Center for Quantitative Economics

Estimation of DSGE models

Bayesian Methods

Fabian Gößling, M.Sc.

Institute of Econometrics
University of Münster
Fabian.Goessling@uni-muenster.de

Winter 2016

Westfälische Wilhelms-Universität Münster

Idea

Bayesian Inference is a way of thinking, not a basket of methods. (Christopher Sims)

- **Approach:** The parameters θ are unknown and data d is given

$\rightarrow$ Make conditional statements based on the likelihood (information on data and model structure) and believes (priors) about the parameters

Bayes-rule:

$$p(\theta|Y) = \frac{p(Y|\theta)p(\theta)}{p(Y)} = \frac{p(Y|\theta)p(\theta)}{\int p(Y|\theta)p(\theta)\ Y\theta} \propto p(Y|\theta)p(\theta)$$

- $p(Y|\theta) = p(Y|\theta)$ is the *likelihood*-function
- $p(Y)$ is the *marginal likelihood* of the data
- $p(\theta)$ is the *prior* density
- $p(\theta|Y)$ is the *posterior* density
- $p(Y|\theta)p(\theta)$ is the *posterior kernel* and a non-normalized density

Point Estimates

- A Bayesian point estimator minimizes

$$\mathbb{E}[L(\hat{\theta},\theta)] = \int L(\hat{\theta},\theta)p(\theta|Y)d\theta,$$

  where $L(\hat{\theta},\theta)$ is a chosen loss function.

- Example: mode of the log posterior density

$$\widehat{\theta}_B = \underset{\theta}{\operatorname{argmax}} \{\log p(\theta|Y)\} = \underset{\theta}{\operatorname{argmax}} \{\log p(Y|\theta) + \log p(\theta)\}$$

$\rightarrow$ With *uninformative* priors this yields the ML-point estimator.

Posterior statistics

- Additionally we want to calculate expressions like

$$\mathbb{E}[g(\theta)|Y] = \int g(\theta)p(\theta|Y)d\theta$$

- Examples:
  - Means
  - Quantiles
  - ...

→ Evaluating the integral by Monte-Carlo Methods requires random draws, but sampling from $p(\theta|Y)$ is not possible in general

Idea

Solution:

Use an algorithm to construct a sequence of draws which mimics the posterior density and use the output for Monte-Carlo Methods

→ The course **Selected Topics: Bayesian Statistics and MCMC** provides in-depth knowledge

→ This course provides intuition and applications to DSGE-Models with Dynare

→ Example

An and Schorfheide (2007, S. 132)

The algorithm constructs a Gaussian approximation around the posterior mode and uses a scaled version of the asymptotic covariance matrix as the covariance matrix for the proposal distribution. This allows for an efficient exploration of the posterior distribution at least in the neighborhood of the mode.

Initialization:

- Specify scaling variable c
- Calculate the mode $\widehat{\theta}_B$ by maximizing the log-posterior density (numerically)
- Specify an initial value $\theta^{(0)}$ or draw it from the proposal density $\mathcal{N}(\widehat{\theta}_B, c^2\Sigma_B)$, where Σ_B is the inverse of the Hessian evaluated at the mode

Metropolis-Hastings-algorithm

1. Draw θ^* from the proposal density $\mathcal{N}(\theta^{(s-1)}, c^2\Sigma_B)$
2. Draw a random number U from the Uniform distribution $U(0,1)$
3. Calculate the acceptance probability α:

$$\alpha \equiv \alpha\left(\theta^{(s-1)}, \theta^*\right) = \min\left\{\frac{p\left(Y|\theta^*\right)\ p\left(\theta^*\right)}{p\left(Y|\theta^{(s-1)}\right)\ p\left(\theta^{(s-1)}\right)}, 1\right\}$$

4. If $U \leq \alpha$ set $\theta^{(s)} = \theta^*$, else $\theta^{(s)} = \theta^{(s-1)}$
5. Return to step 1

Properties

- Asymptotic properties:
  - Priors become irrelevant for the determination of the posterior
  - Posterior converges to a degenerate distribution around the true value (spike)
- Posterior distribution allows
  - Calculating Bayesian confidence intervals (credibility sets)
  - Forecasting using the predictive-density:

$$p(y_{T+1}|Y_{1:T}) = \int p(y_{T+1}|\theta, Y_{1:T})p(\theta|Y_{1:T})d\theta$$

  - Model comparison

Model comparison

- Suppose there are models M_i with $i = \{1, 2\}$ which are "true" with prior probability $p_i = P(M_i)$ and $p_1 + p_2 = 1$
- Each model has a set of parameters θ_i with prior $\wp_i(\theta_i)$ and likelihood $p_i(Y|\theta)$.

