_{\mathcal{A}}$$

- Finally, we have

$$\mathbb{E}\left[\frac{f(\psi_{\mathcal{A}})}{p(\psi_{\mathcal{A}}|\mathcal{A})p(\mathcal{Y}_T^\star|\psi_{\mathcal{A}},\mathcal{A})}\middle|\psi_{\mathcal{A}},\mathcal{A}\right] = \frac{\int_{\Psi} f(\psi)\mathrm{d}\psi}{\int_{\Psi_{\mathcal{A}}} p(\psi_{\mathcal{A}}|A)p(\mathcal{Y}_T^\star|\psi_{\mathcal{A}},\mathcal{A})\mathrm{d}\psi_{\mathcal{A}}}$$

Marginal density (Harmonic mean)

- So that

$$p(\mathcal{Y}_T^\star|\mathcal{A}) = \mathbb{E}\left[\frac{f(\psi_\mathcal{A})}{p(\psi_\mathcal{A}|\mathcal{A})p(\mathcal{Y}_T^\star|\psi_\mathcal{A},\mathcal{A})}\middle|\psi_\mathcal{A},\mathcal{A}\right]^{-1}$$

- This suggests the following estimator of the marginal density

$$\widehat{p}(\mathcal{Y}_T^\star|\mathcal{A}) = \left[\frac{1}{B}\sum_{b=1}^{B}\frac{f(\psi_\mathcal{A}^{(b)})}{p(\psi_\mathcal{A}^{(b)}|\mathcal{A})p(\mathcal{Y}_T^\star|\psi_\mathcal{A}^{(b)},\mathcal{A})}\right]^{-1}$$

- Each drawn vector $\psi_\mathcal{A}^{(b)}$ comes from the Metropolis-Hastings Monte Carlo.

Marginal density (Harmonic mean)

- The preceding proof holds if we replace $f(\theta)$ by 1
  $\looparrowright$ Simple Harmonic Mean estimator. But this estimator may also have a huge variance.
- The density $f(\theta)$ may be interpreted as a weighting function, we want to give less importance to extremal values of θ.
- Geweke (1999) suggests to use a truncated Gaussian function (modified harmonic mean estimator).

Marginal density (Harmonic mean)

$$\overline{\psi} = \frac{1}{B}\sum_{b=1}^{B}\psi_{\mathcal{M}}^{(b)}$$

$$\overline{\Omega} = \frac{1}{B}\sum_{b=1}^{B}(\psi_{\mathcal{M}}^{(b)} - \overline{\psi})'(\psi_{\mathcal{M}}^{(b)} - \overline{\psi})$$

- For some $p \in (0,1)$ we define

$$\widetilde{\Psi} = \left\{\psi_{\mathcal{M}} : (\psi_{\mathcal{M}}^{(b)} - \overline{\psi})'\overline{\Omega}^{-1}(\psi_{\mathcal{M}}^{(b)} - \overline{\psi}) \leq \chi^2_{1-p}(n)\right\}$$

- ... And take

$$f(\psi_{\mathcal{M}}) = p^{-1}(2\pi)^{-\frac{n}{2}}|\overline{\Omega}|^{-\frac{1}{2}}e^{-\frac{1}{2}(\psi_{\mathcal{M}}-\overline{\psi})'\overline{\Omega}^{-1}(\psi_{\mathcal{M}}-\overline{\psi})}\mathbb{I}_{\widetilde{\Psi}}(\psi_{\mathcal{M}})$$

Posterior analysis (Credible set)

- A synthetic way to characterize the posterior distribution is to build something like a confidence interval.
- We define:

$$P(\psi \in C) = \int_C p(\psi)\mathrm{d}\psi = 1 - \alpha$$

  is a $100(1-\alpha)\%$ credible set for ψ with respect to $p(\psi)$ (for instance, with $\alpha = 0.2$ we have a 80% credible set).
- A $100(1-\alpha)\%$ highest probability density (HPD) credible set for ψ with respect to $p(\psi)$ is a $100(1-\alpha)\%$ credible set with the property

$$p(\psi_1) \geq p(\psi_2) \quad \forall \psi_1 \in C \text{ and } \forall \psi_2 \in \bar{C}$$

