DYNARE COURSE
Amsterdam University
Bayesian estimation with Dynare

Michel Juillard

October 10, 2008

Exercice feedback

- Use Dynare version 4.0.1, yesterday's snapshot introduced a bug in forecasting
- Only `./matlab` should be added to Matlab path. For example

```
addpath c:/dynare/4.0.1/matlab
```

Introduction to Bayesian estimation

- Uncertainty and *a priori* knowledge about the model and its parameters are described by prior probabilities
- Confrontation to the data leads to a revision of these probabilities (posterior probabilities)
- Point estimates are obtained by minimizing a loss function (analogous to economic decision under uncertainty)
- Testing and model comparison is done by comparing posterior probabilities

Bayesian ingredients

- Choosing prior density
- Computing posterior mode
- Simulating posterior distribution
- Computing point estimates and confidence regions
- Computing posterior probabilities

Prior density

$$p(\boldsymbol{\theta}_A|A)$$

where A represents the model and $\boldsymbol{\theta}_A$, the parameters of that model.
The prior density describes *a priori* beliefs, before considering the data.

Likelihood function

- Conditional density
$$p(\boldsymbol{y}|\boldsymbol{\theta}_A, A)$$
- Conditional density for dynamic timeseries models
$$p(\mathbf{Y}_T|\boldsymbol{\theta}_A, A) = p(y_0|\boldsymbol{\theta}_A, A)\prod_{t=1}^{T} p(y_t|\mathbf{Y}_{T-1}, \boldsymbol{\theta}_A, A)$$
where $\mathbf{Y}_T$ are the observations until period T
- Likelihood function
$$\mathcal{L}(\boldsymbol{\theta}_A|\mathbf{Y}_T, A) = p(\mathbf{Y}_T|\boldsymbol{\theta}_A, A)$$

Marginal density

$$\begin{aligned} p(\boldsymbol{y}|A) &= \int_{\Theta_A} p(\boldsymbol{y}, \boldsymbol{\theta}_A|A)d\boldsymbol{\theta}_A \\ &= \int_{\Theta_A} p(\boldsymbol{y}|\boldsymbol{\theta}_A, A)p(\boldsymbol{\theta}_A|A)d\boldsymbol{\theta}_A \end{aligned}$$

Posterior density

- Posterior density
$$p(\boldsymbol{\theta}_A|\mathbf{Y}_T, A) = \frac{p(\boldsymbol{\theta}_A|A)p(\mathbf{Y}_T|\boldsymbol{\theta}_A, A)}{p(\mathbf{Y}_T|A)}$$
- Unnormalized posterior density or posterior density kernel
$$p(\boldsymbol{\theta}_A|\mathbf{Y}_T, A) \propto p(\boldsymbol{\theta}_A|A)p(\mathbf{Y}_T|\boldsymbol{\theta}_A, A)$$

Posterior predictive density

$$
\begin{aligned}
p(\tilde{\mathbf{Y}}|\mathbf{Y}_T, A) &= \int_{\Theta_A} p(\tilde{\mathbf{Y}}, \boldsymbol{\theta}_A|\mathbf{Y}_T, A) d\boldsymbol{\theta}_A \\
&= \int_{\Theta_A} p(\tilde{\mathbf{Y}}|\boldsymbol{\theta}_A, \mathbf{Y}_T, A) p(\boldsymbol{\theta}_A|\mathbf{Y}_T, A) d\boldsymbol{\theta}_A
\end{aligned}
$$

Bayes risk function

$$
\begin{aligned}
R(a) &= E\left[L(a, \boldsymbol{\theta})\right] \\
&= \int_{\Theta_A} L(a, \boldsymbol{\theta}_A) p(\boldsymbol{\theta}_A) d\boldsymbol{\theta}_A
\end{aligned}
$$

where $L(a, \boldsymbol{\theta})$ is the loss function associated with decision a when parameters take value $\boldsymbol{\theta}_A$.

