Estimation of DSGE models

Stéphane Adjemian

Université du Maine, GAINS & CEPREMAP

stephane.adjemian@ens.fr

July 2, 2007

- Our model is given by:

$$\mathbb{E}_t \left[\mathcal{F}_\theta (y_{t+1}, y_t, y_{t-1}, \varepsilon_t) \right] = 0 \tag{1}$$

  with $\varepsilon_t \sim \text{iid}\,(0, \Sigma)$ 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.

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

$$y_t = \mathcal{H}_\theta \left(y_{t-1}, \varepsilon_t \right) \tag{2}$$

  The endogenous variables are written as a function of their past levels and the contemporaneous structural shocks. $\mathcal{H}_\theta$ 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 (2).

- Dynare uses a local approximation around the deterministic steady state. Global approximations are not yet implemented in DYNARE.

- Substituting (2) in (1) for y_t and y_{t+1} we obtain:

$$\mathbb{E}_t \left[\mathcal{F}_\theta \left(\left(\mathcal{H}_\theta \left(y_t, \varepsilon_{t+1} \right), \mathcal{H}_\theta \left(y_{t-1}, \varepsilon_t \right), y_{t-1}, \varepsilon_t \right) \right] = 0$$

- Substituting again (for y_t in y_{t+1}) and dropping time we get:

$$\mathbb{E}_t \left[\mathcal{F}_\theta \left(\left(\mathcal{H}_\theta \left(\mathcal{H}_\theta \left(y, \varepsilon \right), \varepsilon' \right), \mathcal{H}_\theta \left(y, \varepsilon \right), y, \varepsilon \right) \right] = 0 \qquad (3)$$

  where y and ε are in the time t information set, but not ε' which is assumed to be iid $(0, \Sigma)$. $\mathcal{F}_\theta$ is known and $\mathcal{H}_\theta$ is the unknown. We are looking for a function $\mathcal{H}_\theta$ satisfying this equation for all possible states (y, ε)...

- This task is far easier if we "solve" only locally (around the deterministic steady state) this functional equation.

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

$$\mathcal{F}_\theta\left(y^*(\theta), y^*(\theta), y^*(\theta), 0\right) = 0$$

- 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 $\mathcal{H}_\theta$ must satisfy the following equality:

$$y^* = \mathcal{H}_\theta\left(y^*, 0\right)$$

- Once the steady state is known, we can compute the jacobian matrix associated to $\mathcal{F}_\theta$...

- Let $\hat{y} = y_{t-1} - y^*$, $\mathcal{F}_{y_+} = \frac{\partial \mathcal{F}_\theta}{\partial y_{t+1}}$, $\mathcal{F}_y = \frac{\partial \mathcal{F}_\theta}{\partial y_t}$, $\mathcal{F}_{y_-} = \frac{\partial \mathcal{F}_\theta}{\partial y_{t-1}}$, $\mathcal{F}_\varepsilon = \frac{\partial \mathcal{F}_\theta}{\partial \varepsilon_t}$, $\mathcal{H}_y = \frac{\partial \mathcal{H}_\theta}{\partial y_{t-1}}$ and $\mathcal{H}_\varepsilon = \frac{\partial \mathcal{H}_\theta}{\partial \varepsilon_t}$.
- $\mathcal{F}_{y_+}$, $\mathcal{F}_y$, $\mathcal{F}_{y_-}$, $\mathcal{F}_\varepsilon$ are known and $\mathcal{H}_y$, $\mathcal{H}_\varepsilon$ are the unknowns.
- With a first order Taylor expansion of the functional equation (3) around y^*:

$$0 \simeq \mathcal{F}_\theta(y^*, y^*, y^*, 0) + \mathcal{F}_{y_+}\left(\mathcal{H}_y\left(\mathcal{H}_y\hat{y} + \mathcal{H}_\varepsilon\varepsilon\right) + \mathcal{H}_\varepsilon\varepsilon'\right)$$
$$+ \mathcal{F}_y\left(\mathcal{H}_y\hat{y} + \mathcal{H}_\varepsilon\varepsilon\right) + \mathcal{F}_{y_-}\hat{y} + \mathcal{F}_\varepsilon\varepsilon$$

  Where all the derivatives are evaluated at the deterministic steady state and $\mathcal{F}_\theta(y^*, y^*, y^*, 0) = 0$.

- Applying the conditional expectation operator, we obtain:

$$\begin{aligned} 0 \simeq & \mathcal{F}_{y_+}\left(\mathcal{H}_y\left(\mathcal{H}_y\hat{y}+\mathcal{H}_\varepsilon\varepsilon\right)\right) \\ & +\mathcal{F}_y\left(\mathcal{H}_y\hat{y}+\mathcal{H}_\varepsilon\varepsilon\right)+\mathcal{F}_{y_-}\hat{y}+\mathcal{F}_\varepsilon\varepsilon \end{aligned}$$

  or equivalently:

$$\begin{aligned} 0 \simeq & \mathcal{F}_{y_+}\mathcal{H}_y\mathcal{H}_y\hat{y}+\mathcal{F}_y\mathcal{H}_y\hat{y}+\mathcal{F}_{y_-}\hat{y} \\ & \mathcal{F}_{y_+}\mathcal{H}_y\mathcal{H}_\varepsilon\varepsilon+\mathcal{F}_y\mathcal{H}_\varepsilon\varepsilon+\mathcal{F}_\varepsilon\varepsilon \end{aligned}$$

- This equation must hold for any state $(\hat{y},\varepsilon)$, so that the unknowns $\mathcal{H}_y$ and $\mathcal{H}_\varepsilon$ must satisfy:

$$\begin{cases} 0 & = \mathcal{F}_{y_+}\mathcal{H}_y\mathcal{H}_y+\mathcal{F}_yH_y+\mathcal{F}_{y_-} \\ 0 & = \mathcal{F}_{y_+}\mathcal{H}_y\mathcal{H}_\varepsilon+\mathcal{F}_y\mathcal{H}_\varepsilon+\mathcal{F}_\varepsilon \end{cases}$$

- This system is triangular ($\mathcal{H}_\varepsilon$ does not appear in the first equation) $\Rightarrow$ "easy" to solve.
- The first equation is a quadratic equation... But the unknown is a squared matrix ($\mathcal{H}_y$). This equation may be solved with any spectral method. DYNARE uses a generalized Schur decomposition. A unique solution exists iff BK conditions are satisfied.
- The second equation is linear in the unknown $\mathcal{H}_\varepsilon$, a unique solution exists iff

$$\mathcal{F}_{y_+}\mathcal{H}_y + \mathcal{F}_y$$

is an inversible matrix ($\not\hookrightarrow$ if $\mathcal{F}_y$ and F_{y_+} are diagonal matrices, each endogenous variable have to appear at time t **or** with a lead).

- Finally the local dynamic is given by:

$$y_t = y^* + \mathcal{H}_y(\theta)\left(y_{t-1} - y^*\right) + \mathcal{H}_\varepsilon(\theta)\varepsilon_t$$

  where y^*, $\mathcal{H}_y(\theta)$ and $\mathcal{H}_\varepsilon(\theta)$ are nonlinear functions of the deep parameters.

- This result can be used to approximate the theoretical moments:

$$\mathbb{E}_\infty[y_t] = y^*(\theta)$$
$$\mathbb{V}_\infty[y_t] = \mathcal{H}_y(\theta)\mathbb{V}_\infty[y_t]\mathcal{H}_y(\theta)' + \mathcal{H}_\varepsilon(\theta)\Sigma\mathcal{H}_\varepsilon(\theta)'$$

  The second equation is a kind of sylvester equation and may be solved using the vec operator and kronecker product.

- This result can also be used to approximate the likelihood.

- A direct estimation approach is to maximize the likelihood with respect to θ and vech (Σ).
- 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:

$$y_t^\star = Z y_t + \eta_t \tag{4a}$$

$$y_t = \mathcal{H}_\theta \left(y_{t-1}, \varepsilon_t \right) \tag{4b}$$

Equation (4b) is the reduced form of the DSGE model $\Rightarrow$ *state equation*. Equation (4a) 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*.

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

$$\mathcal{L}(\psi; \mathcal{Y}_T^\star) = p\left(\mathcal{Y}_T^\star|\psi\right) = p\left(y_0^\star|\psi\right) \prod_{t=1}^{T} p\left(y_t^\star|\mathcal{Y}_{t-1}^\star, \psi\right) \qquad (5)$$

- 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 $p\left(y_t^\star|\mathcal{Y}_{t-1}^\star, \psi\right)$.

