Quantitative Macroeconomics

Winter 2023/24

Week 11

Willi Mutschler
willi@mutschler.eu

Version: 1.0
Latest version available on: GitHub

Contents

1. Bayesian Estimation Basics

Consider a simple univariate model:

$$y_t = \mu + u_t$$

with $t = 1, 2, \ldots, T$ and $u_t \sim \mathcal{N}(0, \sigma^2)$. Assume that σ^2 is known. The objective of an econometrician is to estimate μ.

1. How do classical and Bayesian analysis differ?
2. Name the key ingredients for Bayesian estimation.
3. What are "conjugate priors" and "natural conjugate priors"?
4. What is the idea of Monte Carlo integration in the context of Bayesian estimation?

Readings

- Greenberg (2008, Part I)
- Koop (2003, Ch.1-2)

2. Bayesian Estimation of Multivariate Linear Regression Model

Consider a linear regression model with multiple regressors:

$$Y = X\beta + u$$

with $u \sim \mathcal{N}(0, \sigma^2 I)$.

1. Name the idea and general procedure for estimating this model with Bayesian methods.
2. Provide an expression for the likelihood function $p(Y|\beta, \sigma^2)$.
3. Assume that σ^2 is known and the prior distribution for β is Gaussian with mean β_0 and covariance matrix Σ_0. Derive an expression for the **conditional posterior distribution** $p(\beta|\sigma^2, Y)$.
4. Assume that β is known and the prior distribution for the precision $1/\sigma^2$ is Gamma with shape parameter s_0 and scale parameter v_0. Derive an expression for the **conditional posterior distribution** $p(1/\sigma^2|\beta, Y)$.
5. Now assume that both β and σ^2 are unknown. Since we are able to draw directly from the **conditional posterior distributions** (direct sampling), we can use the Gibbs sampling algorithm to get draws from the **joint posterior distribution** $p(\beta, \sigma^2|Y)$. Provide an overview of the basic steps and algorithm of the Gibbs sampling algorithm.

Readings

- Greenberg (2008, Ch. 7.1)
- Koop (2003, Ch. 3)

3. Bayesian Estimation of Quarterly Inflation

Perform a Bayesian estimation using the Gibbs sampler of an autoregressive model with two lags of quarterly US inflation

$$y_t = c + \phi_1 y_{t-1} + \phi_2 y_{t-2} + u_t = Y_{t-1}\theta + u_t$$

where $Y_{t-1} = (1, y_{t-1}, y_{t-2})$, $u_t \sim WN(0, \sigma_u^2)$ and $\theta = (c, \phi_1, \phi_2)'$. To this end, assume a Gamma distribution for the marginal prior for the precision $1/\sigma_u^2$ and a normal distribution for the conditional prior for the coefficients θ given $1/\sigma_u^2$.

1. Load the dataset `QuarterlyInflation.csv`. It contains a series for US quarterly inflation from 1947Q1 to 2012Q3. Plot the data.
2. Create the matrix of regressors and the corresponding vector of endogenous variables for an AR(2) model with a constant.
3. Set the prior mean for the coefficients to a vector of zeros, $\theta_0 = 0$, and the prior covariance matrix to the identity matrix, $\Sigma_0 = I$.
4. Set the shape parameter for the variance parameter to $s_0 = 1$ and the scale parameter to $v_0 = 0.1$.
5. Set the total number of Gibbs iterations to $R = 50000$ with a burn-in phase of $B = 40000$.
6. Initialize output matrices for the remaining $R - B$ draws of the coefficient estimates and the variance estimate.
7. Initialize the first draw of $1/\sigma_u^2$ to its OLS estimate.
8. For $j = 1, \ldots, R$ do the following
   a) Sample $\phi(j)$ conditional on $1/\sigma_u^2(j)$ from $\mathcal{N}(\theta_1, \Sigma_1)$ where
   $$\Sigma_1 = \left(\Sigma_0^{-1} + \sigma_u^{-2}(j)(X'X)\right)^{-1}$$
   $$\theta_1 = \Sigma_1 \cdot (\Sigma_0^{-1}\phi_0 + \sigma_u^{-2}(j)X'y)$$
   Optionally: check the stability of the draw to avoid an explosive AR processes.
   b) Sample $1/\sigma_u^2(j)$ conditional on $\theta(j)$ from the Gamma distribution $G(s_1, v_1)$ where
   $$s_1 = s_0 + T$$
   $$v_1 = v_0 + \sum_{t=3}^{T} (y_t - Y_{t-1}\theta(j))^2$$
   c) If you passed the burn-in phase ($j > B$), then save the draws of $\theta(j)$ and $\sigma^2(j)$ into the output matrices.
9. Plot the histograms of the draws in your output matrices.