Estimation

Action: deciding that the estimated value of $\boldsymbol{\theta}_A$ is $\widetilde{\boldsymbol{\theta}}_A$

- Point estimate:

$$
\widehat{\boldsymbol{\theta}}_A = \arg\min_{\widetilde{\boldsymbol{\theta}}_A} \int_{\Theta_A} L(\widetilde{\boldsymbol{\theta}}_A, \boldsymbol{\theta}_A) p(\boldsymbol{\theta}_A|\mathbf{Y}_T, A) d\boldsymbol{\theta}_A
$$

- Quadratic loss function:

$$
\widehat{\boldsymbol{\theta}}_A = E(\boldsymbol{\theta}_A|\mathbf{Y}_T, A)
$$

- Zero-one loss function: $\widehat{\boldsymbol{\theta}}_A$ = posterior mode

Credible sets

$$
P(\theta \in C) = \int_C p(\theta) d\theta = 1 - \alpha
$$

is a $100(1-\alpha)\%$ credible set for θ with respect to $p(\theta)$.

A $100(1-\alpha)\%$ highest probability density (HPD) credible set for θ with respect to $p(\theta)$ is a $100(1-\alpha)\%$ credible set with the property

$$
p(\theta_1) \geq p(\theta_2) \quad \forall \theta_1 \in C \text{ and } \forall \theta_2 \in \bar{C}
$$

Numerical integration

$$
\begin{aligned}
E(h(\boldsymbol{\theta}_A)) &= \int_{\Theta_A} h(\boldsymbol{\theta}_A) p(\boldsymbol{\theta}_A | \mathbf{Y}_T, A) d\boldsymbol{\theta}_A \\
&\approx \frac{1}{N} \sum_{k=1}^{N} h(\boldsymbol{\theta}_A^k)
\end{aligned}
$$

where $\boldsymbol{\theta}_A^k$ is drawn from $p(\boldsymbol{\theta}_A | \mathbf{Y}_T, A)$.

Metropolis algorithm

1. Draw a starting point $\boldsymbol{\theta}^\circ$ which $p(\boldsymbol{\theta}) > 0$ from a starting distribution $p^\circ(\boldsymbol{\theta})$.

Metropolis algorithm (continued)

2. For $t = 1, 2, \ldots$
   1. Draw a *proposal* $\boldsymbol{\theta}^*$ from a *jumping* distribution

      $$J(\boldsymbol{\theta}^* | \boldsymbol{\theta}^{t-1}) = N(\boldsymbol{\theta}^{t-1}, c\Sigma_{\text{mode}})$$

   2. Compute the acceptance ratio

      $$r = \frac{p(\boldsymbol{\theta}^*)}{p(\boldsymbol{\theta}^{t-1})}$$

   3. Set

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

In practice . . .

- fix scale factor c so as to obtain a 25% average acceptance ratio
- discard first 50% of the draws

Potential Scale Reduction Factor

If we have simulated m independant sequences of n draws, a particular draw of scalar θ is noted θ_{ij} with $i = 1, \ldots, n$ and $j = 1, \ldots, m$.

$$
\begin{aligned}
B &= \frac{n}{m-1} \sum_{j=1}^{m} \left(\bar{\theta}_{\cdot j} - \bar{\theta}_{\cdot\cdot}\right)^2 \\
W &= \frac{1}{m} \sum_{j=1}^{m} \frac{1}{n-1} \sum_{i=1}^{n} \left(\theta_{ij} - \theta_{\cdot j}\right)^2 \\
\widehat{var}^{+}(\theta|\mathbf{Y}_T, A) &= \frac{n-1}{n} W + \frac{1/n}{B} \\
\hat{R} &= \sqrt{\frac{\widehat{var}^{+}(\theta|\mathbf{Y}_T, A)}{W}}
\end{aligned}
$$

