Bayesian Estimation of DSGE Models

Stéphane Adjemian

Université du Maine, GAINS & CEPREMAP

stephane.adjemian@univ-lemans.fr

http://www.dynare.org/stepan

June 28, 2011
Bayesian paradigm (motivations)

- Bayesian estimation of DSGE models with Dynare.
  1. Data are not informative enough...
  2. DSGE models are misspecified.
  3. Model comparison.
- Prior elicitation.
- Efficiency issues.
Bayesian paradigm (basics)

• A model defines a joint probability distribution parametrized function over a sample of variables:

\[ p(\mathcal{Y}_T^*|\theta) \]  

(1)

⇒ Likelihood.

• We Assume that our prior information about parameters can be summarized by a joint probability density function. Let the prior density be \( p_0(\theta) \).

• The posterior distribution is given by (Bayes theorem squared): 

\[ p_1 (\theta|\mathcal{Y}_T^*) = \frac{p_0 (\theta) p(\mathcal{Y}_T^*|\theta)}{p(\mathcal{Y}_T^*)} \]  

(2)
The denominator is defined by

\[ p(Y_T^*) = \int_{\Theta} p_0(\theta) p(Y_T^*|\theta) d\theta \]  

(3)

⇒ the marginal density of the sample.
⇒ A weighted mean of the sample conditional densities over all the possible values for the parameters.

• The posterior density is proportional to the product of the prior density and the density of the sample.

\[ p_1(\theta|Y_T^*) \propto p_0(\theta) p(Y_T^*|\theta) \]

⇒ That’s all we need for any inference about \( \theta \)!

• The prior density deforms the shape of the likelihood!
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 \( \mathcal{Y}_T \equiv (y_1, \ldots, y_T) \). The likelihood is given by:

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

• And the ML estimator of \( \mu \) is:

\[
\hat{\mu}_{ML,T} = \frac{1}{T} \sum_{t=1}^{T} y_t \equiv \bar{y}
\]
• Note that the variance of this estimator is a simple function of the sample size

\[ \text{Var}[\hat{\mu}_{ML,T}] = \frac{1}{T} \]

• Noting that:

\[ \sum_{t=1}^{T} (y_t - \mu)^2 = \nu s^2 + T(\mu - \hat{\mu})^2 \]

with \( \nu = T - 1 \) and \( s^2 = (T - 1)^{-1} \sum_{t=1}^{T} (y_t - \mu)^2 \).

• The likelihood can be equivalently written as:

\[ p(\mathcal{Y}_T | \mu) = (2\pi)^{-\frac{T}{2}} e^{-\frac{1}{2}(\nu s^2 + T(\mu - \hat{\mu})^2)} \]

The two statistics \( s^2 \) and \( \hat{\mu} \) are summing up the sample information.
A simple example (II, bis)

\[
\sum_{t=1}^{T} (y_t - \mu)^2 = \sum_{t=1}^{T} ([y_t - \hat{\mu}] - [\mu - \hat{\mu}])^2
\]

\[
= \sum_{t=1}^{T} (y_t - \hat{\mu})^2 + \sum_{t=1}^{T} (\mu - \hat{\mu})^2 - \sum_{t=1}^{T} (y_t - \hat{\mu})(\mu - \hat{\mu})
\]

\[
= \nu s^2 + T(\mu - \hat{\mu})^2 - \left( \sum_{t=1}^{T} y_t - T\hat{\mu} \right) (\mu - \hat{\mu})
\]

\[
= \nu s^2 + T(\mu - \hat{\mu})^2
\]

The last term cancels out by definition of the sample mean.
Let our prior be a gaussian distribution with expectation $\mu_0$ and variance $\sigma^2_\mu$.

The posterior density is defined, up to a constant, by:

\[
p (\mu|\mathcal{Y}_T) \propto (2\pi \sigma^2_\mu)^{-\frac{1}{2}} e^{-\frac{1}{2} \frac{(\mu-\mu_0)^2}{\sigma^2_\mu}} \times (2\pi)^{-\frac{T}{2}} e^{-\frac{1}{2} (\nu s^2 + T(\mu - \mu))^2}
\]

where the missing constant (denominator) is the marginal density (does not depend on $\mu$).