→ Then we have for all i:

$$P(M_i|Y) = \frac{P(M_i)p_i(Y|M_i)}{p(Y)} \qquad = \frac{p_i \int p_i(Y|\theta_i, M_i)p_i(\theta_i|M_i)d\theta_i}{p(Y)}$$

Model comparison

- *Marginal-likelihood*:

$$m_i(Y) \equiv p_i(Y|M_i) = \int p_i(Y|\theta_i, M_i) p_i(\theta_i|M_i) d\theta_i$$

- *Posterior-odds*:

$$PO_{12} = \frac{P(M_1|Y)}{P(M_2|Y)} = \underbrace{\frac{p_1}{p_2}}_{\text{Prior-odds-Ratio}} \cdot \underbrace{\frac{m_1(Y)}{m_2(Y)}}_{Bayes-factor}$$

- *Posterior-model-probabilities*:

$$P(M_1|Y) = \frac{PO_{12}}{1 + PO_{12}}$$

Model comparison

- *Marginal likelihood* measures the quality of a model to characterize data
- *Posterior-odds* do not hint to a *physically* true model
- $PO_{12} >> 1$ is an indication that the data as well as the priors prefer model M_1
- Guidelines of Jeffrey (1961):
  - $1:1-3:1$ weak evidence for model 1,
  - $10:1-100:1$ strong evidence for model 1,
  - $>100:1$ decisive evidence for model 1.

- The presented choice of the proposal and its covariance matrix is only one of the possibilities
- Common choice for priors: gaussian, (normal, shifted or inverse) Gamma, Beta or the uniform distribution
- Choosing a proper prior one has to consider lower and upper bounds as well as the skewness and kurtosis of the distribution
- Therefore one has to check the robustness (sensitivity) of the results:
  - Different parametrization
  - More general priors
  - Noninformative priors

- More restrictive assumptions are needed compared to the limited information estimation: specification of the distribution of the schocks, i.e. the likelihood.
- Advantages of a *Maximum-Likelihood*-estimation lie in the full characterization of the data-generating-process and the exact, consistent and efficient estimation of the parameters.
- Dilemma of absurd parameter estimates: Problem of the ML-estimation due to wrong distributional assumptions, problems in the optimization algorithm or non-separable identifiable parameters.
- Even transformations, upper and lower bounds, etc. are only limited to help overcome this problem, when the likelihood is flat.

Full information estimators

- Considering priors one can incorporate additional information into a model.
- Dilemma of absurd parameter estimates: Even with Bayesian means it is not possible to estimate these parameters (the posterior looks almost the same as the prior), but one can assign probability such that these parameters are very unlikely.

⇒ Using priors one can exclude these absurd parameter estimates.

- Nevertheless the point of robustness and identification of the parameters remains a critical topic.

Motivation

- So far we assumed that the likelihood $p(Y_{1:T}|\theta)$ is easily available
- This is not true in general!
- The state-space representation of the DSGE model provides us with the *joint* likelihood of observables and states

$$p(Y_{1:T}, S_{1:T}|\theta) = \prod_{t=1}^{T} p(y_t|s_t, \theta)p(s_t|s_{t-1}, \theta)$$

- We need the likelihood $p(Y_{1:T}|\theta)$, i.e. we need to integrate out $S_{1:T}$

$\rightarrow$ *Filtering* (e.g. Herbst and Schorfheide 2015)

Concepts

- Denote $p(s_0|Y_{1:0},\theta) = p(s_0|\theta)$
- Recall:
  - Likelihood:

$$p(Y_{1:T}|\theta) = \prod_{t=1}^{T} p(y_t|Y_{1:t-1},\theta)$$

  - Bayes:

$$p(s_t|y_t) = \frac{p(y_t|s_t)p(s_t)}{p(y_t)}$$

  - Marginalizing:

$$p(s_t) = \int p(s_t|y_t)p(y_t)dy_t$$

- Now we can build a simple filter...