Multivariate PSRF

$$
\begin{aligned}
\hat{V} &= \frac{n-1}{n} W + \left(1 + \frac{1}{m}\right) B/n \\
W &= \frac{1}{m(n-1)} \sum_{j=1}^{m} \sum_{i=1}^{n} (\boldsymbol{\theta}_{ij} - \bar{\boldsymbol{\theta}}_{\cdot j})(\boldsymbol{\theta}_{ij} - \bar{\boldsymbol{\theta}}_{\cdot j})' \\
B/n &= \frac{1}{m-1} \sum_{j=1}^{m} (\bar{\boldsymbol{\theta}}_{\cdot j} - \overline{\boldsymbol{\theta}_{\cdot\cdot}})(\bar{\boldsymbol{\theta}}_{\cdot j} - \overline{\boldsymbol{\theta}_{\cdot\cdot}})' \\
\hat{R}^p &= \frac{n-1}{n} + \frac{m+1}{m} \lambda_1
\end{aligned}
$$

λ_1 is the largest eigenvalue of $W^{-1}B/n$

Model comparison

The ratio of posterior probabilities of two models is

$$
\frac{P(A_j|\mathbf{Y}_T)}{P(A_k|\mathbf{Y}_T)} = \frac{P(A_j)}{P(A_k)} \frac{p(\mathbf{Y}_T|A_j)}{p(\mathbf{Y}_T|A_k)}
$$

In favor of the model A_j versus the model A_k:

- the **prior odds ratio** is $P(A_j)/P(A_k)$
- the **Bayes factor** is $p(\mathbf{Y}_T|A_j)/p(\mathbf{Y}_T|A_k)$
- the **posterior odds ratio** is $P(A_j|\mathbf{Y}_T)/P(A_k|\mathbf{Y}_T)$

Laplace approximation

$$
\begin{aligned}
p(\mathbf{Y}_T, A) &= \int_{\boldsymbol{\theta}_A} p(\boldsymbol{\theta}_A|\mathbf{Y}_T, A) p(\boldsymbol{\theta}_A|A) d\boldsymbol{\theta}_A \\
\hat{p}(\mathbf{Y}_T|A) &= (2\pi)^{\frac{k}{2}} |\Sigma_{\boldsymbol{\theta}^M}|^{-\frac{1}{2}} p(\boldsymbol{\theta}_A^M|\mathbf{Y}_T, A) p(\boldsymbol{\theta}_A^M|A)
\end{aligned}
$$

where $\boldsymbol{\theta}_A^M$ is the posterior mode.

Geweke (1999) modified harmonic mean

$$
\begin{aligned}
p(\mathbf{Y}_T|A) &= \int_{\boldsymbol{\theta}_A} p(\boldsymbol{\theta}_A|\mathbf{Y}_T,A)p(\boldsymbol{\theta}_A|A)d\boldsymbol{\theta}_A \\
\hat{p}(\mathbf{Y}_T|A) &= \left[\frac{1}{n}\sum_{i=1}^{n}\frac{f(\boldsymbol{\theta}_A^{(i)})}{p(\boldsymbol{\theta}_A^{(i)}|\mathbf{Y}_T,A)p(\boldsymbol{\theta}_A^{(i)}|A)}\right]^{-1} \\
f(\boldsymbol{\theta}) &= p^{-1}(2\pi)^{\frac{k}{2}}|\Sigma_{\boldsymbol{\theta}}|^{-\frac{1}{2}}\exp\left\{-\frac{1}{2}(\boldsymbol{\theta}-\overline{\boldsymbol{\theta}})'\Sigma_{\boldsymbol{\theta}}{}^{-1}(\boldsymbol{\theta}-\overline{\boldsymbol{\theta}})\right\} \\
&\quad \times\left\{(\boldsymbol{\theta}-\overline{\boldsymbol{\theta}})'\Sigma_{\boldsymbol{\theta}}{}^{-1}(\boldsymbol{\theta}-\overline{\boldsymbol{\theta}})\leq F^{-1}_{\chi^2_k(p)}\right\}
\end{aligned}
$$

with p an arbitrary probability and k, the number of estimated parameters.