We also have:

\[
p(\mu|\mathcal{Y}_T) \propto \exp \left\{ -\frac{1}{2} \left( T(\mu - \mu)^2 + \frac{1}{\sigma^2_\mu} (\mu - \mu_0)^2 \right) \right\}
\]
A simple example (IV)

\[ A(\mu) = T(\mu - \hat{\mu})^2 + \frac{1}{\sigma^2_{\mu}}(\mu - \mu_0)^2 \]

\[ = T(\mu^2 + \hat{\mu}^2 - 2\mu\hat{\mu}) + \frac{1}{\sigma^2_{\mu}}(\mu^2 + \mu_0^2 - 2\mu\mu_0) \]

\[ = (T + \frac{1}{\sigma^2_{\mu}})\mu^2 - 2\mu\left(T\hat{\mu} + \frac{1}{\sigma^2_{\mu}}\mu_0\right) + \left(T\hat{\mu}^2 + \frac{1}{\sigma^2_{\mu}}\mu_0^2\right) \]

\[ = \left(T + \frac{1}{\sigma^2_{\mu}}\right)\left[\mu^2 - 2\mu\frac{T\hat{\mu} + \frac{1}{\sigma^2_{\mu}}\mu_0}{T + \frac{1}{\sigma^2_{\mu}}}\right] + \left(T\hat{\mu}^2 + \frac{1}{\sigma^2_{\mu}}\mu_0^2\right) \]

\[ = \left(T + \frac{1}{\sigma^2_{\mu}}\right)\left[\mu - \frac{T\hat{\mu} + \frac{1}{\sigma^2_{\mu}}\mu_0}{T + \frac{1}{\sigma^2_{\mu}}}\right]^2 + \left(T\hat{\mu}^2 + \frac{1}{\sigma^2_{\mu}}\mu_0^2\right) \]

\[ - \frac{\left(T\hat{\mu} + \frac{1}{\sigma^2_{\mu}}\mu_0\right)^2}{T + \frac{1}{\sigma^2_{\mu}}} \]
Finally we have:

\[
p(\mu | Y_T) \propto \exp \left\{ -\frac{1}{2} \left( T + \frac{1}{\sigma^2_\mu} \right) \left[ \mu - \frac{T \hat{\mu} + \frac{1}{\sigma^2_\mu} \mu_0}{T + \frac{1}{\sigma^2_\mu}} \right]^2 \right\}
\]

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

\[
E[\mu] = \frac{T \hat{\mu} + \frac{1}{\sigma^2_\mu} \mu_0}{T + \frac{1}{\sigma^2_\mu}}
\]

and (posterior) variance:

\[
\mathbb{V}[\mu] = \frac{1}{T + \frac{1}{\sigma^2_\mu}}
\]
A simple example (VI, The bridge)

- The posterior mean is a convex combination of the prior mean and the ML estimate.
  - If $\sigma^{2}_{\mu} \to \infty$ (no prior information) then $\mathbb{E}[\mu] \to \hat{\mu}$ (ML).
  - If $\sigma^{2}_{\mu} \to 0$ (calibration) then $\mathbb{E}[\mu] \to \mu_0$.
- If $\sigma^{2}_{\mu} < \infty$ 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 (Gaussian) approximation.
  - Simulation based approach (MCMC, Metropolis-Hastings, ...).
• Comparison of marginal densities of the (same) data across models.

• \( p(\mathcal{Y}_T^*|\mathcal{I}) \) measures the fit of model \( \mathcal{I} \).

• Suppose we have a prior distribution over models \( \mathcal{A}, \mathcal{B}, ... \): \( p(\mathcal{A}), p(\mathcal{B}), ... \)

• Again, using the Bayes theorem we can compute the posterior distribution over models:

\[
p(\mathcal{I}|\mathcal{Y}_T^*) = \frac{p(\mathcal{I})p(\mathcal{Y}_T^*|\mathcal{I})}{\sum_{\mathcal{I}} p(\mathcal{I})p(\mathcal{Y}_T^*|\mathcal{I})}
\]
Estimation of DSGE models (I, Reduced form)

- Compute the steady state of the model (a system of non linear recurrence equations.