Posterior analysis (density)

- To obtain a complete view about the posterior distribution we can estimate each the marginal posterior densities (for each parameter of the model).
- We use a non parametric estimator:

$$\hat{f}(\psi) = \frac{1}{Nh} \sum_{i=1}^{N} K\left(\frac{\psi - \psi^{(i)}}{h}\right)$$

  where N is the number of draws in the Metropolis, ψ is a point where we want to evaluate the posterior density, $\psi^{(i)}$ is a draw from the Metropolis, $K(\bullet)$ is a kernel (Gaussian by default in DYNARE) and h is a bandwidth parameter.
- In DYNARE the bandwidth parameter is optimally chosen considering the Silverman's rule of thumb.

Posterior analysis (Predictive density)

- Knowing the posterior distribution of the model's parameters, we can forecast the endogenous variables of the model.
- We define the posterior predictive density as follows:

$$p(\tilde{\mathbf{Y}}|\mathcal{Y}_T^\star) = \int_{\Psi_\mathcal{A}} p(\tilde{\mathbf{Y}}, \psi_\mathcal{A}|\mathcal{Y}_T^\star, \mathcal{A})\mathrm{d}\psi_\mathcal{A}$$

where, for instance, $\tilde{\mathbf{Y}}$ might be y_{T+1}. Knowing that $p(\tilde{\mathbf{Y}}, \psi_\mathcal{A}|\mathcal{Y}_T^\star, \mathcal{A}) = p(\tilde{\mathbf{Y}}|\psi_\mathcal{A}, \mathcal{Y}_T^\star, \mathcal{A})p(\psi_\mathcal{A}|\mathcal{Y}_T^\star, \mathcal{A})$ we have:

$$p(\tilde{\mathbf{Y}}|\mathcal{Y}_T^\star) = \int_{\Psi_\mathcal{A}} p(\tilde{\mathbf{Y}}|\psi_\mathcal{A}, \mathcal{Y}_T^\star, \mathcal{A})p(\psi_\mathcal{A}|\mathcal{Y}_T^\star, \mathcal{A})\mathrm{d}\psi_\mathcal{A}$$

Numerical integration with MH-draws

- The Metropolis draws can be used to estimate any moments of the parameters (or function of the parameters).
- We have

$$
\begin{aligned}
\mathbb{E}\left[h(\psi_{\mathcal{A}})\right] &= \int_{\Psi_{\mathcal{A}}} h(\psi_{\mathcal{A}}) p(\psi_{\mathcal{A}}|\mathcal{Y}_T^{\star}, \mathcal{A}) \mathrm{d}\psi_{\mathcal{A}} \\
&\approx \frac{1}{N} \sum_{i=1}^{N} h\left(\psi_{\mathcal{A}}^{(i)}\right)
\end{aligned}
$$

where $\psi_{\mathcal{A}}^{(i)}$ is a Metropolis draw and h is any continuous function.

Posterior analysis (Point estimation)

- The Metropolis-Hastings allows us to estimate the posterior distribution of each deep parameters of a model... But we may be interested in a point estimate (like in classical inference) instead of the entire distribution.
- We have to choose a point in the posterior distribution.
- We define a Bayes risk function:

$$\begin{aligned} R(a) &= \mathbb{E}\left[L(a,\psi)\right] \\ &= \int_{\Psi} L(a,\psi)p(\psi)\mathrm{d}\psi \end{aligned}$$

where $L(a,\psi)$ is the loss function associated with decision a when parameters take value ψ.