- The state-space model (4), or the reduced form (2), 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 $\Rightarrow$ **example with an AR(1)**).
- 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:

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

- Unit roots (diffuse kalman filter).

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

$$p\left(y_t^\star|\mathcal{Y}_{t-1}^\star,\psi\right)=\int_\Lambda p\left(y_t^\star|y_t,\psi\right)p(y_t|\mathcal{Y}_{t-1}^\star,\psi)\mathrm{d}y_t \qquad (6)$$

  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 (4a).
- 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*].

- The Kalman filter can be seen as a bayesian recursive estimation device:

$$p\left(y_t|\mathcal{Y}_{t-1}^{\star},\psi\right)=\int_{\Lambda}p\left(y_t|y_{t-1},\psi\right)p\left(y_{t-1}|\mathcal{Y}_{t-1}^{\star},\psi\right)\mathrm{d}y_{t-1} \tag{7a}$$

$$p\left(y_t|\mathcal{Y}_{t}^{\star},\psi\right)=\frac{p\left(y_t^{\star}|y_t,\psi\right)p\left(y_t|\mathcal{Y}_{t-1}^{\star},\psi\right)}{\int_{\Lambda}p\left(y_t^{\star}|y_t,\psi\right)p\left(y_t|\mathcal{Y}_{t-1}^{\star},\psi\right)\mathrm{d}y_t} \tag{7b}$$

- Equation (7a) 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 (4b), weigthed by the density y_{t-1} conditional on $\mathcal{Y}_{t-1}^{\star}$ (given by (7b)).
- The update equation (7b) is a direct application of the Bayes theorem and tells us how to update our knowledge about the latent variables when new information (data) is available.

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

- $p\left(y_t|\mathcal{Y}_{t-1}^\star,\psi\right)$ 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 (4a)) $\Rightarrow$ 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.

- 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.

The linear–gaussian Kalman filter recursion is given by:

$$\begin{aligned} v_t &= y_t^\star - y(\theta)^* - Z\hat{y}_t \\ F_t &= ZP_tZ' + \mathbb{V}[\eta] \\ K_t &= \mathcal{H}_y(\theta)P_t\mathcal{H}_y(\theta)'F_t^{-1} \\ \hat{y}_{t+1} &= \mathcal{H}_y(\theta)\hat{y}_t + K_t v_t \\ P_{t+1} &= \mathcal{H}_y(\theta)P_t(\mathcal{H}_y(\theta) - K_t Z)' + \mathcal{H}_\varepsilon(\theta)\Sigma H_\varepsilon(\theta)' \end{aligned}$$

for $t = 1, \ldots, T$, with $\hat{y}_0$ and P_0 given.

Finally the (log)-likelihood is:

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

- We generally do not have an analytical expression for the likelihood, but a numerical evaluation is possible.
- Experience shows that it is quite difficult to estimate a model by maximum likelihood.
- The main reason is that data are not informative enough... The likelihood is flat in some directions (identification problems).
- This suggests that (when possible) we should use other sources information ⇒ Bayesian approach.
- A second (practical) motivation for the bayesian estimation is that (DSGE) models are mispecified.

- When a misspecified model is estimated (for instance an RBC model) 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 third motivation is related to the precision of the ML estimator. Using informative priors we reduce the posterior uncertainty (variance).

- Let the prior density be $p_0(\psi)$.
- The posterior density is given by (Bayes theorem):

$$p_1\left(\psi|\mathcal{Y}_T^\star\right) = \frac{p_0\left(\psi\right) p(\mathcal{Y}_T^\star|\psi)}{p(\mathcal{Y}_T^\star)} \tag{8}$$

  where

$$p\left(\mathcal{Y}_T^\star\right) = \int_\Psi p_0\left(\psi\right) p(\mathcal{Y}_T^\star|\psi)\mathrm{d}\psi \tag{9}$$

  is the marginal density of the sample (model comparison).
- The posterior density is proportional to the product of the prior density and the likelihood.

$$p_1\left(\psi|\mathcal{Y}_T^\star\right) \propto p_0\left(\psi\right) p(\mathcal{Y}_T^\star|\psi)$$

- The prior affects the shape of the likelihood!...

- Data Generating Process

$$y_t = \mu + \varepsilon_t$$

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

- Let $\mathcal{Y}_T \equiv (y_1, \ldots, y_T)$. 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:

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

- 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.

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

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

- 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:

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

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

- 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:

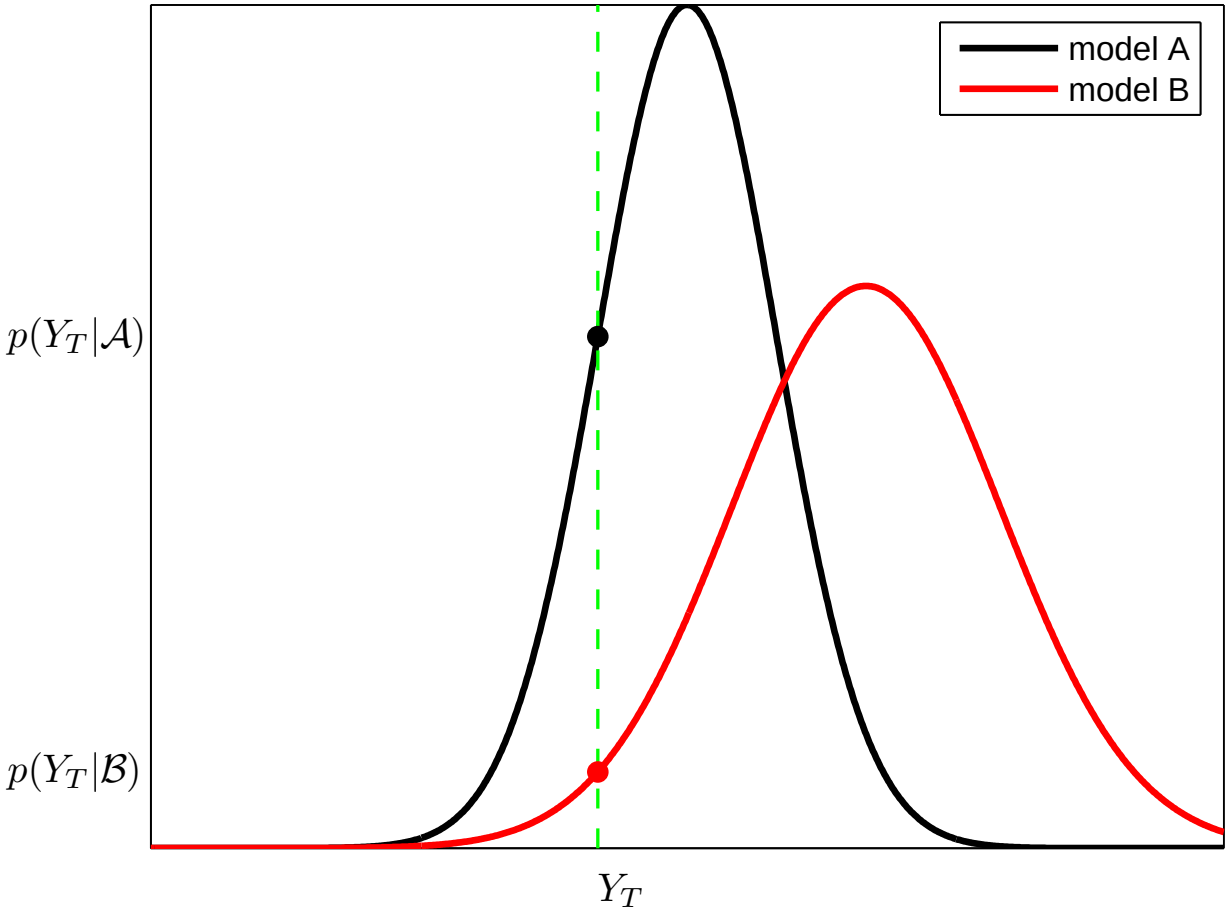
$$\mathbb{V}\left[\mu\right] = \frac{1}{T+\frac{1}{\sigma_\mu^2}}$$

- The posterior mean is a convex combination of the prior mean and the ML estimate.
  - If $\sigma_\mu^2 \to \infty$ (no prior information) then $\mathbb{E}[\mu] \to \widehat{\mu}$ (ML).
  - If $\sigma_\mu^2 \to 0$ (calibration) then $\mathbb{E}[\mu] \to \mu_0$.
- 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.

