ESTIMATION of a DSGE MODEL

Paris, October 17 2005

STÉPHANE ADJEMIAN

stephane.adjemian@ens.fr

UNIVERSITÉ DU MAINE & CEPREMAP

Bayesian approach

- A bridge between calibration and maximum likelihood.
- Uncertainty and *a priori* knowledge about the model and its parameters are described by prior probabilities.
- Confrontation to the data, through the likelihood function, leads to a revision of these probabilities ($\longrightarrow$ posterior probabilities).
- Testing and model comparison is done by comparing posterior probabilities (over models).

Prior density

- The prior density describes *a priori* beliefs, before observing the data:

$$p(\boldsymbol{\theta}_{\mathcal{A}}|\mathcal{A})$$

- $\mathcal{A}$ stands for a specific model.
- $\boldsymbol{\theta}_{\mathcal{A}}$ represents the parameters of model $\mathcal{A}$.
- $p(\bullet) \rightarrow$ normal, gamma, shifted gamma, inverse gamma, beta, generalized beta, uniform...

Likelihood function

- The likelihood describes the density of the observed data, given the model and its parameters:

$$\mathcal{L}(\boldsymbol{\theta}_{\mathcal{A}}|\mathbf{Y}_T, \mathcal{A}) \equiv p(\mathbf{Y}_T|\boldsymbol{\theta}_{\mathcal{A}}, \mathcal{A})$$

- $\mathbf{Y}_T$ are the observations until period T.

- In our case, the likelihood is recursive:

$$p(\mathbf{Y}_T|\boldsymbol{\theta}_{\mathcal{A}}, \mathcal{A}) = p(y_0|\boldsymbol{\theta}_{\mathcal{A}}, \mathcal{A}) \prod_{t=1}^{T} p(y_t|\mathbf{Y}_{t-1}, \boldsymbol{\theta}_{\mathcal{A}}, \mathcal{A})$$

Bayes theorem

- We have a prior density $p(\boldsymbol{\theta})$ and the likelihood $p(\mathbf{Y}_T|\boldsymbol{\theta})$.
- We are interested in $p(\boldsymbol{\theta}|\mathbf{Y}_T)$, **the posterior density**. Using the Bayes theorem twice we obtain the density of the parameters knowing the data:

We have:

$$p(\boldsymbol{\theta}|\mathbf{Y}_T) = \frac{p(\boldsymbol{\theta};\mathbf{Y}_T)}{p(\mathbf{Y}_T)}$$

and also

$$p(\mathbf{Y}_T|\boldsymbol{\theta}) = \frac{p(\boldsymbol{\theta};\mathbf{Y}_T)}{p(\boldsymbol{\theta})} \Leftrightarrow p(\boldsymbol{\theta};\mathbf{Y}_T) = p(\mathbf{Y}_T|\boldsymbol{\theta}) \times p(\boldsymbol{\theta})$$

Posterior kernel and density

- By substitution, we get the posterior density:

$$p(\boldsymbol{\theta}_{\mathcal{A}}|\mathbf{Y}_T,\mathcal{A}) = \frac{p(\mathbf{Y}_T|\boldsymbol{\theta}_{\mathcal{A}},\mathcal{A})p(\boldsymbol{\theta}_{\mathcal{A}}|\mathcal{A})}{p(\mathbf{Y}_T|\mathcal{A})}$$

- Where $p(\mathbf{Y}_T|\mathcal{A})$ is the marginal density of the data conditional on the model:

$$p(\mathbf{Y}_T|\mathcal{A}) = \int_{\Theta_{\mathcal{A}}} p(\boldsymbol{\theta}_{\mathcal{A}};\mathbf{Y}_T|\mathcal{A})d\boldsymbol{\theta}_{\mathcal{A}}$$

- The posterior kernel (un-normalized posterior density):

$$p(\boldsymbol{\theta}_{\mathcal{A}}|\mathbf{Y}_T,\mathcal{A}) \propto p(\mathbf{Y}_T|\boldsymbol{\theta}_{\mathcal{A}},\mathcal{A})p(\boldsymbol{\theta}_{\mathcal{A}}|\mathcal{A}) \equiv \mathcal{K}(\boldsymbol{\theta}_{\mathcal{A}}|\mathbf{Y}_T,\mathcal{A})$$

A simple example (I)

- Data Generating Process

$$y_t = \mu + \varepsilon_t, \text{ with } t = 1, ..., T$$

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

- The likelihood is given by

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

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

A simple example (II)

- Note that the variance of this estimator is a simple function of the sample size

$$\mathbb{V}[\widehat{\mu}_{ML,T}] = \frac{1}{T}$$

- 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:

$$p\left(\mu|\mathbf{Y}_T\right) \propto \left(2\pi\sigma_\mu^2\right)^{-\frac{1}{2}} e^{-\frac{1}{2}\frac{(\mu-\mu_0)^2}{\sigma_\mu^2}} \times \left(2\pi\right)^{-\frac{T}{2}} e^{-\frac{1}{2}\sum_{t=1}^{T}(y_t-\mu)^2}$$

A simple example (III)

- ... Or equivalently

$$p\left(\mu|\mathbf{Y}_T\right) \propto e^{-\frac{(\mu-\mathbb{E}[\mu])^2}{\mathbb{V}[\mu]}}$$

$\rightarrow$ the posterior distribution is gaussian, with:

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

and

$$\mathbb{E}[\mu] = \frac{\left(\frac{1}{T}\right)^{-1} \widehat{\mu}_{ML,T} + \sigma_\mu^{-2}\mu_0}{\left(\frac{1}{T}\right)^{-1} + \sigma_\mu^{-2}}$$

A simple example (The bridge, IV)

$$\mathbb{E}[\mu] = \frac{\left(\frac{1}{T}\right)^{-1}\widehat{\mu}_{ML,T} + \sigma_{\mu}^{-2}\mu_0}{\left(\frac{1}{T}\right)^{-1} + \sigma_{\mu}^{-2}}$$