Hints

- Use `mvnrnd(theta1,Sigma1)` to draw from a multivariate normal distribution with mean θ_1 and covariance matrix Σ_1.
- Use `gamrnd(s1,1/v1,1,1)` to draw from a Gamma distribution with shape parameter s_1 and scale parameter v_1.

Readings

- Chib and Greenberg (1994)
- Greenberg (2008, Ch. 10.1)

References

Chib, Siddhartha and Edward Greenberg (Sept. 1994). “Bayes Inference in Regression Models with ARMA (p, q) Errors”. In: *Journal of Econometrics* 64.1-2, pp. 183–206. DOI: 10.1016/0304-4076(94)90063-9.

Greenberg, Edward (2008). *Introduction to Bayesian Econometrics*. Cambridge ; New York: Cambridge University Press. ISBN: 978-0-521-85871-7.

Koop, Gary (2003). *Bayesian Econometrics*. Chichester ; Hoboken, N.J: J. Wiley. ISBN: 978-0-470-84567-7.

A. Solutions

1 Solution to Bayesian Estimation Basics

1. In Quantitative Macroeconomics and Econometrics we are concerned with using data to learn about a phenomenon, e.g. the relationship between two macroeconomic variables. That is: we want to learn about something *unknown* (the parameter μ) given something *known* (the data y_t). Let's use the sample mean as our estimating function: $\hat{\mu} = 1/T \sum_{t=1}^{T} y_t$. Due to the law of large numbers and the central limit theorem we can derive that $\hat{\mu} \sim N(\mu, \frac{\sigma^2}{T})$ and conduct inference such as computing confidence intervals $[\hat{\mu} \pm 1.96 \frac{\sigma}{\sqrt{T}}]$.
   Classical/Frequentist approach: μ is a fixed unknown quantity, that is we think there exists a *true value* that is not random. On the other hand, the estimating function, $\hat{\mu}$, is a random variable and is evaluated via repeated sampling. In a thought experiment, we would be able to generate a large number of datasets (given the true μ) and our confidence interval will contain the true value in 95% of cases. The estimator $\hat{\mu}$ is *best* in the sense of having the highest probability of being close to the true μ.
   Bayesian approach: μ is treated as a *random variable*; that is, there is NO true unknown value. Instead our knowledge about the model parameter μ is summarized by a *probability distribution*. In more detail, this distribution summarizes two sources of information:
   a) prior information: subjective beliefs about how likely different parameter values are (information BEFORE seeing the data)
   b) sample information: AFTER seeing the data, we update/revise our prior beliefs

   In a sense we explicitly make use of (subjective) probabilities to quantify uncertainty about the parameter.

2. The key ingredients are based on the rules of probability, which imply for two events A and B: $p(A, B) = p(A|B)p(B)$, where $p(A, B)$ is the joint probability of both events happing simultaneously. $p(A|B)$ is the probability of A occurring conditional on B having occurred; and $p(B)$ is the marginal probability of B. Alternatively, we can reverse A and B to get: $p(A, B) = p(B|A)p(A)$. Equating the two expressions gives you **Bayes' rule**:

$$p(B|A) = \frac{p(A|B)p(B)}{p(A)}$$

   This rule also holds for continuous variables such as parameters θ and data y:

$$p(\theta|y) = \frac{p(y|\theta)p(\theta)}{p(y)}$$

   That is, the key object of interest is the **posterior** $p(\theta|y)$ distribution, which is the product of the **likelihood function** $p(y|\theta)$ and the **prior density** $p(\theta)$, divided by the **marginal data density** $p(y)$. In other words, the prior contains our prior (non-data) information, whereas the likelihood function is the density of the data conditional on the parameters. Note that the marginal data density $p(y)$ can be ignored as it does not depend on the parameters (it is just a normalization constant as a probability density integrates to one). Therefore, we can use the proportional $\propto$ sign, that is the posterior is proportional to the likelihood times the prior:

$$p(\theta|y) \propto p(y|\theta)p(\theta)$$

   The posterior summarizes all we know about θ after seeing the data. It combines both data and non-data information. The equation can be viewed as an updating rule, where data allows us to update our prior views about θ.

   Note that Bayesians are upfront and rigorous about including non-data information! The idea is that more information (even if subjective) tends to be better than less.

3. In principle any distribution can be combined with the likelihood to form the posterior. Some priors are, however, more convenient than others to make use of analytical results.
   Conjugate priors: If a prior is conjugate, then the posterior has the same density as the prior. This eases analytical derivations.
   Natural conjugate priors: A conjugate prior is called a natural conjugate prior, if the posterior and the prior have the same functional form as the likelihood function. That is, the prior can be interpreted as arising from a fictitious dataset from the same data-generating process.