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

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

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


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

- 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:

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

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

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

- 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.

- Estimation of the Phillips curve :

$$\pi_t = \beta \mathbb{E}\pi_{t+1} + \frac{(1-\xi_p)(1-\beta\xi_p)}{\xi_p}\Big((\sigma_c + \sigma_l)y_t + \tau_t\Big)$$

- ξ_p is the (Calvo) probability (for an intermediate firm) of being able to optimally choose its price at time t. With probability $1-\xi_p$ the price is indexed on past inflation an/or steady state inflation.
- Let $\alpha_p \equiv \frac{1}{1-\xi_p}$ be the expected period length during which a firm will not optimally adjust its price.
- Let $\lambda = \frac{(1-\xi_p)(1-\beta\xi_p)}{\xi_p}$ be the slope of the Phillips curve.
- Suppose that β, σ_c and σ_l are known.

- The prior may be defined on ξ_p, α_p or the slope λ.
- Say we choose a uniform prior for the Calvo probability:

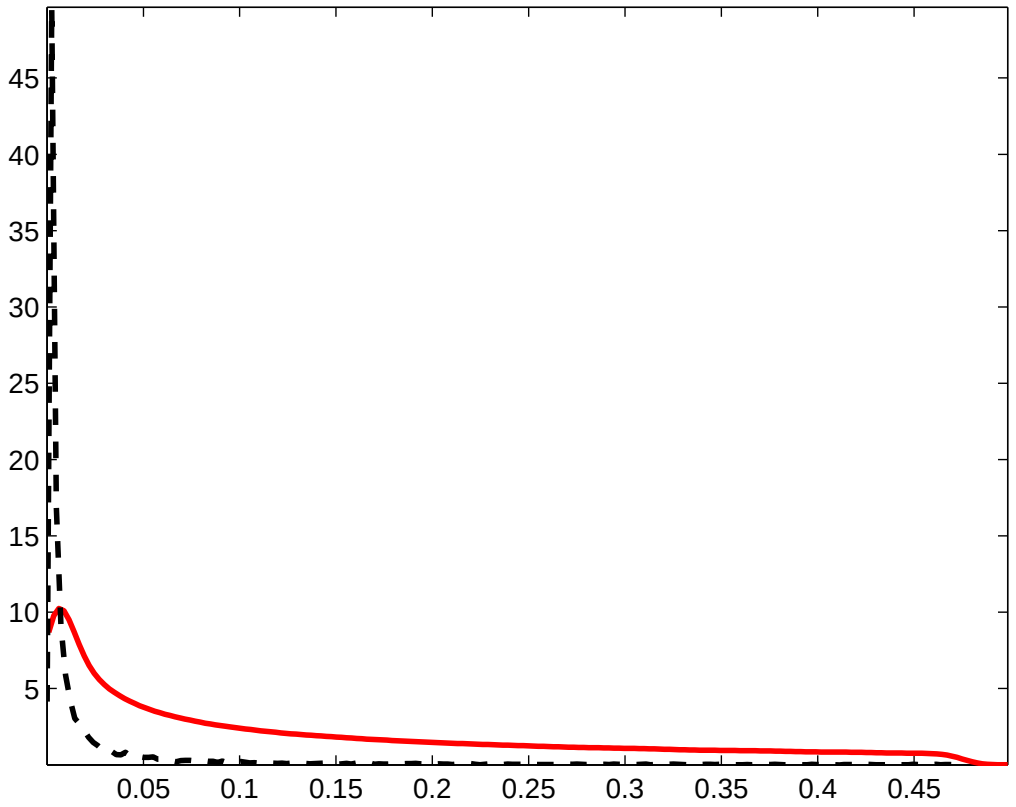
$$\xi_p \sim \mathcal{U}_{[.51,.99]}$$

  The prior mean is .75 (so that the implied value for α_p is 4 quarters). This prior is often think as a non informative prior...

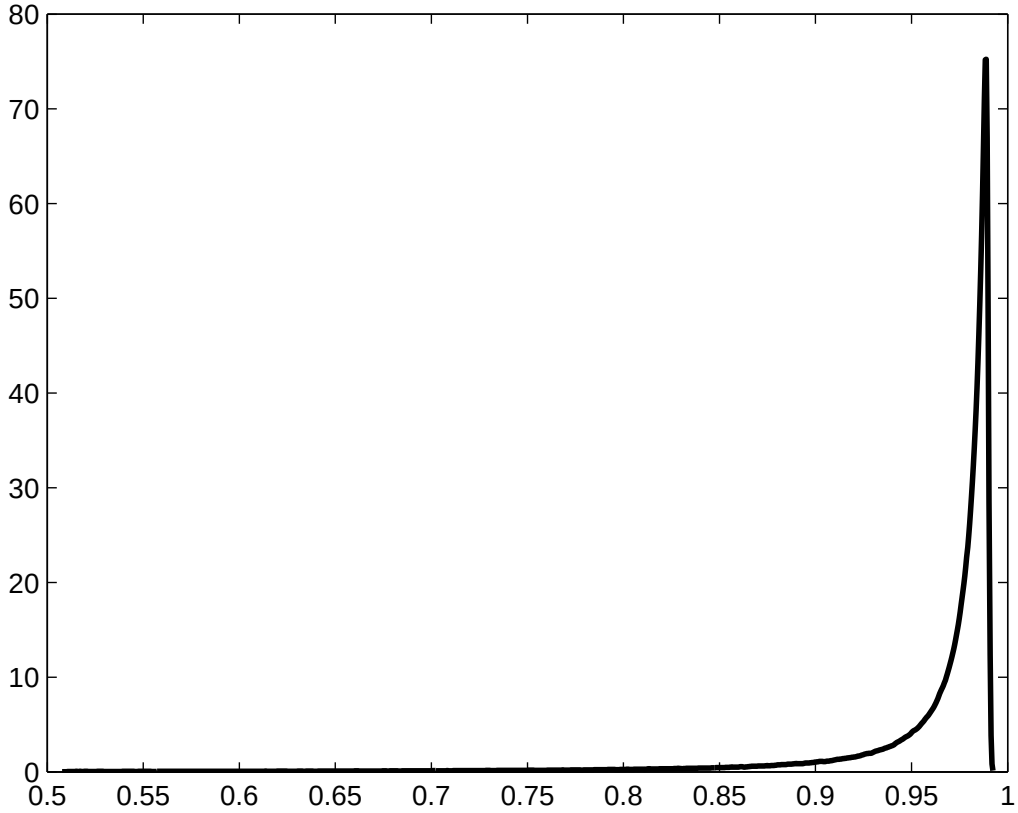
- An alternative would be to choose a uniform prior for α_p:

$$\alpha_p \sim \mathcal{U}_{\left[1-\frac{1}{.51},1-\frac{1}{.99}\right]}$$

- **These two priors are very different!**



↬ The prior on α_p is much more informative than the prior on ξ_p.



Implied prior density of $\xi_p = 1 - \frac{1}{\alpha_p}$ if the prior density of α_p is uniform.

- Robustness of the results may be evaluated by considering a more general prior density.
- For instance, in our simple example we could assume a student prior density for μ instead of a gaussian density.

- If a parameter, say μ, can take values between $-\infty$ and ∞, the flat prior is a uniform density between $-\infty$ and ∞.
- If a parameter, say σ, can take values between 0 and ∞, the flat prior is a uniform density between $-\infty$ and ∞ for $\log\sigma$:

$$p_0(\log\sigma) \propto 1 \Leftrightarrow p_0(\sigma) \propto \frac{1}{\sigma}$$

- Invariance.
- Why is this prior non informative ?... $\int p_0(\mu)\mathrm{d}\mu$ is not defined! $\Rightarrow$ **Improper prior**.
- Practical implications for DSGE estimation.

- An alternative, proposed by Jeffrey, is to use the Fisher information matrix:

$$p_0(\psi) \propto |I(\psi)|^{\frac{1}{2}}$$

with

$$I(\psi) = \mathbb{E}\left[\left(\frac{\partial p(\mathcal{Y}_T^\star|\psi)}{\partial \psi}\right)\left(\frac{\partial p(\mathcal{Y}_T^\star|\psi)}{\partial \psi}\right)'\right]$$

- The idea is to mimic the information in the data...
- Automatic choice of the prior.
- Invariance to any continuous transformation of the parameters.
- Very different results (compared to the flat prior) $\Rightarrow$ Unit root controverse.