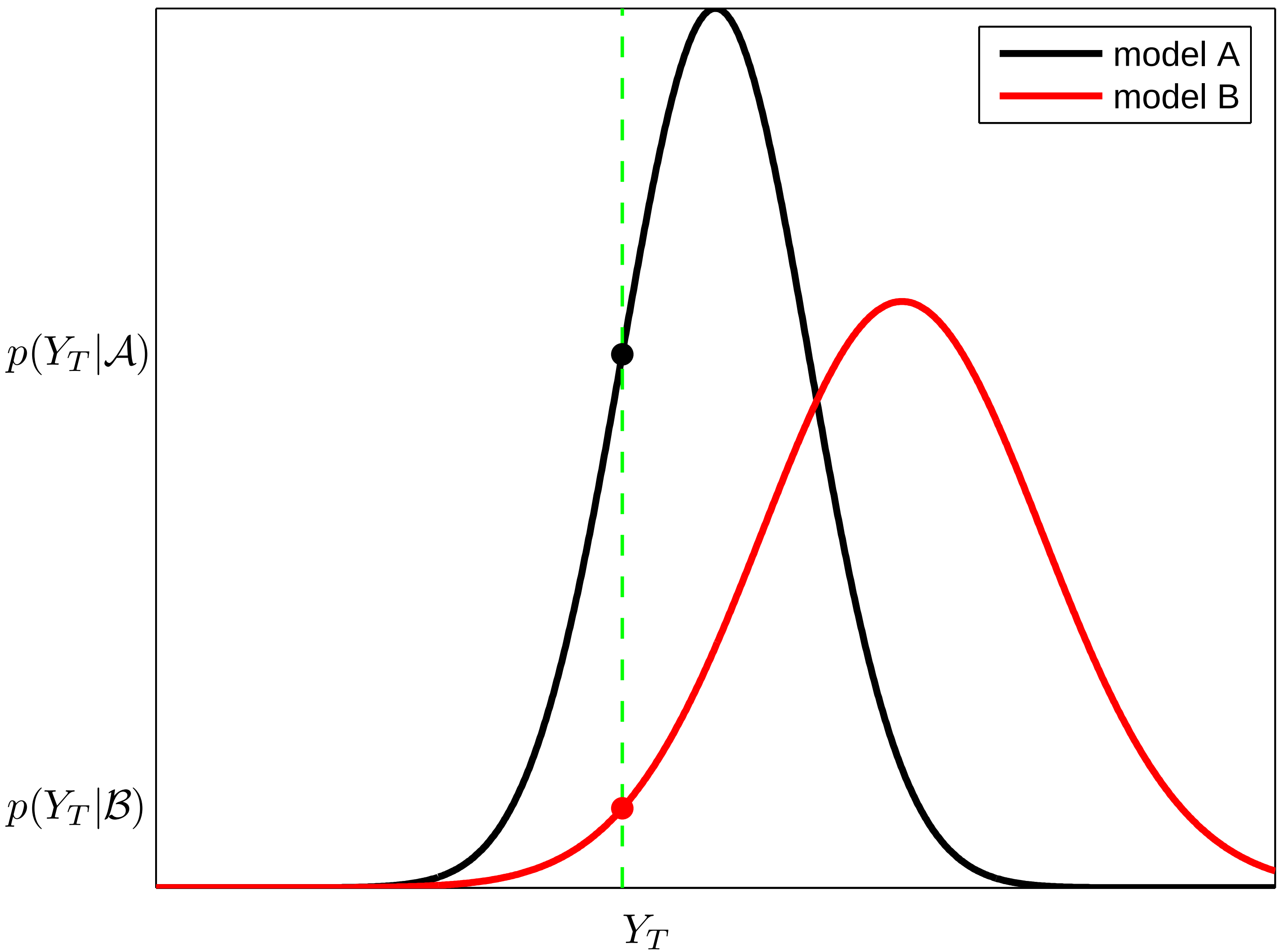
- The posterior mean is a convex combination of the prior mean and the ML estimate.
- If $\sigma_{\mu}^2 \rightarrow \infty$ (no prior information) then $\mathbb{E}[\mu] \rightarrow \widehat{\mu}_{ML,T}$.
- If $\sigma_{\mu}^2 \rightarrow 0$ (calibration) then $\mathbb{E}[\mu] \rightarrow \mu_0$.
- Not so simple if the model is non linear in the estimated parameters... In general we have to follow a simulation based approach to get the posterior distributions and moments.

Model comparison (I)

- Suppose we have two models $\mathcal{A}$ and $\mathcal{B}$ (with two associated vectors of deep parameters $\boldsymbol{\theta}_{\mathcal{A}}$ and $\boldsymbol{\theta}_{\mathcal{B}}$) estimated using the same sample $\mathbf{Y}_T$.
- 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:

$$p(\mathbf{Y}_T|\mathcal{I}) = \int_{\Theta_{\mathcal{I}}} p(\boldsymbol{\theta}_{\mathcal{I}}|\mathcal{I}) \times p(\mathbf{Y}_T|\boldsymbol{\theta}_{\mathcal{I}}, \mathcal{I}) d\boldsymbol{\theta}_{\mathcal{I}}$$

  by integrating out the deep parameters $\boldsymbol{\theta}_{\mathcal{I}}$ from the posterior kernel.
- $p(\mathbf{Y}_T|\mathcal{I})$ measures the fit of model $\mathcal{I}$.


model A
model B
$p(Y_T|\mathcal{A})$
$p(Y_T|\mathcal{B})$
Y_T

Model comparison (II)

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

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

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

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

In practice . . .

- We may be interested in:
  - Estimating some posterior moments.
  - Estimating the posterior marginal density to compare models.
- We can do it by hand for linear models (for instance VAR models) by carefully choosing the priors. . .
- . . . But it's impossible for DSGE models.
- We use a simulation based approach.