4. The posterior is typically not analytically available and needs to be approximated unless for special cases using e.g. natural conjugate priors. But, typically we are not interested in the exact shape of the posterior, but in certain statistics of the posterior distribution such as:

$$E[\theta|y] = \int_{-\infty}^{\infty} \theta p(\theta|y) d\theta$$
$$V[\theta|y] = \int_{-\infty}^{\infty} \theta^2 p(\theta|y) d\theta - (E(\theta|y))^2$$

   So we only need to approximate the integrals using numerical methods such as Monte Carlo integration. That is, IF we had iid draws from the posterior, we can make use of the law of large numbers and could approximate the posterior mean and variance as:

$$E[\theta|y] \approx \frac{1}{S} \sum_{i=1}^{S} \theta_i$$
$$V[\theta|y] \approx \frac{1}{S} \sum_{i=1}^{S} \theta_i^2 - \left(\frac{1}{N} \sum_{i=1}^{N} \theta_i \right)^2$$

   Or in general for any function:

$$E[f(\theta)|y] = \int_{-\infty}^{\infty} f(\theta) p(\theta|y) d\theta \approx \frac{1}{S} \sum_{s=1}^{S} f(\theta_s)$$

   This is the key idea of Monte Carlo integration, i.e. replace the integral by a sum over S draws from the posterior. The Central Limit Theorem can then be used to asses the accuracy of this approximation. But there are two challenges:

   a) How to draw from the posterior?

   b) How to make sure that the draws are iid?

   The first question can be answered by using suitable *posterior sampling algorithms* such as direct sampling, importance sampling, Metropolis-Hastings sampling, Gibbs sampling, or Sequential Monte-Carlo sampling. The second question is more difficult to answer and requires some knowledge about the sampling algorithm and suitable diagnostics.

2 Solution to Bayesian Estimation of Multivariate Linear Regression Model

1. The parameter vector $[\beta, \sigma^2]'$ is a random variable with a probability distribution. A Bayesian estimation of this distribution combines prior beliefs and information from the data:

   a) Prior distribution $p(\beta, \sigma^2)$

   b) Likelihood $p(Y|\beta, \sigma^2)$

   c) Bayes' rule gives the joint posterior distribution

   Some useful relationships:

   - joint posterior distribution of β and σ^2:

   $$p(\beta, \sigma^2|Y) = \frac{p(Y|\beta, \sigma^2)p(\beta, \sigma^2)}{p(Y)} \propto p(Y|\beta, \sigma^2)p(\beta, \sigma^2)$$

   - marginal posterior distributions of β and σ^2:

   $$p(\beta|Y) = \int_0^\infty p(\beta, \sigma^2|Y)d\sigma^2 \propto \int_0^\infty p(Y|\beta, \sigma^2)p(\beta, \sigma^2)d\sigma^2$$
   $$p(\sigma^2|Y) = \int_{-\infty}^\infty p(\beta, \sigma^2|Y)d\beta \propto \int_{-\infty}^\infty p(Y|\beta, \sigma^2)p(\beta, \sigma^2)d\beta$$

   - conditional posterior distribution of β given σ^2:

   $$p(\beta|\sigma^2, Y) = \frac{p(\beta, \sigma^2|Y)}{p(\sigma^2|Y)} \propto p(Y|\beta, \sigma^2)p(\beta, \sigma^2)$$

   Lastly, the following relationship is useful for simulations:

   $$p(\beta, \sigma^2|Y) = p(\beta|\sigma^2, Y)p(\sigma^2|Y)$$

2. Because u is normally distributed, Y is also Gaussian; hence, we can derive the precise form of the likelihood function:

   $$p(Y|\beta, \sigma^2) = \frac{1}{(2\pi\sigma^2)^{T/2}} e^{\left\{-\frac{1}{2\sigma^2}(Y-X\beta)'(Y-X\beta)\right\}}$$

3. First, assuming that σ^2 is known and the prior for β is Gaussian, we have $p(\beta) \sim N(\beta_0, \Sigma_0)$; that is, the prior density is given by:

   $$p(\beta) = (2\pi)^{-K/2}|\Sigma_0|^{-1/2} e^{\left\{-\frac{1}{2}(\beta-\beta_0)\Sigma_0^{-1}(\beta-\beta_0)\right\}}$$

   Note that the prior for β is independent of σ^2, so we can also write:

   $$p(\beta) = p(\beta|\sigma^2)$$

   Second, conditional on σ^2 the likelihood is proportional to:

   $$p(Y|\beta, \sigma^2) \propto e^{\left\{-\frac{1}{2\sigma^2}(Y-X\beta)'(Y-X\beta)\right\}}$$