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

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

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.

- 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:

$$p_1(\psi) \doteq (2\pi)^{-\frac{k}{2}} |\mathcal{H}(\psi^*)|^{-\frac{1}{2}} e^{-\frac{1}{2}(\psi-\psi^*)'[\mathcal{H}(\psi^*)]^{-1}(\psi-\psi^*)} \quad (10)$$

- 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.

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

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

  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.

- 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 $\left(\varphi(\psi^{(1)}), \varphi(\psi^{(2)}), \ldots, \varphi(\psi^{(n)})\right)$, 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.

- Imagine we want to obtain some draws from a $\mathcal{N}(0, 4)$ distribution...
- But we are only able to draw from $\mathcal{N}(0, 1)$ and we don't realize that we should simply multiply by 2 the draws from a standard normal distribution.
- The idea is to build a stochastic process whose limiting distribution is $\mathcal{N}(0, 4)$.
- We define the following AR(1) process:

$$x_t = \rho x_{t-1} + \epsilon_t$$

  with $\epsilon_t \sim \mathcal{N}(0, 1)$, $|\rho| < 1$ and $x_0 = 0$.
- We just have to choose ρ such that the asymptotic distribution of $\{x_t\}$ is $\mathcal{N}(0, 4)$.

We have:

- $x_1 = \epsilon_1 \sim \mathcal{N}(0, 1)$
- $x_2 = \rho\epsilon_1 + \epsilon_2 \sim \mathcal{N}\left(0, 1 + \rho^2\right)$
- $x_3 = \rho^2\epsilon_1 + \rho\epsilon_2 + \epsilon_3 \sim \mathcal{N}\left(0, 1 + \rho^2 + \rho^4\right)$
- $x_T = \rho^{T-1}\epsilon_1 + \rho^{T-2}\epsilon_2 + \cdots + \epsilon_T \sim \mathcal{N}\left(0, 1 + \rho^2 + \ldots \rho^{2(T-1)}\right)$
- And

$$x_\infty \sim \mathcal{N}\left(0, \frac{1}{1-\rho^2}\right)$$

So that $\mathbb{V}_\infty[x_t] = 4$ iff $\rho = \pm\frac{\sqrt{3}}{2}$.

- If we simulate enough draws from this gaussian autoregressive stochastic process, we can replicate the targeted distribution.
- In this case it is very simple because we know exactly the targeted distribution **and** we are able to obtain some draws from its standardized version.
- This is far from true with DSGE models. For instance, we even don't have an analytical expression for the posterior distribution.

1. Choose a starting point Ψ^0 & run a loop over 2-3-4.

2. Draw a *proposal* $\Psi^\star$ from a *jumping* distribution

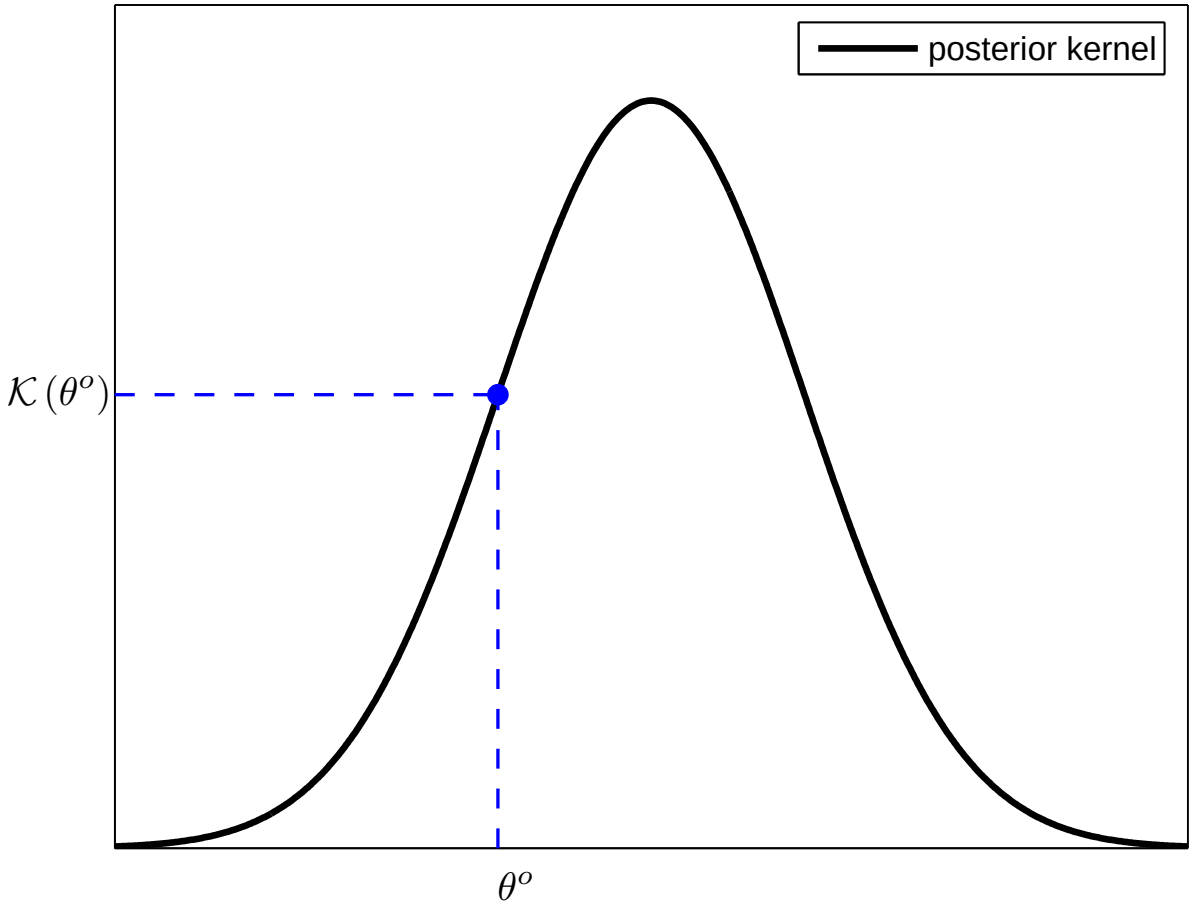
$$J(\Psi^\star|\Psi^{t-1}) = \mathcal{N}(\Psi^{t-1}, c \times \Omega_m)$$