Metropolis algorithm (I)

1. Choose a starting point $\boldsymbol{\theta}^\circ$ & run a loop over 2-3-4.

2. Draw a *proposal* $\boldsymbol{\theta}^*$ from a *jumping* distribution

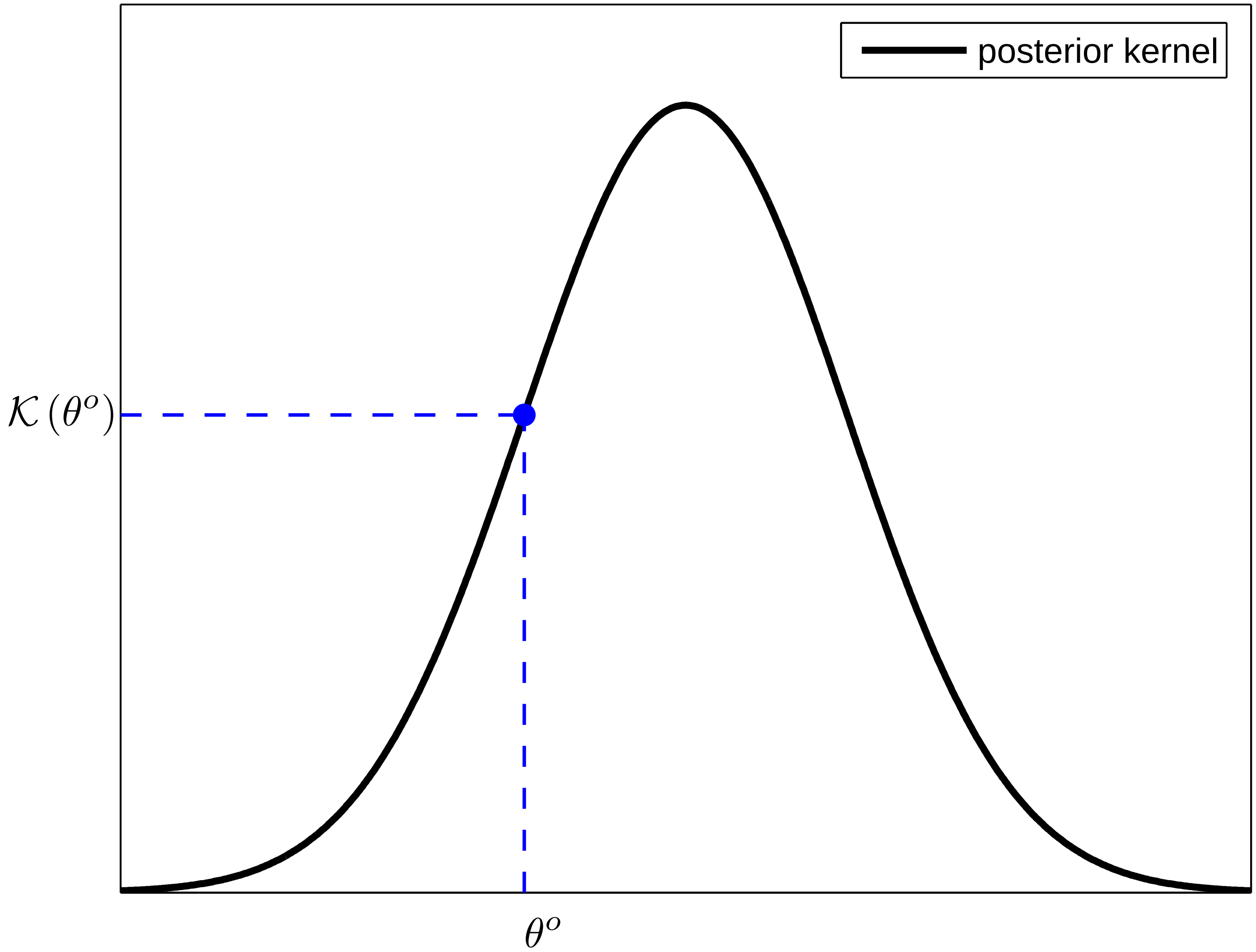
$$J(\boldsymbol{\theta}^*|\boldsymbol{\theta}^{t-1}) = \mathcal{N}(\boldsymbol{\theta}^{t-1}, c\Sigma_m)$$