- Compute linear approximation of the model.

- Solve the linearized model:

$$y_t - \bar{y}(\theta) = T(\theta)(y_{t-1} - \bar{y}(\theta)) + R(\theta) \varepsilon_t$$

where $n$ is the number of endogenous variables, $q$ is the number of structural innovations.

- The reduced form model is non linear w.r.t the deep parameters.

- We do not observe all the endogenous variables.
Estimation of DSGE models (II, SSM)

• Let $y_t^*$ be a subset of $y_t$ gathering $p$ observed variables.

• To bring the model to the data, we use a state-space representation:

$$y_t^* = Z (y_t + \bar{y}(\theta)) + \eta_t \quad (5a)$$

$$\hat{y}_t = T(\theta) \hat{y}_{t-1} + R(\theta) \varepsilon_t \quad (5b)$$

where $\hat{y}_t = y_t - \bar{y}(\theta)$.

• Equation (5b) is the reduced form of the DSGE model.  
  $\Rightarrow$ state equation

• Equation (5a) selects a subset of the endogenous variables,  
  $Z$ is a $p \times n$ matrix filled with zeros and ones.  
  $\Rightarrow$ measurement equation
• Let $\mathcal{Y}_T^* = \{y_1^*, y_2^*, \ldots, y_T^*\}$ be the sample.

• Let $\psi$ be the vector of parameters to be estimated ($\theta$, the covariance matrices of $\varepsilon$ and $\eta$).

• The likelihood, that is the density of $\mathcal{Y}_T^*$ conditionally on the parameters, is given by:

$$
\mathcal{L}(\psi; \mathcal{Y}_T^*) = p(\mathcal{Y}_T^* | \psi) = p(y_0^* | \psi) \prod_{t=1}^{T} p(y_t^* | \mathcal{Y}_{t-1}^*, \psi) \quad (6)
$$

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

• The state-space model (5), describes the evolution of the endogenous variables’ distribution.

• The distribution of the initial condition \( (y_0) \) is set equal to the ergodic distribution of the stochastic difference equation (so that the distribution of \( y_t \) is time invariant).

• Because we consider a linear(ized) reduce form model and the disturbances are supposed to be gaussian (say \( \varepsilon \sim \mathcal{N}(0, \Sigma) \)) then the initial (ergodic) distribution is also gaussian:

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

• Unit roots (diffuse kalman filter).
• Evaluation of the density of $y_t^*|\mathcal{Y}_{t-1}^*$ is not trivial, because $y_t^*$ also depends on unobserved endogenous variables.

• The following identity can be used:

$$p\left(y_t^*|\mathcal{Y}_{t-1}^*, \psi\right) = \int_{\Lambda} p\left(y_t^*|y_t, \psi\right) p(y_t|\mathcal{Y}_{t-1}^*, \psi) dy_t$$  \hspace{1cm} (7)

The density of $y_t^*|\mathcal{Y}_{t-1}^*$ is the mean of the density of $y_t^*|y_t$ weighted by the density of $y_t|\mathcal{Y}_{t-1}^*$.

• The first conditional density is given by the measurement equation (5a).

• A Kalman filter is used to evaluate the density of the latent variables ($y_t$) conditional on the sample up to time $t-1$ ($\mathcal{Y}_{t-1}^*$) \[\Rightarrow\] predictive density \].
The Kalman filter can be seen as a bayesian recursive estimation routine:

\[
p(y_t | \mathcal{Y}_{t-1}^*, \psi) = \int_{\Lambda} p(y_t | y_{t-1}, \psi) p(y_{t-1} | \mathcal{Y}_{t-1}^*, \psi) \, dy_{t-1} \quad (8a)
\]

\[
p(y_t | \mathcal{Y}^*, \psi) = \frac{p(y_t^* | y_t, \psi) p(y_t | \mathcal{Y}_{t-1}^*, \psi)}{\int_{\Lambda} p(y_t^* | y_t, \psi) p(y_t | \mathcal{Y}_{t-1}^*, \psi) \, dy_t} \quad (8b)
\]

Equation (8a) 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 (5b), weighted by the density \( y_{t-1} \) conditional on \( \mathcal{Y}_{t-1}^* \) (given by (8b)).