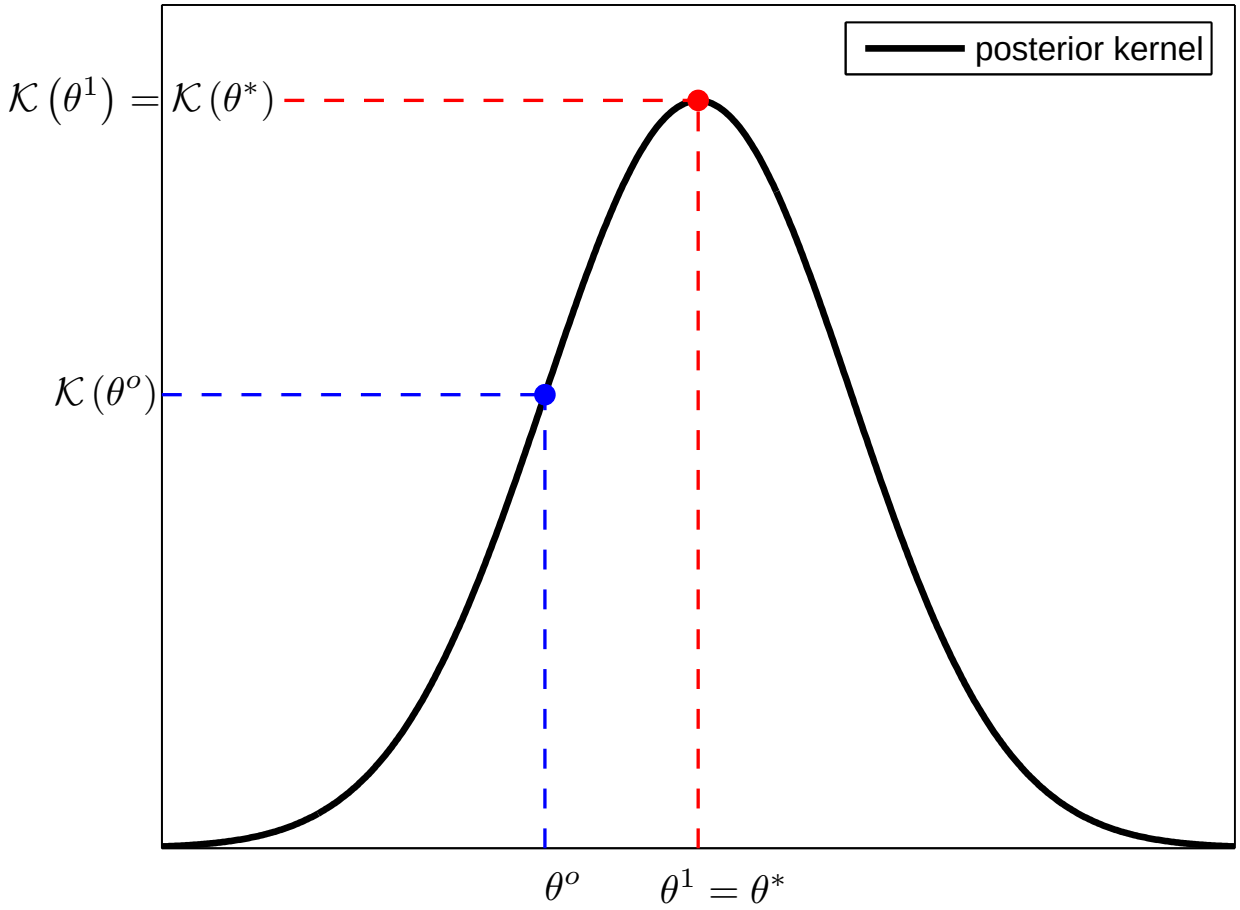
3. Compute the acceptance ratio

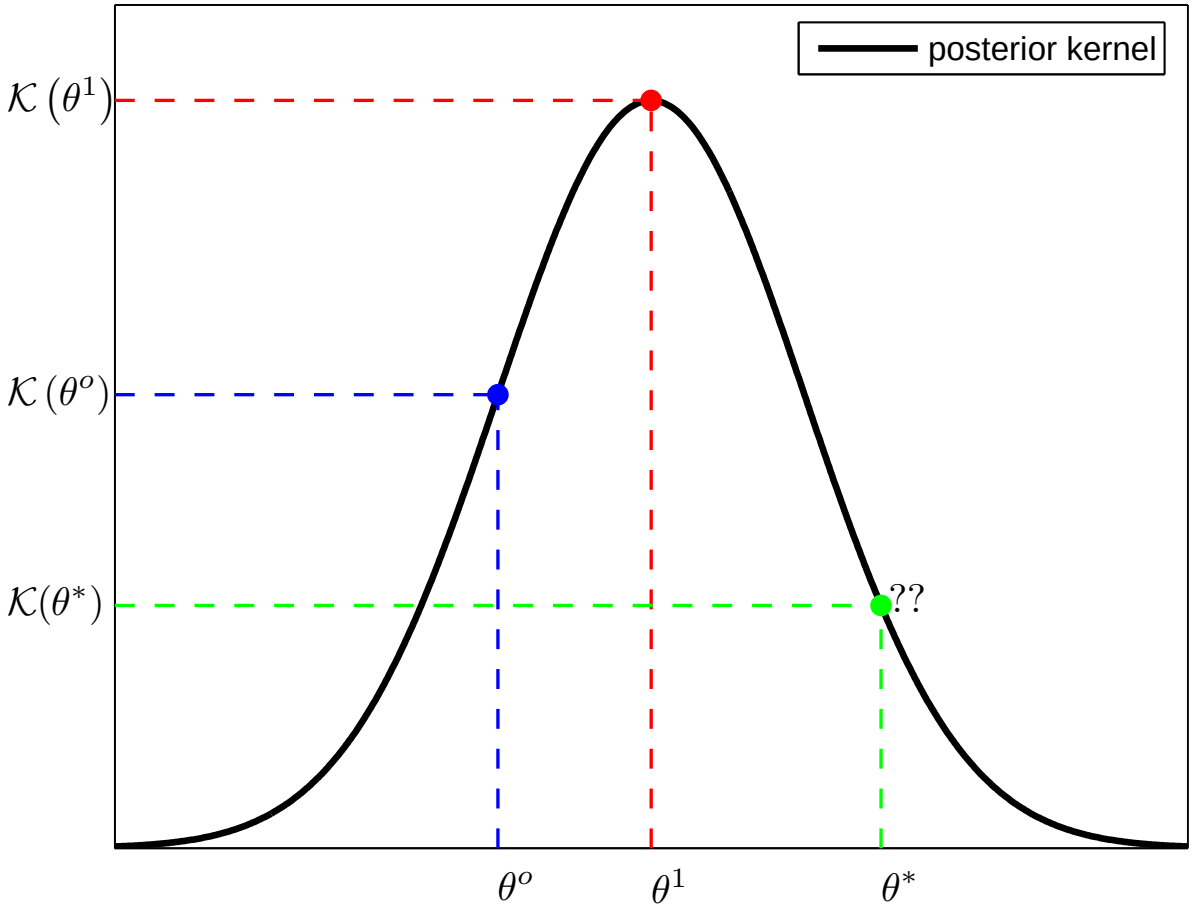
$$r = \frac{p_1(\Psi^\star|\mathcal{Y}_T^\star)}{p(\Psi^{t-1}|\mathcal{Y}_T^\star)} = \frac{\mathcal{K}(\Psi^\star|\mathcal{Y}_T^\star)}{\mathcal{K}(\Psi^{t-1}|\mathcal{Y}_T^\star)}$$

4. Finally

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


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


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


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

- 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**.

- A Markov chain is a sequence of continuous random variables $\left(\psi^{(0)}, \ldots, \psi^{(n)}\right)$, generated by an order one Markov process (*ie* the distribution of $\psi^{(s)}$ depends only on $\psi^{(s-1)}$.
- A Markov chain is defined by a transition kernel that specify the probability to move from $\eta \in \Psi$ to $S \subseteq \Psi$.
- Let $P(\eta, S)$ be the transition kernel. We have $P(\eta, \Psi) = 1$ for all η in Ψ. If the Markov chain defined by the kernel P converge toward an invariant distribution π, then the kernel must also satisfy the following equation:

$$\pi(S) = \int_{\Psi} P(\eta, S)\pi\left(\mathrm{d}\eta\right)$$

for all measurable set S de Ψ.

- Before the ergodic distribution π, if $P^{(s)}(\eta, S)$ denotes the probability that $\psi^{(s)}$ be in S knowing that $\psi^{(s-1)} = \eta$, we have:

$$P^{(s)}(\eta, S) = \int_{\Psi} P(\nu, S) P^{(s-1)}(\eta, \mathrm{d}\nu)$$

  At each iteration the distribution of ψ changes, asymptotically the chain attains the ergodic distribution:

$$\lim_{s \to \infty} P^{(s)}(\eta, S) = \pi(S)$$

- The idea is to choose the transition kernel such that the invariant distribution is the posterior density.

  Let $p(\eta, \nu)$ and $\tilde{\pi}$ be the densities associated to the kernel

P and the invariant distribution π[a].

- Tierney (1994) shows that if the density $p(\eta, \mu)$ satisfy the reversibility condition:

$$\tilde{\pi}(\eta)p(\eta,\nu) = \tilde{\pi}(\nu)p(\nu,\eta)$$

[a]The kernel $P(\eta, S)$ defines the probability to move from η to S. In a favorable case, ψ is in S at the next iteration, two possibilities may be considered : (i) η moves effectively and goes in region S at the next iteration, (ii) η does not move but η is already in region S. The density associated to P is thus a discrete-continuous density, Tierney (1994) adopts the following definition:

$$P(\eta, \mathrm{d}\nu) = p(\eta,\nu)\mathrm{d}\nu + (1 - r(\eta))\delta_\eta(\mathrm{d}\nu)$$

where $p(\eta, \nu) \equiv p(\nu|\eta)$ is the density associated to the transition from η to ν, $r(\eta) = \int p(\eta,\nu)d\nu < 1$, $1 - r(\eta)$ is the probability to stay at the position $\psi = \eta$, $\delta_\eta(S)$ is a dirac fonction equal to one iff $\eta \in S$.

then π is the invariant distribution associated to P

- Equivalently, this condition says that:

$$\frac{\tilde{\pi}(\eta)}{\tilde{\pi}(\nu)} = \frac{p(\nu,\eta)}{p(\eta,\nu)} > 1$$

if the density of $\psi = \eta$, $\tilde{\pi}(\eta)$, dominate the density associated to $\psi = \nu$, $\tilde{\pi}(\nu)$, then it must be "easier" to go from ν to η than from η to ν.