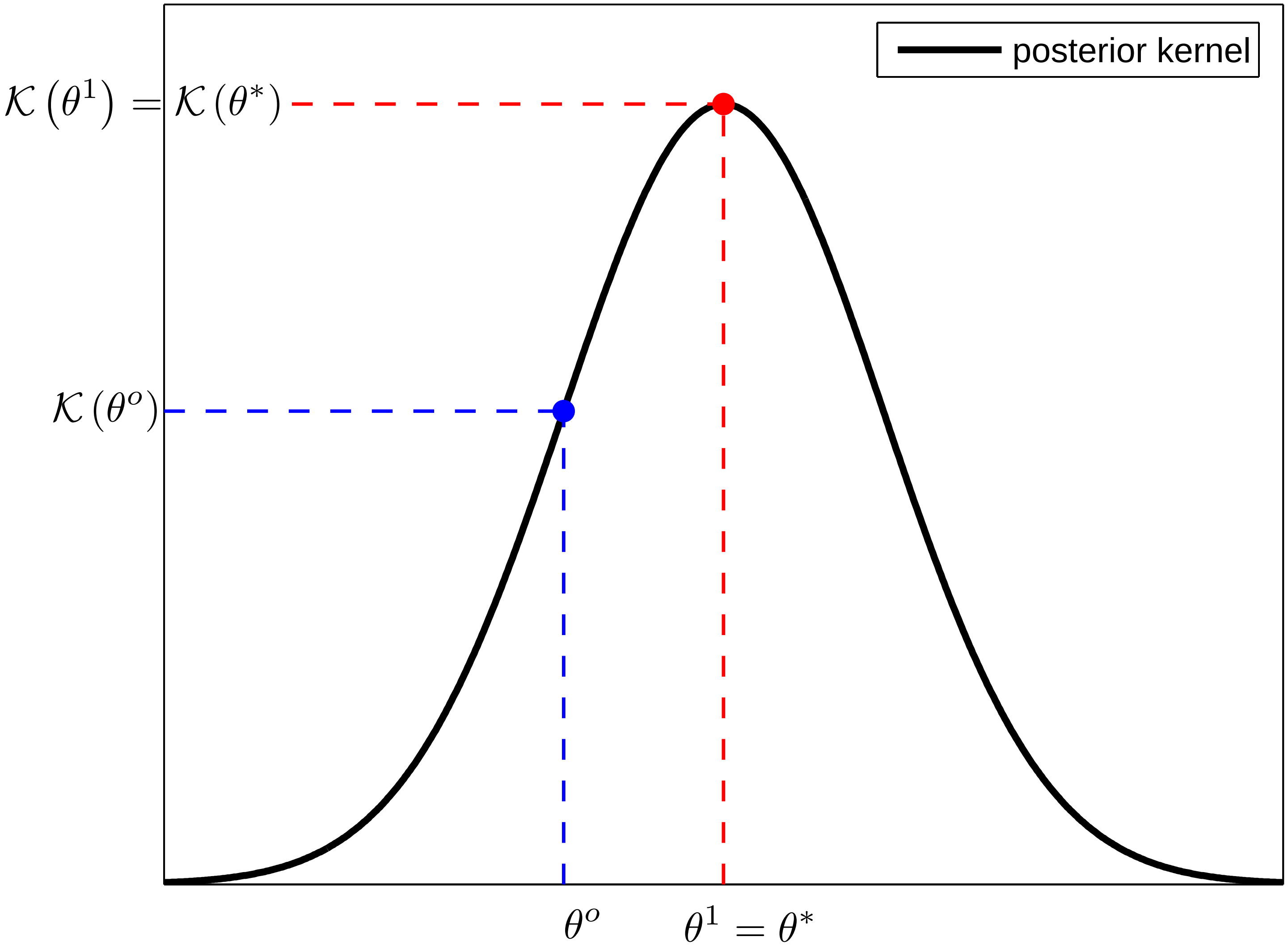
3. Compute the acceptance ratio

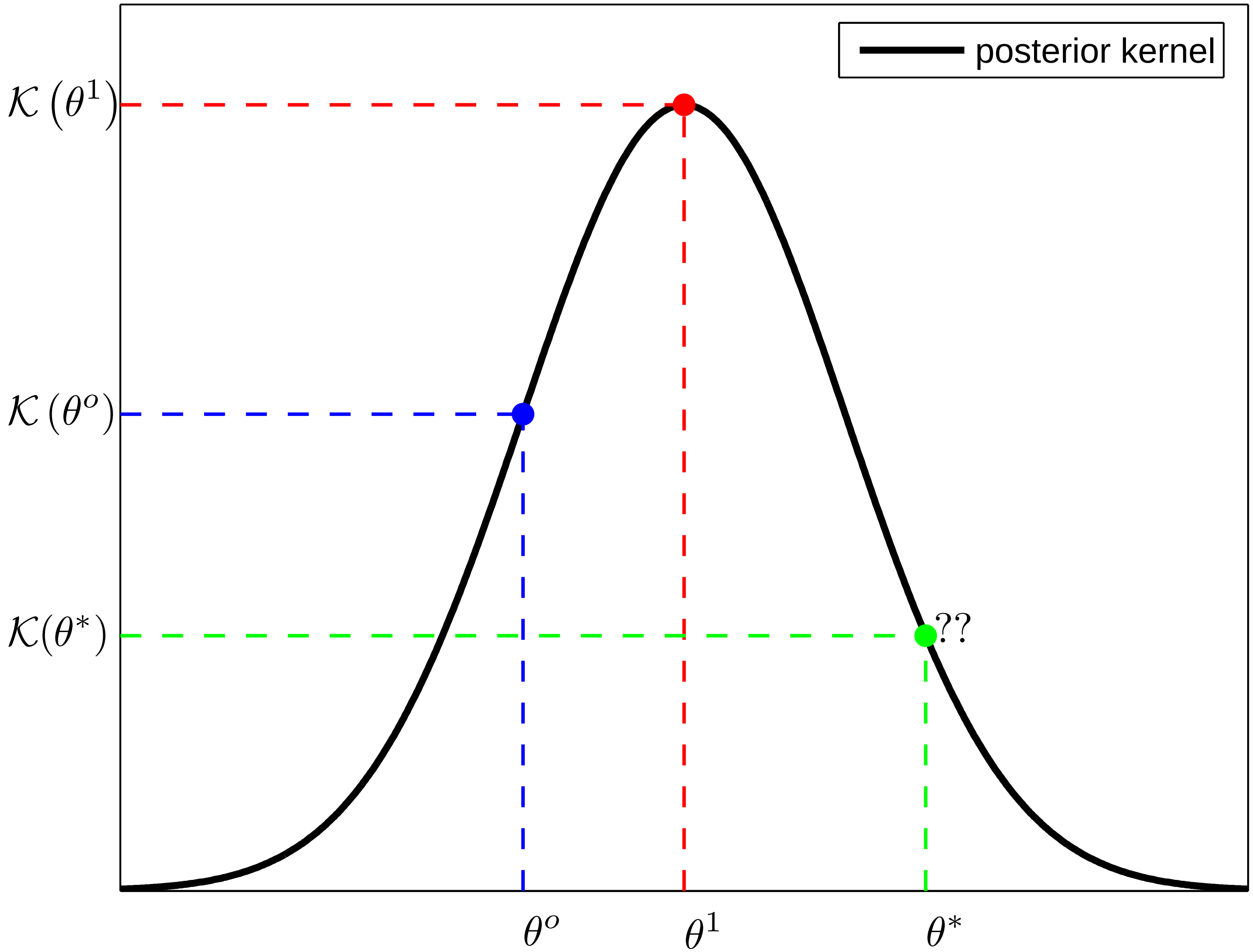
$$r = \frac{p(\boldsymbol{\theta}^*|\mathbf{Y}_T)}{p(\boldsymbol{\theta}^{t-1}|\mathbf{Y}_T)} = \frac{\mathcal{K}(\boldsymbol{\theta}^*|\mathbf{Y}_T)}{\mathcal{K}(\boldsymbol{\theta}^{t-1}|\mathbf{Y}_T)}$$

4. Finally

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


posterior kernel
$\mathcal{K}(\theta^o)$
θ^o


posterior kernel
$\mathcal{K}\left(\theta^1\right) = \mathcal{K}\left(\theta^*\right)$
$\mathcal{K}\left(\theta^o\right)$
θ^o
$\theta^1 = \theta^*$


posterior kernel
$\mathcal{K}(\theta^1)$
$\mathcal{K}(\theta^o)$
$\mathcal{K}(\theta^*)$
??
θ^o
θ^1
θ^*

Metropolis algorithm (II)

- 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**.