Posterior analysis (Point estimation)

Action: deciding that the estimated value of ψ is $\widehat{\psi}$ such that:

$$\widehat{\psi} = \arg\min_{\widetilde{\psi}} \int_{\Psi} L(\widetilde{\psi}, \psi) p(\psi | \mathcal{Y}_T^\star, \mathcal{M}) \mathrm{d}\psi$$

- Quadratic loss function (L_2 norm):

$$\widehat{\psi} = \mathbb{E}(\psi | \mathcal{Y}_T^\star, \mathcal{M})$$

- Absolute value loss function (L_1 norm):

$$\widehat{\psi} = \text{ median of the posterior distribution}$$

- Zero-one loss function: $\widehat{\psi} =$ posterior mode

Rabanal & Rubio-Ramirez 2001 (I)

- New Keynesian models.
- Common equations :
  - $y_t = \mathbb{E}_t y_{t+1} - \sigma(r_t - \mathbb{E}_t \Delta p_{t+1} + \mathbb{E}_t g_{t+1} - g_t)$
  - $y_t = a_t + (1-\delta) n_t$
  - $mc_t = w_t - p_t + n_t - y_t$
  - $mrs_t = \frac{1}{\sigma} y_t + \gamma n_t - g_t$
  - $r_t = \rho_r r_{t-1} + (1-\rho_r)\left[\gamma_\pi \Delta p_t + \gamma_y y_t\right] + z_t$
  - $w_t - p_t = w_{t-1} - p_{t-1} + \Delta w_t - \Delta p_t$
  - $a_t, g_t \sim AR(1)$, z_t, λ_t are Gaussian white noises.

Rabanal & Rubio-Ramirez 2001 (II)

- Baseline sticky prices model (BSP) :
  - $\Delta p_t = \beta \mathbb{E} \left[\Delta p_{t+1} + \kappa_p (mc_t + \lambda_t) \right]$
  - $w_t - p_t = mrs_t$
- Sticky prices & Price indexation (INDP) :
  - $\Delta p_t = \gamma_b \Delta p_{t-1} + \gamma_f \mathbb{E} \left[\Delta p_{t+1} + \kappa'_p (mc_t + \lambda_t) \right]$
  - $w_t - p_t = mrs_t$
- Sticky prices & wages (EHL) :
  - $\Delta p_t = \beta \mathbb{E}_t \left[\Delta p_{t+1} + \kappa_p (mc_t + \lambda_t) \right]$
  - $\Delta w_t = \beta \mathbb{E}_t \left[\Delta w_{t+1} \right] + \kappa_w \left[mrs_t - (w_t - p_t) \right]$
- Sticky prices & wages + Wage indexation (INDW) :
  - $\Delta w_t - \alpha \Delta p_{t-1} = \beta \mathbb{E}_t \left[\Delta w_{t+1} \right] - \alpha \beta \Delta p_t + \kappa_w \left[mrs_t - (w_t - p_t) \right]$

Rabanal & Rubio-Ramirez 2001 (III, with Dynare)

```
var a g mc mrs n pie r rw winf y;

varexo e_a e_g e_lam e_ms;

parameters invsig delta .... ;

model(linear);
 y=y(+1)-(1/invsig)*(r-pie(+1)+g(+1)-g);
 y=a+(1-delta)*n;
 mc=rw+n-y;
     ....
end;
```

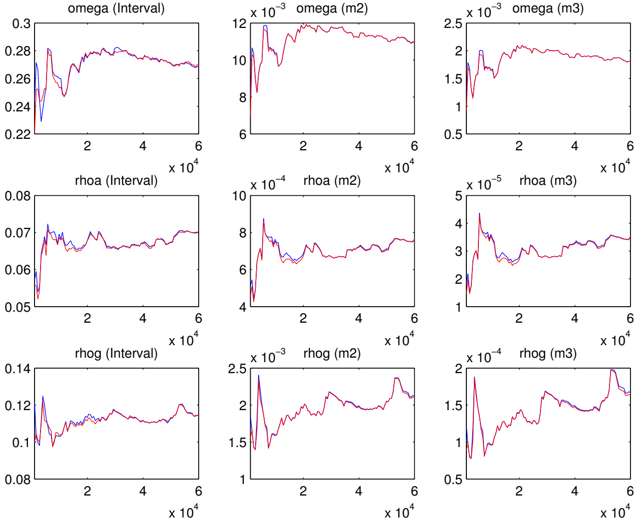
Rabanal & Rubio-Ramirez 2001 (III, with Dynare)

```
estimated_params;
  stderr e_a, uniform_pdf,,,0,1;
  stderr e_lam, uniform_pdf,,,0,1;
    ....
  gampie, normal_pdf, 1.5, 0.25;
    ....
end;

varobs pie r y rw;

estimation(datafile=dataraba,first_obs=10,
  ....,mh_jscale=0.5);
```