The likelihood of DSGE models

A reduced form state space representation:

$$
\begin{aligned}
y_t^* &= M\bar{y}(\theta)+M\hat{y}_t+N(\theta)x_t+\eta_t \\
\hat{y}_t &= g_y(\theta)\hat{y}_{t-1}+g_u(\theta)u_t \\
E(\eta_t\eta_t') &= V(\theta) \\
E(u_tu_t') &= Q(\theta)
\end{aligned}
$$

The log–likelihood is computed with the Kalman filter.

Kalman filter

For $t=1,\dots,T$

$$
\begin{aligned}
v_t &= y_t^*-\bar{y}^*-M\hat{y}_t-Nx_t \\
F_t &= MP_tM'+V \\
K_t &= g_yP_tg_y'F_t^{-1} \\
\hat{y}_{t+1} &= g_y\hat{y}_t+K_tv_t \\
P_{t+1} &= g_yP_t(g_y-K_tM)'+g_uQg_u'
\end{aligned}
$$

with y_1 and P_1 given.

$$
\ln L\left(\theta|Y_T^*\right)=-\frac{Tk}{2}\ln(2\pi)-\frac{1}{2}\sum_{t=1}^{T}|F_t|-\frac{1}{2}v_t'F_t^{-1}v_t
$$

Priors in DYNARE

NORMAL_PDF	$N(\mu,\sigma)$	R
GAMMA_PDF	$G_2(\mu,\sigma,p_3)$	$[p_3,+\infty)$
BETA_PDF	$B(\mu,\sigma,p_3,p_4)$	$[p_3,p_4]$
INV_GAMMA_PDF	$IG_1(\mu,\sigma)$	R^+
UNIFORM_PDF	$U(p_3,p_4)$	$[p_3,p_4]$

By default, $p_3=0$, $p_4=1$.

How to choose priors

- the shape should be consistent with the domain of definition of the parameter
- use values obtained in other studies (micro or macro)
- check the graph of the priors
- check the implication of your priors by running stoch_simul with parameters set at prior mean
- compare moments of endogenous variables in previous simulation with empirical moments of observed variables
- do sensitivity tests by widening your priors

estim_params

Estimated parameters are declared in a `estim_params; ...end;`.
For each estimated parameter, declare the initial value and, optionally, a lower and upper bound.
Example

```
estim_params;
ALPHA1, NORMAL_PDF,0.5, 0.1;
ALPHA2, UNIFORM_PDF,,, 0.2, 0.8;
end;
```

varobs and estimation

Observed variables are declared in `varobs`.
Computing the estimation is triggered by `estimation`.
Required option: `datafile`
Example

```
estim_params;
ALPHA1, NORMAL_PDF,0.5, 0.1;
ALPHA2, UNIFORM_PDF,,, 0.2, 0.8;
end;

varobs Y, PIE, RS;

estimation(datafile=ddd);
```

Usefull options

first_obs=n : first observation (default: 1)
nobs=n or nobs=([n1:n2)]: number of observations (default: the entire data file)
mode_file : filename of previous results (default: none)
compute_mode : optimization algorithm
- 0 : no optimization
- 1 : Maltab's fmincon
- 3 : Matlab's fminunc
- 4 : Chris Sims' csminwel (default)
- 5 : Marco Ratto's robust optpimizer

mode_check : draws objective function in each parameter direction.

More options

- prefilter : 0, no prefiltering; 1, the data are demeaned before estimation (default: 0).
- presample : number of initial periods that don't enter into likelihood computation (default: 0).
- loglinear : computes a log–linear approximation of the model instead of a linear (default) approximation.
- optim=() : changes options for Matlab optimizer (see Matlab `optimset` command).