- Say we use $Q(\eta, S)$ as a transition kernel. If we target the posterior distribution it is most likely that the reversibility condition won't be satisfied, *ie*

$$p_1(\eta)q(\eta,\nu) \neq p_1(\nu)q(\nu,\eta)$$

- The Metropolis-Hastings is a general algorithm that corrects the transition kernel so that the reversibility condition holds.
- Suppose that $p_1(\eta)q(\eta,\nu) > p_1(\nu)q(\nu,\eta)$, the MC does not provide enough transitions from $\psi = \nu$ to $\psi = \eta$ so that the reversibility condition is not satisfied.
- The MH algorithm corrects this error by not accepting systematically the jumps proposed by the transition kernel.

1. Choose an initial condition $\psi^{(0)}$ such that $p_1(\psi^{(0)}) > 0$ and let s be equal to 1.
2. Generate a draw (proposition) $\psi^\star$ from $q\left(\psi^{(s-1)}, \psi^\star\right)$.
3. Generate u from a uniform distribution between 0 and 1.
4. Apply the following rule:

$$\psi^{(s)} = \begin{cases} \psi^\star & \text{if } \alpha\left(\psi^{(s-1)}, \psi^\star\right) > u \\ \psi^{(s-1)} & \text{otherwise.} \end{cases}$$

   where the probability of acceptation is:

$$\alpha(\psi^{(s-1)}, \psi^\star) = \min\left\{1, \frac{\mathcal{K}\left(\psi^\star \mid \mathcal{Y}_T^\star\right)}{\mathcal{K}\left(\psi^{(s-1)} \mid \mathcal{Y}_T^\star\right)} \frac{q\left(\psi^{(s-1)} \mid \psi^\star\right)}{q\left(\psi^\star \mid \psi^{(s-1)}\right)}\right\}$$

5. Loop over (2-4) for $s = 2, \ldots, n$

- 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.

- 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.

- 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.

- $\widehat{p}(\mathcal{Y}_T^\star|\mathcal{A}) \underset{B\to\infty}{\longrightarrow} p(\mathcal{Y}_T^\star|\mathcal{A})$.

- 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).

- Note that

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

  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

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

- 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.

- 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).

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

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

- 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

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

- 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

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

- 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.

- 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:

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

- 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.

- 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.

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):

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

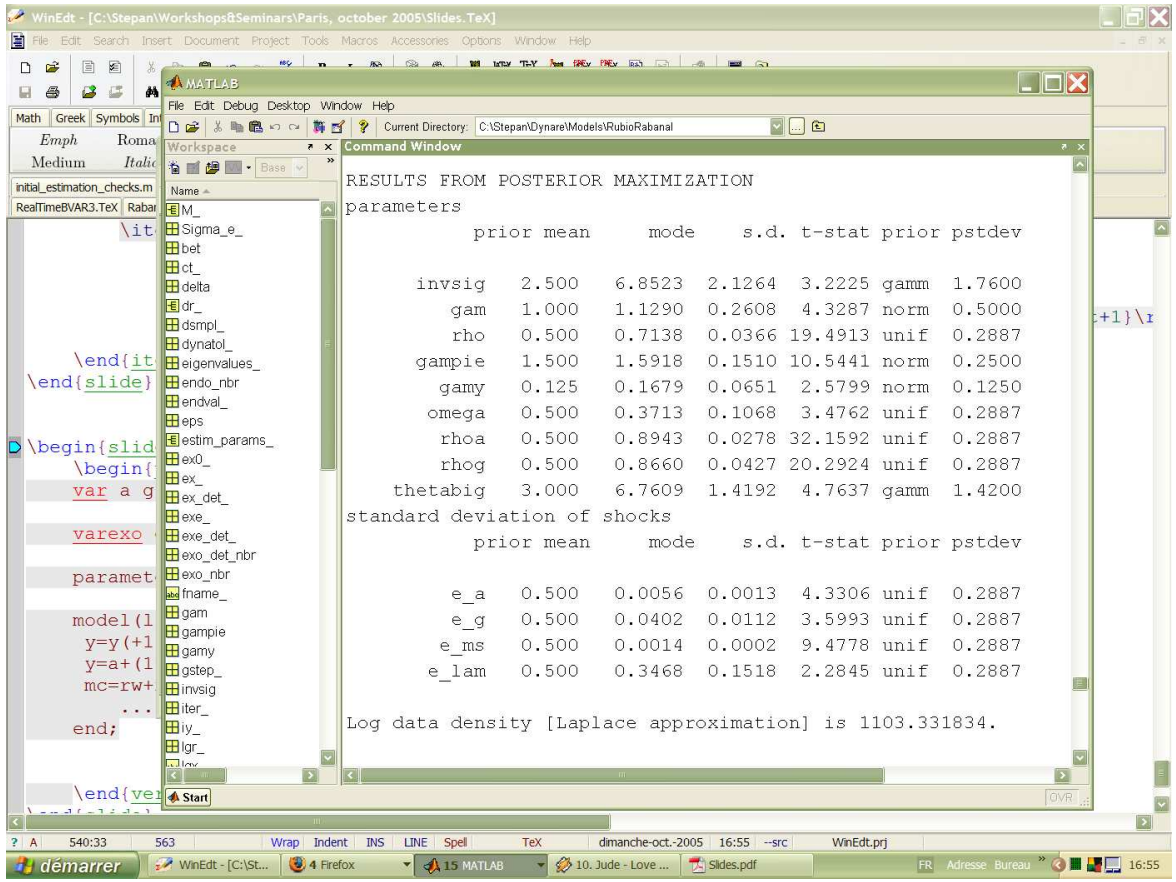
- Zero-one loss function: $\widehat{\psi}$ = posterior mode

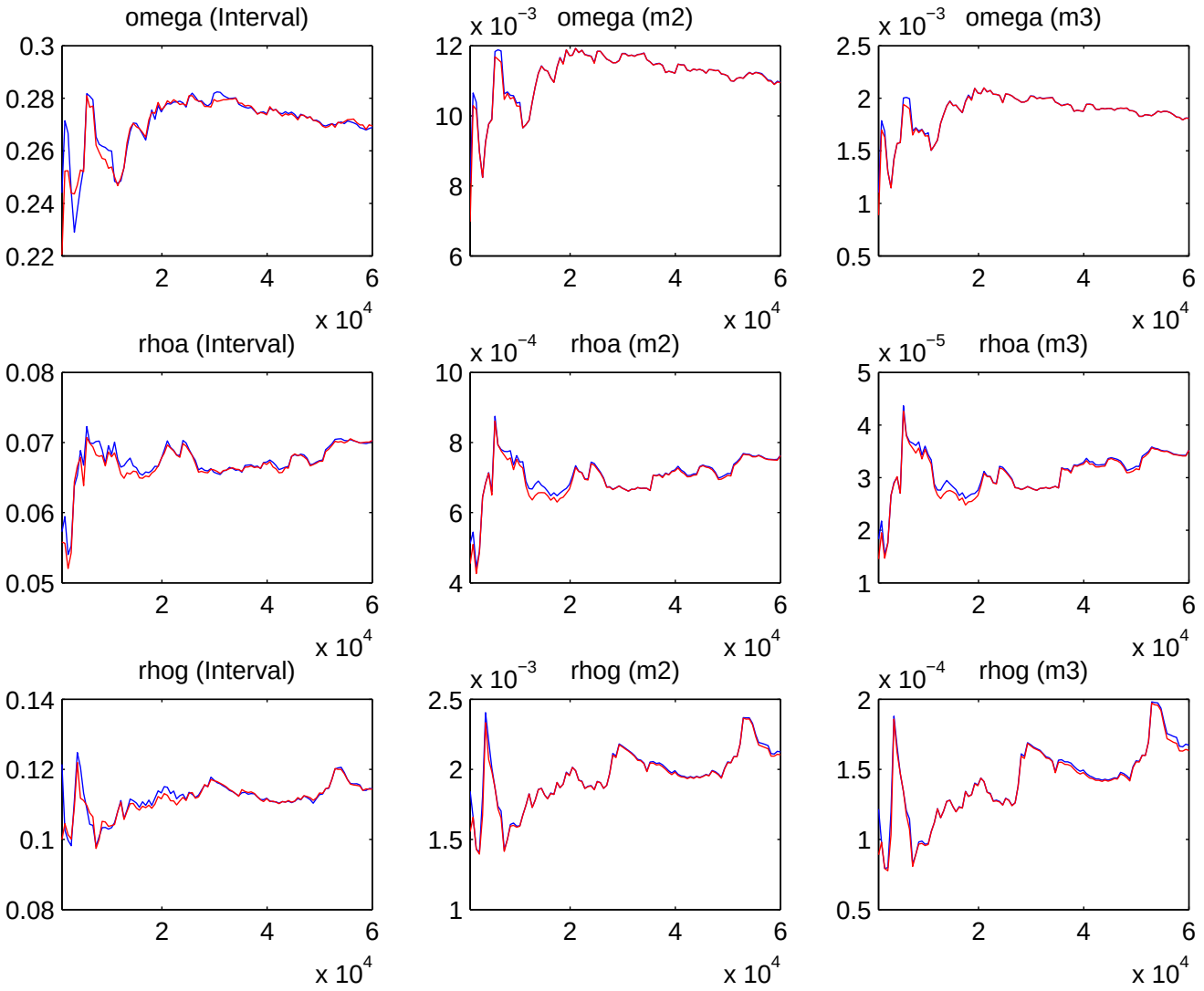
- New keynesian models.
- Common equations :
  - $y_t = \mathbb{E}_t y_{t+1} - \sigma(r_t - \mathbb{E}_t \Delta p_{t+1} + \mathbb{E}_t g_{t+1} - g_t)$
  - $y_t = a_t + (1-\delta) n_t$
  - $mc_t = w_t - p_t + n_t - y_t$
  - $mrs_t = \frac{1}{\sigma} y_t + \gamma n_t - g_t$
  - $r_t = \rho_r r_{t-1} +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