Approximation of the marginal density

- By definition we have:

$$p(\mathbf{Y}_T|\mathcal{A}) = \int_{\Theta_{\mathcal{A}}} p(\boldsymbol{\theta}_{\mathcal{A}}|\mathbf{Y}_T, \mathcal{A})p(\boldsymbol{\theta}_{\mathcal{A}}|\mathcal{A})d\boldsymbol{\theta}_{\mathcal{A}}$$

- By assuming that the posterior density is gaussian (Laplace approximation) we have the following estimator:

$$\widehat{p}(\mathbf{Y}_T|\mathcal{A}) = (2\pi)^{\frac{k}{2}}|\Sigma_{\boldsymbol{\theta}^m_{\mathcal{A}}}|^{-\frac{1}{2}}p(\boldsymbol{\theta}^m_{\mathcal{A}}|\mathbf{Y}_T, \mathcal{A})p(\boldsymbol{\theta}^m_{\mathcal{A}}|\mathcal{A})$$

where $\boldsymbol{\theta}^m_{\mathcal{A}}$ is the posterior mode.

Harmonic mean (I)

- Note that

$$\mathbb{E}\left[\frac{f(\boldsymbol{\theta}_{\mathcal{A}})}{p(\boldsymbol{\theta}_{\mathcal{A}}|\mathcal{A})p(\mathbf{Y}_T|\boldsymbol{\theta}_{\mathcal{A}},\mathcal{A})}\middle|\boldsymbol{\theta}_{\mathcal{A}},\mathcal{A}\right] = \int_{\Theta_{\mathcal{A}}} \frac{f(\boldsymbol{\theta}_{\mathcal{A}})p(\boldsymbol{\theta}_{\mathcal{A}}|\mathbf{Y}_T,\mathcal{A})}{p(\boldsymbol{\theta}_{\mathcal{A}}|\mathcal{A})p(\mathbf{Y}_T|\boldsymbol{\theta}_{\mathcal{A}},\mathcal{A})} d\boldsymbol{\theta}_{\mathcal{A}}$$

  where f is a density function.

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

$$\int_{\Theta_{\mathcal{A}}} \frac{f(\boldsymbol{\theta}_{\mathcal{A}})}{p(\boldsymbol{\theta}_{\mathcal{A}}|\mathcal{A})p(\mathbf{Y}_T|\boldsymbol{\theta}_{\mathcal{A}},\mathcal{A})} \frac{p(\boldsymbol{\theta}_{\mathcal{A}}|\mathcal{A})p(\mathbf{Y}_T|\boldsymbol{\theta}_{\mathcal{A}},\mathcal{A})}{\int_{\Theta_{\mathcal{A}}} p(\boldsymbol{\theta}_{\mathcal{A}}|\mathcal{A})p(\mathbf{Y}_T|\boldsymbol{\theta}_{\mathcal{A}},\mathcal{A})d\boldsymbol{\theta}_{\mathcal{A}}} d\boldsymbol{\theta}_{\mathcal{A}}$$

- Finally, we have

$$\mathbb{E}\left[\frac{f(\boldsymbol{\theta}_{\mathcal{A}})}{p(\boldsymbol{\theta}_{\mathcal{A}}|\mathcal{A})p(\mathbf{Y}_T|\boldsymbol{\theta}_{\mathcal{A}},\mathcal{A})}\middle|\boldsymbol{\theta}_{\mathcal{A}},\mathcal{A}\right] = \frac{\int_{\Theta_{\mathcal{A}}} f(\boldsymbol{\theta})d\boldsymbol{\theta}}{\int_{\Theta_{\mathcal{A}}} p(\boldsymbol{\theta}_{\mathcal{A}}|A)p(\mathbf{Y}_T|\boldsymbol{\theta}_{\mathcal{A}},\mathcal{A})d\boldsymbol{\theta}_{\mathcal{A}}}$$

Harmonic mean (II)

- So that

$$p(\mathbf{Y}_T|\mathcal{A}) = \mathbb{E}\left[\frac{f(\boldsymbol{\theta}_\mathcal{A})}{p(\boldsymbol{\theta}_\mathcal{A}|\mathcal{A})p(\mathbf{Y}_T|\boldsymbol{\theta}_\mathcal{A},\mathcal{A})}\Big|\boldsymbol{\theta}_\mathcal{A},\mathcal{A}\right]^{-1}$$

- This suggests the following estimator of the marginal density

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

- Each drawn vector $\boldsymbol{\theta}_\mathcal{A}^{(b)}$ comes from the Metropolis-Hastings monte-carlo.

Harmonic mean (III)

- 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).

DSGE model (I)

- Any DSGE model may be represented as follows

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

- with:
  - $\mathbb{E}_t \left[\epsilon_{t+1} \right] = 0$
  - $\mathbb{E}_t \left[\epsilon_{t+1} \epsilon'_{t+1} \right] = \Sigma_\epsilon$
- where:
  - y is the vector of endogenous variables,
  - ϵ is the vector of exogenous stochastic shocks,
  - θ is the vector of deep parameters.

DSGE model (II, solution)

- We linearize the model around a deterministic steady state $\bar{y}(\theta)$ (such that $f_\theta(\bar{y}, \bar{y}, \bar{y}, 0) = 0$).
- Solving the linearized version of this model (with dynare for instance), we get a state space reduced form representation:

$$
\begin{aligned}
y_t^* &= M\bar{y}(\theta) + M\hat{y}_t + \eta_t \\
\hat{y}_t &= g_y(\theta)\hat{y}_{t-1} + g_u(\theta)\epsilon_t \\
\mathbb{E}\left[\epsilon_t \epsilon_t'\right] &= \Sigma_\epsilon \\
\mathbb{E}\left[\eta_t \eta_t'\right] &= V
\end{aligned}
$$

- The reduced form is non linear in the deep parameters...