The update equation (8b) is an application of the Bayes theorem → how to update our knowledge about the latent variables when new information (data) becomes available.
Estimation (III, Likelihood) – e –

\[
p(y_t|\mathcal{Y}_t^*, \psi) = \frac{p(y_t^*|y_t, \psi) p(y_t|\mathcal{Y}_{t-1}^*, \psi)}{\int_{\Lambda} p(y_t^*|y_t, \psi) p(y_t|\mathcal{Y}_{t-1}^*, \psi) \, dy_t}
\]

• \(p(y_t|\mathcal{Y}_{t-1}^*, \psi)\) is the a priori density of the latent variables at time \(t\).

• \(p(y_t^*|y_t, \psi)\) is the density of the observation at time \(t\) knowing the state and the parameters (this density is obtained from the measurement equation (5a)) \(\Rightarrow\) the likelihood associated to \(y_t^*\).

• \(\int_{\Lambda} p(y_t^*|y_t, \psi) p(y_t|\mathcal{Y}_{t-1}^*, \psi) \, dy_t\) is the marginal density of the new information.
The linear–gaussian Kalman filter recursion is given by:

\[ v_t = y_t^* - Z(\hat{y}_t + \bar{y}(\theta)) \]
\[ F_t = ZP_tZ' + \nabla [\eta] \]
\[ K_t = T(\theta)P_tT(\theta)'F_t^{-1} \]
\[ \hat{y}_{t+1} = T(\theta)\hat{y}_t + K_tv_t \]
\[ P_{t+1} = T(\theta)P_t(T(\theta) - K_tZ)' + R(\theta)\Sigma R(\theta)' \]

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

Finally the (log)-likelihood is:

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

**References:** Harvey, Hamilton.
Simulations for exact posterior analysis

• Noting that:

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

we can use the empirical mean of
\((\varphi(\psi^{(1)}), \varphi(\psi^{(2)}), \ldots, \varphi(\psi^{(n)}))\), where \(\psi^{(i)}\) are draws from the posterior distribution to evaluate the expectation of \(\varphi(\psi)\). The approximation error goes to zero when \(n \to \infty\).

• We need to simulate draws from the posterior distribution
  \(\Rightarrow\) Metropolis-Hastings.

• We build a stochastic recurrence whose limiting distribution is the posterior distribution.
1. Choose a starting point \( \Psi^0 \) & run a loop over 2-3-4.

2. Draw a proposal \( \Psi^* \) from a jumping distribution

\[
J(\Psi^* | \Psi^{t-1}) = \mathcal{N}(\Psi^{t-1}, c \times \Omega_m)
\]

3. Compute the acceptance ratio

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

4. Finally

\[
\Psi^t = \begin{cases} 
\Psi^* & \text{with probability min}(r, 1) \\
\Psi^{t-1} & \text{otherwise}.
\end{cases}
\]
Simulations (Metropolis-Hastings) – b –
Simulations (Metropolis-Hastings) – c –

\[ \mathcal{K}(\theta^1) = \mathcal{K}(\theta^*) \]

\[ \mathcal{K}(\theta^o) \]

\[ \theta^o \quad \theta^1 = \theta^* \]

posterior kernel
Simulations (Metropolis-Hastings) – d –

\[ K(\theta^0) \]

\[ K(\theta^1) \]

\[ K(\theta^*) \]

\[ \theta^0 \]

\[ \theta^1 \]

\[ \theta^* \]
• How should we choose the scale factor $c$ (variance of the jumping distribution) ?

• The acceptance rate 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**.
Dynare syntax (I)

var A B C;

varexo E;

parameters a b c d e f;

model(linear);
    A=A(+1)-b/e*(B-C(+1)+A(+1)-A);
    C=f*A+(1-d)*C(-1);
    ....
end;
estimated_params;
  stderr e, uniform_pdf,,,0,1;
  ....
  a, normal_pdf, 1.5, 0.25;
  b, gamma_pdf, 1, 2;
  c, gamma_pdf, 2, 2, 1;
  ....
end;

varobs pie r y rw;

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

- The results may depend heavily on our choice for the prior density or the parametrization of the model (not asymptotically).