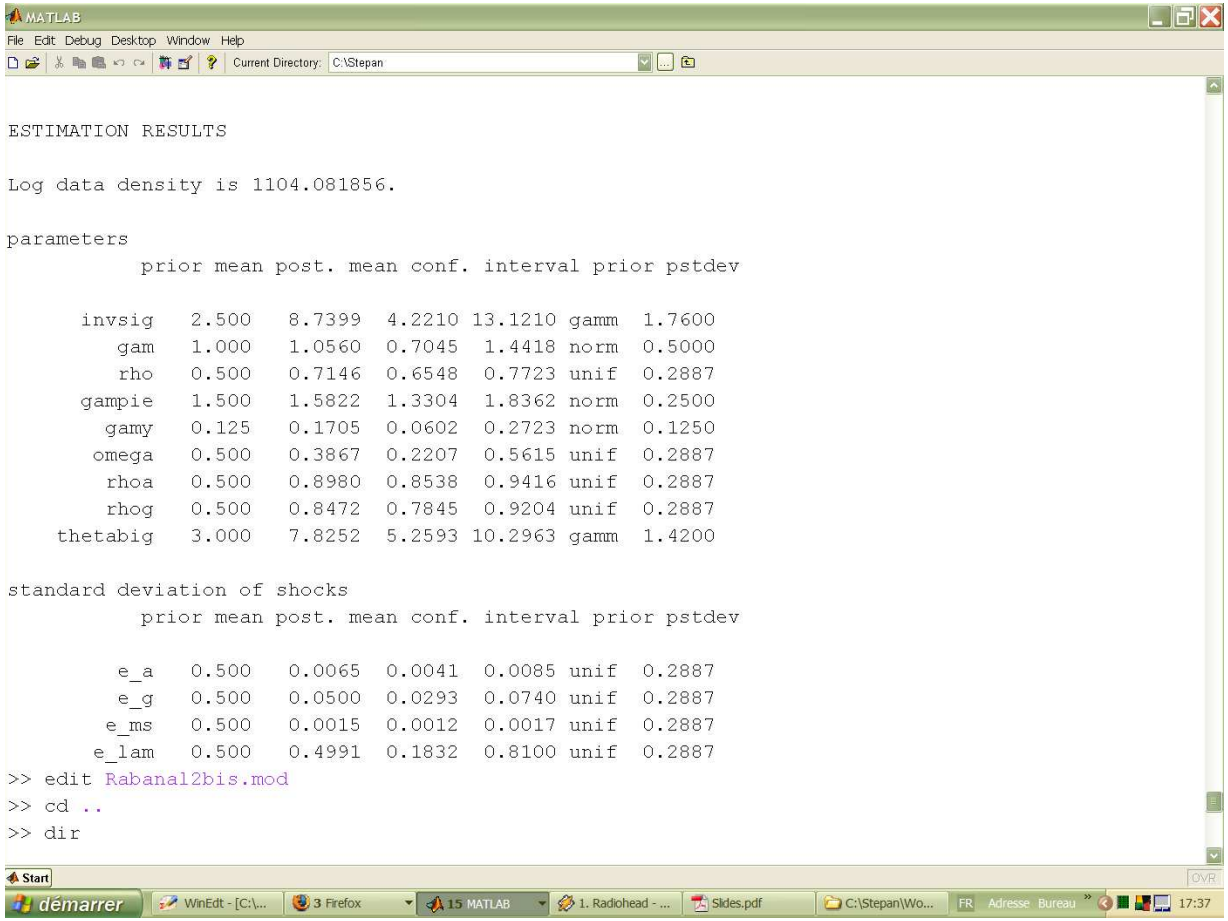
- Baseline sticky prices model (BSP) :
  - $\Delta p_t = \beta \mathbb{E}\left[\Delta p_{t+1} + \kappa_p (mc_t + \lambda_t)\right]$
  - $w_t - p_t = mrs_t$
- Sticky prices & Price indexation (INDP) :
  - $\Delta p_t = \gamma_b \Delta p_{t-1} + \gamma_f \mathbb{E}\left[\Delta p_{t+1} + \kappa'_p (mc_t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

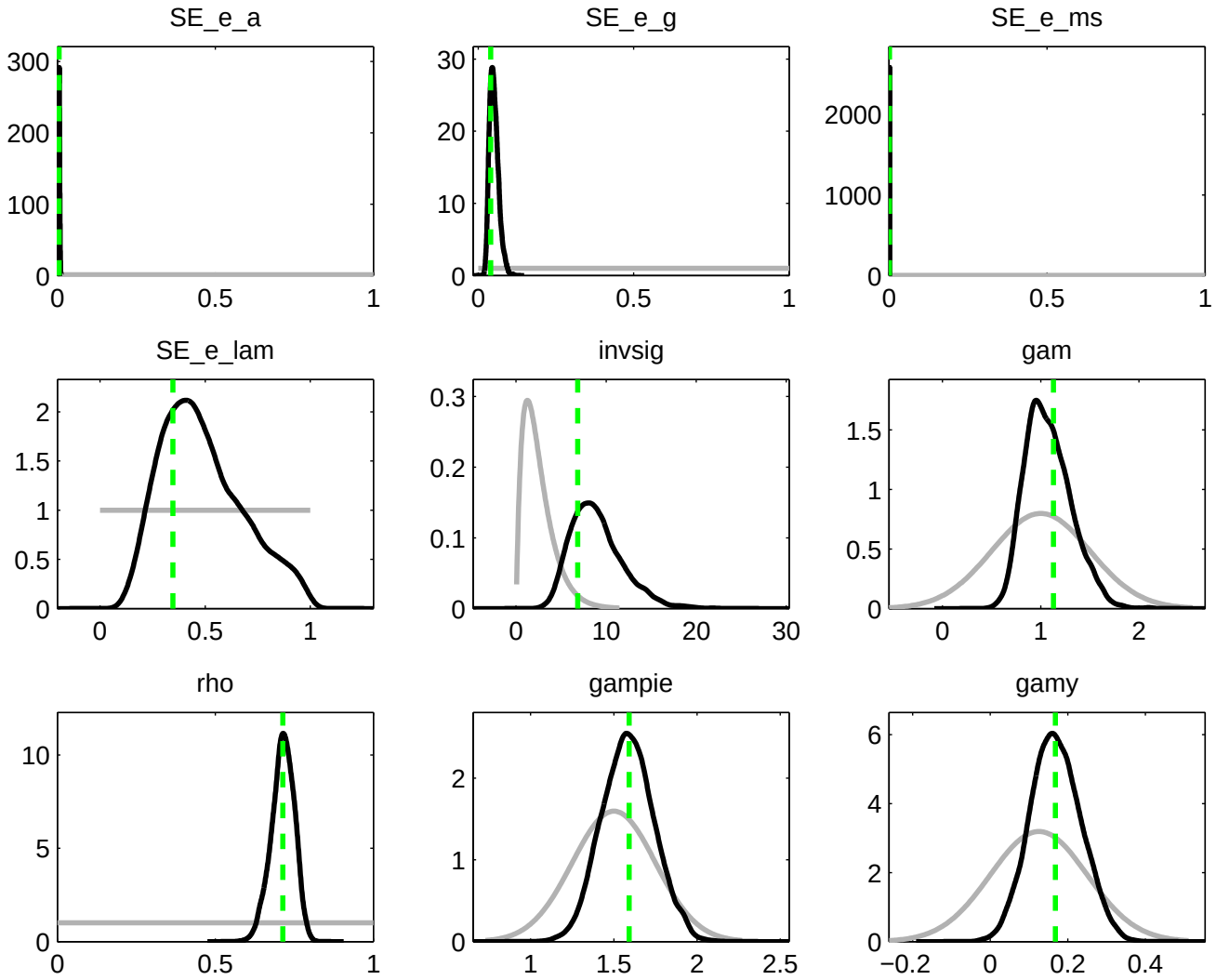
```
var a g mc mrs n pie r rw winf y;

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;
```