DSGE model (III, Kalman filter)

- ... But linear in the endogenous and exogenous variables. So that the likelihood may be evaluated with a linear prediction error algorithm.
- Kalman Filter recursion, for $t = 1, \dots, T$:

$$\begin{aligned}
v_t &= y_t^* - \bar{y}^*(\theta) - M\hat{y}_t \\
F_t &= MP_tM' + V \\
K_t &= g_y(\theta)P_tM'F_t^{-1} \\
\hat{y}_{t+1} &= g_y(\theta)\hat{y}_t + K_tv_t \\
P_{t+1} &= g_y(\theta)P_t(g_y(\theta) - K_tM)' + g_u(\theta)\Sigma_\epsilon g_u(\theta)'
\end{aligned}$$

with initial conditions y_1 and P_1.

DSGE model (IV, likelihood & posterior)

- We have to estimate θ, Σ_ϵ and V
- The log-likelihood is defined by :

$$\log \mathcal{L}(\boldsymbol{\theta}|\mathbf{Y}_T^*) = \frac{Tk}{2}\log 2\pi - \frac{1}{2}\sum_{t=1}^{T}\log|F_t| - \frac{1}{2}\sum_{t=1}^{T} v_t' F_t^{-1} v_t$$

where the vector $\boldsymbol{\theta}$ contains the k parameters to be estimated (θ, $vech\,(\Sigma_\epsilon)$ and $vech\,(V)$).

- The log posterior kernel is then defined by :

$$\log \mathcal{K}(\boldsymbol{\theta}|\mathbf{Y}_T^*) = \log \mathcal{L}(\boldsymbol{\theta}|\mathbf{Y}_T^*) + \log p(\boldsymbol{\theta})$$

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) :
  - $\Delta w_t - \alpha \Delta p_{t-1} =$
    $\beta \mathbb{E}_t\left[\Delta w_{t+1}\right] - \alpha\beta\Delta p_t + \kappa_w\left[mrs_t - (w_t - p_t)\right]$

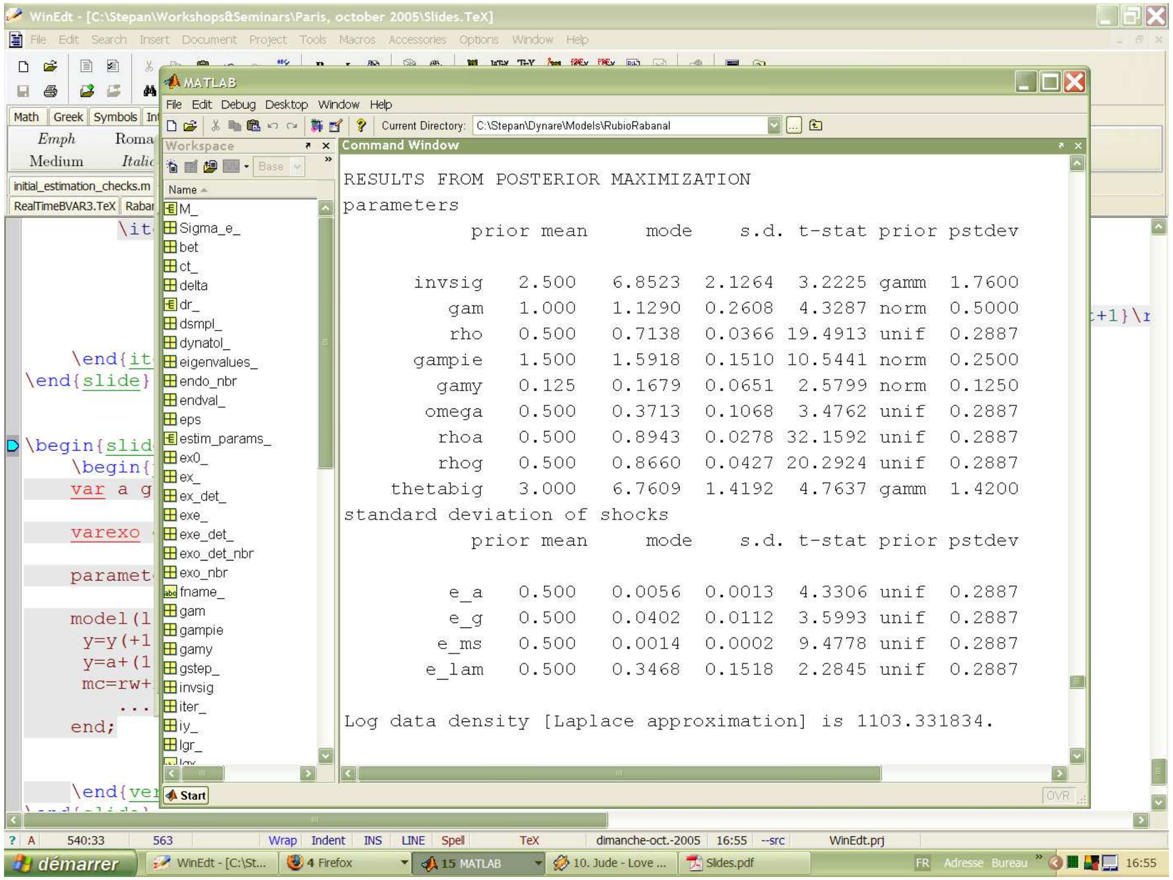
DYNARE (I)

```
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;
```