Generic Filter

For $t = 1, ..., T$:

1. Forecast s_t:

$$p(s_t|Y_{1:t-1}, \theta) = \int p(s_t|s_{t-1}, Y_{1:t-1}, \theta)p(s_{t-1}|Y_{1:t-1}, \theta)ds_{t-1}$$

2. Forecast y_t:

$$p(y_t|Y_{1:t-1}, \theta) = \int p(y_t|s_t, Y_{t-1}, \theta)p(s_t|Y_{1:t-1}, \theta)ds_t$$

3. Update s_t:

$$p(s_t|Y_{1:t}, \theta) = p(s_t|y_t, Y_{1:t-1}, \theta) = \frac{p(y_t|s_t, Y_{1:t-1}, \theta)p(s_t|Y_{1:t-1}, \theta)}{p(y_t|Y_{1:t-1}, \theta)}$$

Generic Filter

- Likelihood is a by-product of the filter (product of the predictive densities)
- Kalman filter:
  - All densities are normal
  - Transition/Observation equations are linear
  - $\rightarrow$ All densities are *analytically* available
- If we depart from linearity/Gaussians:
  - Extended Kalman Filter
  - Sigma Point Filter
  - Particle Filter
  - $\rightarrow$ Course: **Econometrics of Filtering**
- Extensive computational time ...

Exercise 1: Bayesian Statistics

(a) Derive the formula for the posterior-odds.

(b) Derive the formula for the posterior-model-probabilites.

(c) Give an interpretation of the posterior-odds and probabilites.

Exercise 2: Estimation with Bayesian methods

Consider the following simplified RBC-model:

$$\max_{\{c_{t+j}, l_{t+j}, k_{t+j}\}_{j=0}^{\infty}} W_t = \sum_{j=0}^{\infty} \beta^j u(c_{t+j}, l_{t+j})$$

$$\begin{aligned} s.t. \quad y_t &= c_t + i_t, & A_t &= Ae^{a_t}, \\ y_t &= A_t f(k_{t-1}, l_t), & a_t &= \rho a_{t-1} + \varepsilon_t, \\ k_t &= i_t + (1-\delta)k_{t-1}, & \varepsilon_t &\sim N(0, \sigma_\varepsilon^2), \end{aligned}$$

where preferences and technology follow:

$$u(c_t, l_t) = \frac{\left[c_t^\theta (1-l_t)^{1-\theta}\right]^{1-\tau}}{1-\tau},$$

$$f(k_{t-1}, l_t) = \left[\alpha k_{t-1}^\psi + (1-\alpha) l_t^\psi\right]^{1/\psi}.$$

Optimality is given by:

$$
\begin{aligned}
u_c(c_t, l_t) - \beta E_t \left\{ u_c(c_{t+1}, l_{t+1}) \left[A_{t+1} f_k(k_t, l_{t+1}) + 1 - \delta \right] \right\} &= 0, \\
-\frac{u_l(c_t, l_t)}{u_c(c_t, l_t)} - A_t f_l(k_{t-1}, l_t) &= 0, \\
c_t + k_t - A_t f(k_{t-1}, l_t) - (1 - \delta) k_{t-1} &= 0.
\end{aligned}
$$

Hint: *Use the Dynare user guide, which provides all necessary commands.*

(a) Write a mod-file for themodel (with a sensible calibration and a steady-state/initival block).

(b) Simulate a sample of 10000 observations for c_t, l_t and y_t using `stoch_simul` and save it in a mat-file.

(c) Define priors for α, θ and τ (or a different set of parameters).***Hint***: *Don't forget to adjust the model equations as well as the steady-state block.*

(d) Estimate the posterior mode using the `estimation` command and a limited sample with 200 observations. How many observable variables do you need? Check the posterior mode using `mode_check`. If you get errors due to a non-positive definite Hessian, try a different optimization algorithm or change the initial values.

Exercise 2: Estimation with Bayesian methods

(e) If you are satisfied with the posterior mode, approximate the posterior distribution using the the Metropolis- Hastings-Algorithm with 3×5000 iterations. If it does not converge to the posterior-distribution, repeat the algorithm without discarding the previous draws.

(f) How robust are the results regarding the specification of the priors? Repeat the estimation of the posterior-mode for different priors.

(g) Use the same dataset to estimate the parameters of a misspecified model. Use the same model, but with a Cobb-Douglas production function. ***Hint***: *Don't forget to adjust the model equations as well as the steady-state block.*

(h) Compare the estimation of the common parameters as well as the marginal densities of the different models. Calculate the *posterior-odds* and the *posterior-model-probabilities*.