Rabanal & Rubio-Ramirez 2001 (IV, Dynare output)

```
RESULTS FROM POSTERIOR MAXIMIZATION
parameters
           prior mean     mode    s.d. t-stat prior pstdev

      invsig   2.500   2.5841  0.3633  7.1131 gamm  1.7600
         gam   1.000   1.5158  0.2874  5.2743 norm  0.5000
         rho   0.500   0.4476  0.0775  5.7789 unif  0.2887
      gampie   1.500   1.7466  0.1938  9.0101 norm  0.2500
        gamy   0.125   0.0678  0.0508  1.3355 gamm  0.0750
        rhoa   0.500   0.5795  0.0451 12.8518 unif  0.2887
        rhog   0.500   0.3903  0.0855  4.5668 unif  0.2887
    thetabig   3.000   1.9085  0.0873 21.8614 gamm  1.4200
standard deviation of shocks
           prior mean     mode    s.d. t-stat prior pstdev

         e_a   0.100   0.7681  0.1157  6.6366 invg     Inf
         e_g   0.100   0.7997  0.1266  6.3183 invg     Inf
        e_ms   0.100   1.0855  0.1049 10.3481 invg     Inf
       e_lam   0.100   2.2880  0.3284  6.9672 invg     Inf

Log data density [Laplace approximation] is -379.287210.
```

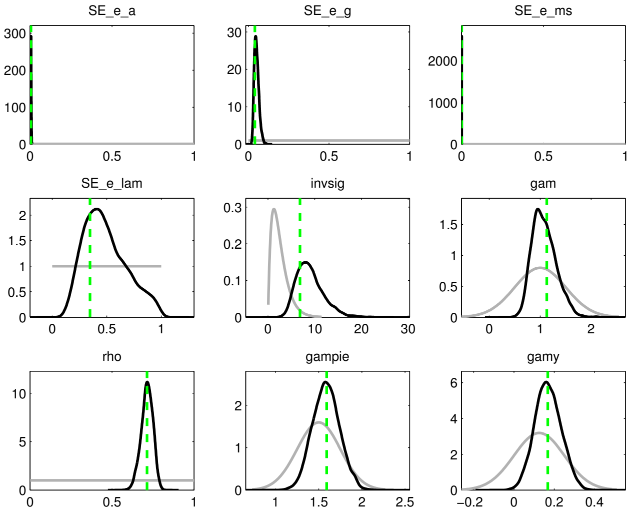
Rabanal & Rubio-Ramirez 2001 (IV, Dynare output)



Rabanal & Rubio-Ramirez 2001 (IV, Dynare output)

```
ESTIMATION RESULTS

Log data density is -380.794630.

parameters
           prior mean post. mean conf. interval prior pstdev

      invsig   2.500   2.6946  2.0995  3.3963 gamm  1.7600
         gam   1.000   1.5024  0.8929  2.0420 norm  0.5000
         rho   0.500   0.4727  0.3405  0.5840 unif  0.2887
      gampie   1.500   1.8074  1.4763  2.0891 norm  0.2500
        gamy   0.125   0.0921  0.0189  0.1790 gamm  0.0750
        rhoa   0.500   0.5527  0.4721  0.6413 unif  0.2887
        rhog   0.500   0.4060  0.2288  0.5576 unif  0.2887
    thetabig   3.000   1.9148  1.7880  2.1413 gamm  1.4200

standard deviation of shocks
           prior mean post. mean conf. interval prior pstdev

         e_a   0.100   0.8318  0.5883  1.1194 invg     Inf
         e_g   0.100   0.8871  0.6838  1.1401 invg     Inf
        e_ms   0.100   1.1076  0.9626  1.2520 invg     Inf
       e_lam   0.100   2.3382  1.7908  3.0452 invg     Inf
```

Rabanal & Rubio-Ramirez 2001 (IV, Dynare output)



Rabanal & Rubio-Ramirez 2001 (IV, Dynare output)