More options

- moments_varendo : computes posterior distribution of moments of endogenous variables.
- bayesian_irf : computes posterior distribution of IRF's.
- smoother : computes posterior distribution of smoothed variables.
- filtered_vars : computes posterior distribution of filtered variables.
- forcast=n : computes forecasts for n periods.

observation_trends

Linear trends in the observed variables, if they exist, are declared in `observation_trends;...end;`
For each observation variables, the trend is expressed as a function of model parameters.
Example

```
observation_trends;
Y (gam);
P (mu/gam);
end;
```

Exemple

Model from Rabanal et Rubio (2005) "Comparing New Keynesian models of the business cycle" JME.
Compares different types of nominal rigidities

1. Calvo pricing
2. Calvo pricing with indexation
3. Calvo pricing and Calvo wage setting
4. Calvo pricing and Calvo wage setting with indexation

Model

$$
\begin{aligned}
y_t &= y_{t+1|t} - \sigma\left(r_t - \pi_{t+1|t} + g_{t+1|t} - g_t\right) \\
y_t &= a_t + (1-\delta)n_t \\
mc_t &= rw_t + n_t - y_t \\
mrs_t &= \frac{1}{\sigma}y_t + \gamma n_t - g_t \\
r_t &= \rho r_{t-1} + (1-\rho)\left(\gamma_\pi \pi_t + \gamma_y y_t\right) + e_{ms\,t} \\
rw_t &= rw_{t-1} + winf_t - \pi_t \\
a_t &= \rho_a a_{t-1} + e_{a\,t} \\
g_t &= \rho_g g_{t-1} + e_{g\,t} \\
\pi_t &= \beta\pi_{t+1|t} + (1-\delta)(1-\theta_p\beta)\frac{1-\theta_p}{\theta_p(1+\delta(\epsilon-1))}\left(mc_t + e_{\lambda\,t}\right) \\
rw_t &= mrs_t
\end{aligned}
$$

rabanal.mod

```
var y r pie g a n mc rw mrs winf;
varexo e_a e_g e_ms e_lam;

parameters sig delta gam rho gampie gamy rhoa rhog
           bet thetabig eps;

eps=6;
bet=0.99;
delta=0.36;
```

rabanal.mod (continued)

```
model(linear);
#theta_p = thetabig/(thetabig+1);
y=y(+1)-sig*(r-pie(+1)+g(+1)-g);
y=a+(1-delta)*n;
mc=rw+n-y;
mrs=(1/sig)*y+gam*n-g;
r=rho*r(-1)+(1-rho)*(gampie*pie+gamy*y)+e_ms;
rw=rw(-1)+winf-pie;
a=rhoa*a(-1)+e_a;
g=rhog*g(-1)+e_g;
pie=bet*pie(+1)+(1-delta)*(1-theta_p*bet)*(1-theta_p)/
    (theta_p*(1+delta*(eps-1)))*(mc+e_lam);
rw=mrs;
end;
```

rabanal.mod (continued)

```
estimated_params;
stderr e_a, uniform_pdf,,,0,1;
stderr e_g, uniform_pdf,,,0,1;
stderr e_ms, uniform_pdf,,,0,1;
stderr e_lam, uniform_pdf,,,0,1;
sig, inv_gamma_pdf, 0.67, 0.9;
gam, normal_pdf, 1, 0.5;
rho, uniform_pdf,,,0,1;
gampie, normal_pdf,1.5 ,0.5;
gamy, normal_pdf,0.125 ,0.125;
rhoa, uniform_pdf,,,0,1;
rhog, uniform_pdf,,,0,1;
thetabig, gamma_pdf, 2, 1.42;
end;
```

rabanal.mod (continued)

```
varobs pie r y rw;

estimation(datafile=datarabanal,nobs=75,
           mh_replic=20000,mh_jscale=0.6);
```