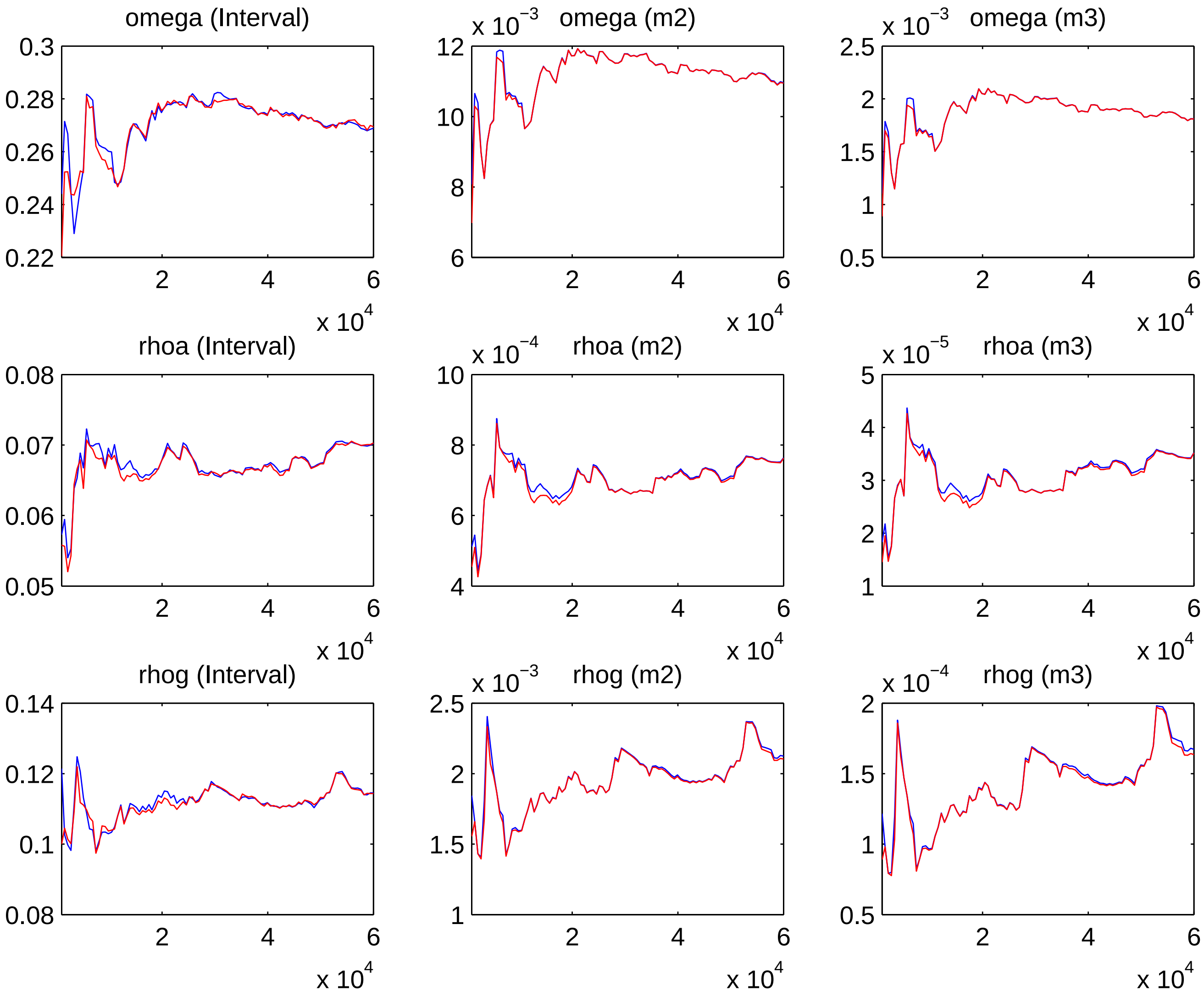
DYNARE (II)

```
estimated_params;
  stderr e_a, uniform_pdf,,,0,1;
  stderr e_lam, uniform_pdf,,,0,1;
    ....
  rho,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);
```

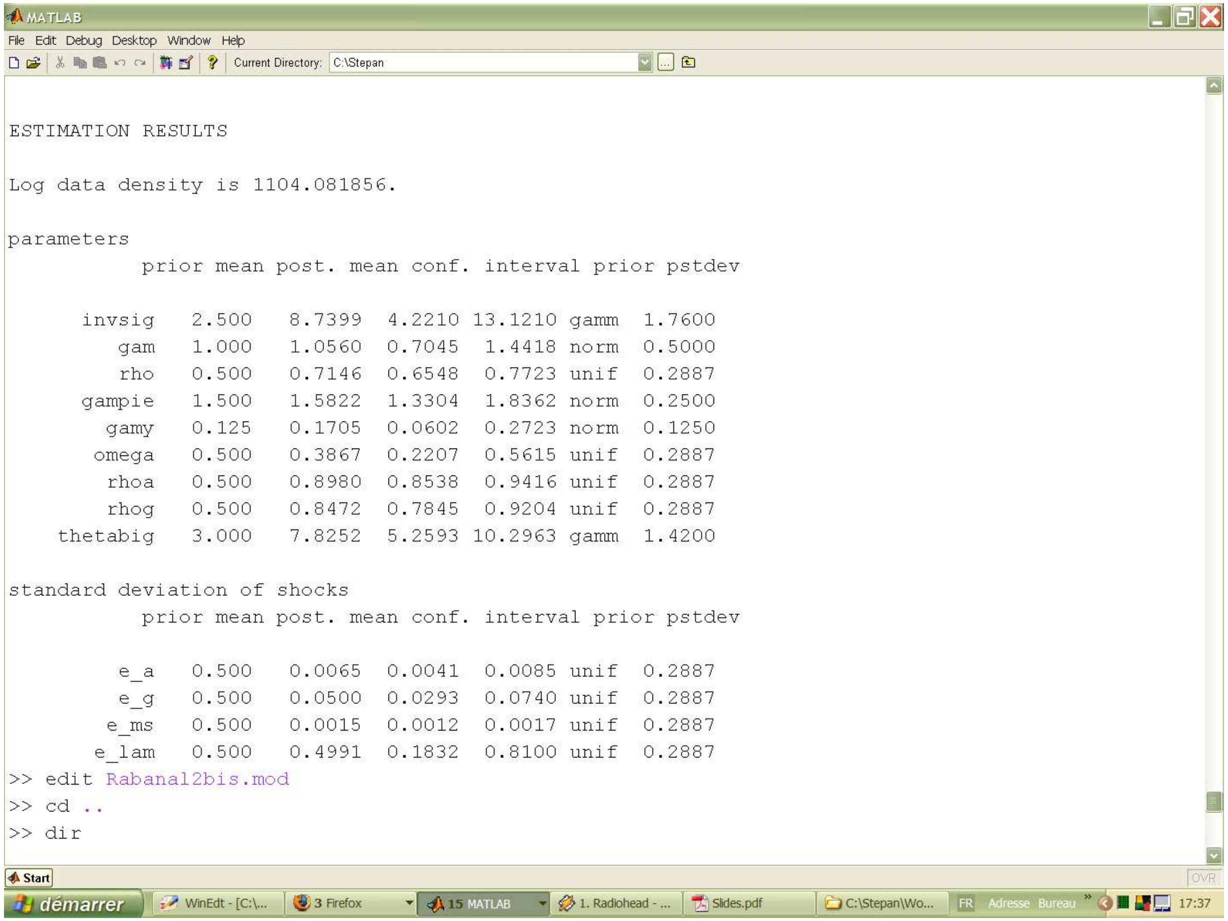
Posterior mode



Posterior mode



Posterior distributions (I)