- How to choose the prior?
  - Subjective choice (data driven or theoretical), example: the Calvo parameter for the Phillips curve.

- Robustness of the results must be evaluated:
  - Try different parametrization.
  - Use more general prior densities.
  - Uninformative priors.
Prior Elicitation (parametrization of the model) – a –

- Estimation of the Phillips curve:

\[
\pi_t = \beta \mathbb{E}\pi_{t+1} + \frac{(1 - \xi_p)(1 - \beta \xi_p)}{\xi_p} \left( (\sigma_c + \sigma_l) y_t + \tau_t \right)
\]

- \( \xi_p \) 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 \( \beta, \sigma_c \) and \( \sigma_l \) are known.
• The prior may be defined on $\xi_p$, $\alpha_p$ or the slope $\lambda$.

• Say we choose a uniform prior for the Calvo probability:

$$\xi_p \sim U[.51,.99]$$

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

• An alternative would be to choose a uniform prior for $\alpha_p$:

$$\alpha_p \sim U[1-1.51,1-1.99]$$

• These two priors are very different!
The prior on $\alpha_p$ is much more informative than the prior on $\xi_p$. 
Implied prior density of $\xi_p = 1 - \frac{1}{\alpha_p}$ if the prior density of $\alpha_p$ 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 \( \mu \) instead of a gaussian density.
• If a parameter, say $\mu$, can take values between $-\infty$ and $\infty$, the flat prior is a uniform density between $-\infty$ and $\infty$.

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

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

• Invariance.

• Why is this prior non informative?... $\int p_0(\mu) d\mu$ is not defined! $\Rightarrow$ Improper prior.

• Practical implications for DSGE estimation.
Prior Elicitation (non informative prior)

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

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

with

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

- 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) ⇒ Unit root controversy.
Effective prior mass

- Dynare excludes parameters such that the steady state does not exist, or such that the BK conditions are not satisfied.
- The effective prior mass can be less than 1.
- Comparison of marginal densities of the data is not informative if the prior mass is not invariant across models.
- The estimation of the posterior mode is more difficult if the effective prior mass is less than 1.
• If possible use options `use_dll` (need a compiler, gcc)

• Do not let dynare compute the steady state!

• Even if there is no closed form solution for the steady state, the static model can be concentrated and reduced to a small nonlinear system of equations, which can be solved by using standard newton algorithm in the `steastate` file. This numerical part can be done in a `mex` routine called by the `steastate` file.

• Alternative initialization of the Kalman filter (`lik_init=4`).
### Table 1: Estimation of fs2000.

In percentage the maximum deviation is less than 0.45.

<table>
<thead>
<tr>
<th></th>
<th>Fixed point of the state equation</th>
<th>Fixed point of the Riccati equation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Execution time</td>
<td>7.28</td>
<td>3.31</td>
</tr>
<tr>
<td>Likelihood</td>
<td>-1339.034</td>
<td>-1338.915</td>
</tr>
<tr>
<td>Marginal density</td>
<td>1296.336</td>
<td>1296.203</td>
</tr>
<tr>
<td>$\alpha$</td>
<td>0.3526</td>
<td>0.3527</td>
</tr>
<tr>
<td>$\beta$</td>
<td>0.9937</td>
<td>0.9937</td>
</tr>
<tr>
<td>$\gamma_a$</td>
<td>0.0039</td>
<td>0.0039</td>
</tr>
<tr>
<td>$\gamma_m$</td>
<td>1.0118</td>
<td>1.0118</td>
</tr>
<tr>
<td>$\rho$</td>
<td>0.6769</td>
<td>0.6738</td>
</tr>
<tr>
<td>$\psi$</td>
<td>0.6508</td>
<td>0.6508</td>
</tr>
<tr>
<td>$\delta$</td>
<td>0.0087</td>
<td>0.0087</td>
</tr>
<tr>
<td>$\sigma_{\varepsilon_a}$</td>
<td>0.0146</td>
<td>0.0146</td>
</tr>
<tr>
<td>$\sigma_{\varepsilon_m}$</td>
<td>0.0043</td>
<td>0.0043</td>
</tr>
</tbody>
</table>