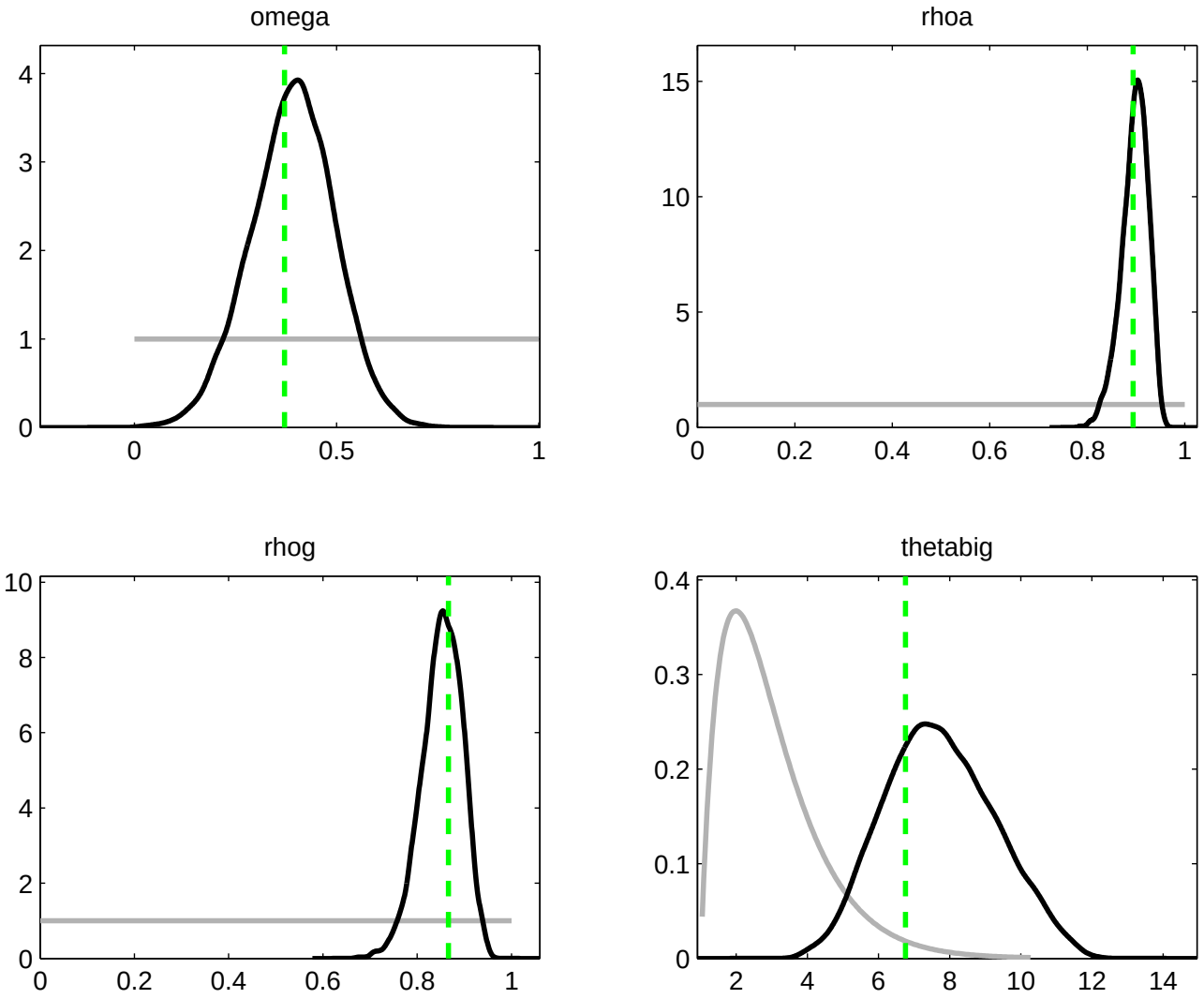
```
estimated_params;
  stderr e_a, uniform_pdf,,,0,1;
  stderr e_lam, uniform_pdf,,,0,1;
    ....
  gampie, normal_pdf, 1.5, 0.25;
    ....
end;

varobs pie r y rw;

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


WinEdt - [C:\Stepan\Workshops&Seminars\Paris, october 2005\Slides.TeX]
MATLAB
Command Window
RESULTS FROM POSTERIOR MAXIMIZATION
parameters
prior mean mode s.d. t-stat prior pstdev
invsig 2.500 6.8523 2.1264 3.2225 gamm 1.7600
gam 1.000 1.1290 0.2608 4.3287 norm 0.5000
rho 0.500 0.7138 0.0366 19.4913 unif 0.2887
gampie 1.500 1.5918 0.1510 10.5441 norm 0.2500
gamy 0.125 0.1679 0.0651 2.5799 norm 0.1250
omega 0.500 0.3713 0.1068 3.4762 unif 0.2887
rhoa 0.500 0.8943 0.0278 32.1592 unif 0.2887
rhog 0.500 0.8660 0.0427 20.2924 unif 0.2887
thetabig 3.000 6.7609 1.4192 4.7637 gamm 1.4200
standard deviation of shocks
prior mean mode s.d. t-stat prior pstdev
e_a 0.500 0.0056 0.0013 4.3306 unif 0.2887
e_g 0.500 0.0402 0.0112 3.5993 unif 0.2887
e_ms 0.500 0.0014 0.0002 9.4778 unif 0.2887
e_lam 0.500 0.3468 0.1518 2.2845 unif 0.2887
Log data density [Laplace approximation] is 1103.331834.


omega (Interval)
0.3
0.28
0.26
0.24
0.22
2
4
6
x 10^4
x 10^-3 omega (m2)
12
10
8
6
2
4
6
x 10^4
x 10^-3 omega (m3)
2.5
2
1.5
1
0.5
2
4
6
x 10^4
rhoa (Interval)
0.08
0.07
0.06
0.05
2
4
6
x 10^4
x 10^-4 rhoa (m2)
10
8
6
4
2
4
6
x 10^4
x 10^-5 rhoa (m3)
5
4
3
2
1
2
4
6
x 10^4
rhog (Interval)
0.14
0.12
0.1
0.08
2
4
6
x 10^4
x 10^-3 rhog (m2)
2.5
2
1.5
1
2
4
6
x 10^4
x 10^-4 rhog (m3)
2
1.5
1
0.5
2
4
6
x 10^4


MATLAB
File Edit Debug Desktop Window Help
Current Directory: C:\Stepan
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
Start
démarrer
WinEdt - [C:\...
3 Firefox
15 MATLAB
1. Radiohead - ...
Slides.pdf
C:\Stepan\Wo...
Adresse Bureau
17:37


SE_e_a
SE_e_g
SE_e_ms
SE_e_lam
invsig
gam
rho
gampie
gamy


omega
rhoa
rhog
thetabig