ESTIMATION RESULTS

Log data density is 1104.081856.

parameters

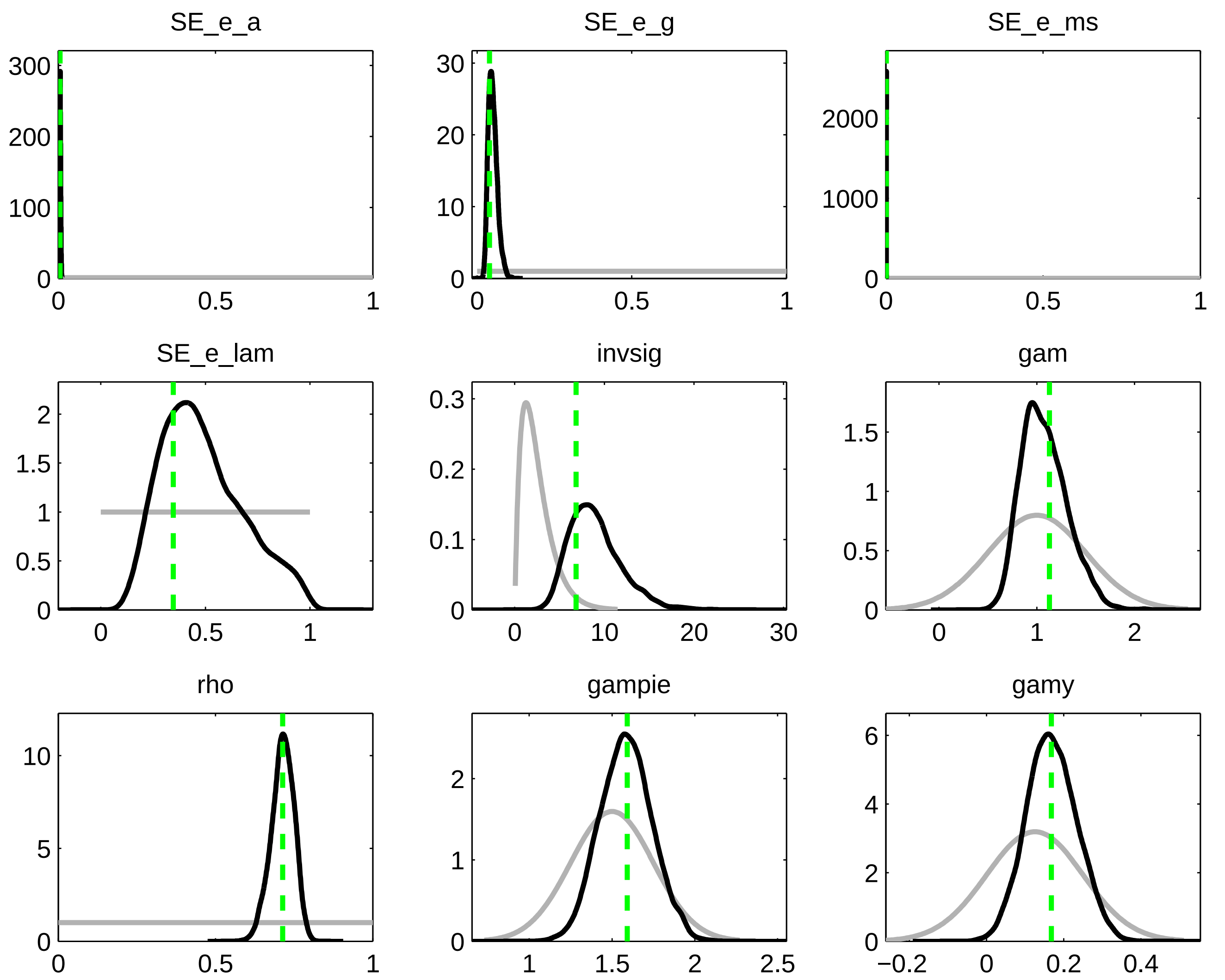
	prior mean	post. mean	conf. interval		prior	pstdev
invsig	2.500	8.7399	4.2210	13.1210	gamm	1.7600
gam	1.000	1.0560	0.7045	1.4418	norm	0.5000
rho	0.500	0.7146	0.6548	0.7723	unif	0.2887
gampie	1.500	1.5822	1.3304	1.8362	norm	0.2500
gamy	0.125	0.1705	0.0602	0.2723	norm	0.1250
omega	0.500	0.3867	0.2207	0.5615	unif	0.2887
rhoa	0.500	0.8980	0.8538	0.9416	unif	0.2887
rhog	0.500	0.8472	0.7845	0.9204	unif	0.2887
thetabig	3.000	7.8252	5.2593	10.2963	gamm	1.4200

standard deviation of shocks

	prior mean	post. mean	conf. interval		prior	pstdev
e_a	0.500	0.0065	0.0041	0.0085	unif	0.2887
e_g	0.500	0.0500	0.0293	0.0740	unif	0.2887
e_ms	0.500	0.0015	0.0012	0.0017	unif	0.2887
e_lam	0.500	0.4991	0.1832	0.8100	unif	0.2887

```
>> edit Rabanal2bis.mod
>> cd ..
>> dir
```

Posterior distributions (II)



Posterior distributions (III)