   Third, combining prior and likelihood yields:

   $$p(\beta|\sigma^2, Y) \propto e^{\left\{-\frac{1}{2}(\beta-\beta_0)'\Sigma_0^{-1}(\beta-\beta_0)-\frac{1}{2\sigma^2}(Y-X\beta)'(Y-X\beta)\right\}}$$

   One can show (see the references), that this is a Gaussian distribution

   $$p(\beta|\sigma^2, Y) \sim N(\beta_1, \Sigma_1)$$

   with

   $$\beta_1 = \left(\Sigma_0^{-1} + \sigma^{-2}(X'X)\right)^{-1}\left(\Sigma_0^{-1}\beta_0 + \sigma^{-2}(X'Y)\right)$$
   $$\Sigma_1 = \left(\Sigma_0^{-1} + \sigma^{-2}(X'X)\right)^{-1}$$

4. Assuming that β is known and the prior for $1/\sigma^2$ is Gamma, we have $p(1/\sigma^2) = p(1/\sigma^2|\beta) \sim \Gamma(s_0, v_0)$; that is, the prior density is given by:

$$p(1/\sigma^2|\beta) \propto \left(\frac{1}{\sigma^2}\right)^{s_0-1} e^{\left\{-\frac{1}{v_0\sigma^2}\right\}}$$

Conditional on β the likelihood is proportional to:

$$p(Y|\sigma^2, \beta) \propto (\sigma^2)^{-T/2} e^{\left\{-\frac{1}{2\sigma^2}(Y-X\beta)'(Y-X\beta)\right\}}$$

Combining prior and likelihood yields (see the readings for the algebra):

$$p(1/\sigma^2|\beta, Y) \sim \Gamma(s_1, v_1)$$

where

$$\begin{aligned} s_1 &= s_0 + T \\ v_1 &= v_0 + (Y - X\beta)'(Y - X\beta) \end{aligned}$$

5. In the previous exercises we have derived the conditional posteriors in closed-form. When both β and σ^2 are unknown, we can specify the **joint prior** distribution for these parameters assuming a Gamma distribution for the marginal prior for $1/\sigma^2$ and a normal distribution for the conditional prior for $\beta|1/\sigma^2$. That is, the joint prior is then $p(\beta, 1/\sigma^2) \propto p(\beta|1/\sigma^2)p(1/\sigma^2)$. It can then be shown that the joint posterior density is:

$$p(\beta, 1/\sigma^2|Y) = p(\beta|1/\sigma^2, Y)p(1/\sigma^2|Y)$$

To make inference on β, we need to know the marginal posterior

$$p(\beta|Y) = \int_0^\infty p(\beta, 1/\sigma^2|Y)d(1/\sigma^2)$$

This integration is very hard, but we can make use of a numerical Monte Carlo integration approach: **Gibbs sampling**.

The idea of Gibbs sampling is to repeatedly sample from the conditional posterior distributions to get an approximation of the marginal and joint posterior distributions of the parameters.

Basic steps of the Gibbs sampling algorithm:

- Set priors and initial guess for σ^2
- Sample β conditional on $1/\sigma^2$
- Sample $1/\sigma^2$ conditional on β
- Repeat (2) and (3) a large number of times R and keep the last L draws.
- Use the L draws to make inference on β and σ.

3 Solution to Bayesian Estimation of Quarterly Inflation

progs/matlab/BayesianQuarterlyInflation.m

```
% ------------------------------------------------------------------------
% Bayesian estimation of an AR(2) model of quarterly inflation
% ------------------------------------------------------------------------
% Willi Mutschler, January 23, 2024
% willi@mutschler.eu
% ------------------------------------------------------------------------

clearvars; clc;close all;

%% data handling
QuarterlyInflation = importdata('../../data/QuarterlyInflation.csv'); % Load data
data = QuarterlyInflation.data;
T = size(data,1);                       % determine sample length, i.e. how many quarters
X = [ones(T,1) lagmatrix(data,1:2)];   % matrix of regressors, i.e. c, y(t-1) and y(t-2) for AR(2) model with constant
X = X(3:end,:);                         % remove the first 2 observations in regressors
y = data(3:end,1);                      % remove the first 2 observations in dependent variable
T = size(y,1);                          % sample length after adjustment

%% plot data
% create x axis with dates
sampl = datetime('1947-Q4','InputFormat','yyyy-QQQ'):calquarters(1):datetime('2012-Q3','InputFormat','yyyy-QQQ');
figure('name','US Quarterly Inflation')
plot(sampl,y,'LineWidth',2);
title('Quarterly US Inflation');

%% set priors
% priors for theta ~ N(theta0,Sigma0)
theta0 = zeros(3,1); % prior mean for coefficients
Sigma0 = eye(3);     % prior variance for coefficients
invSigma0 = inv(Sigma0); % as we need the inverse of Sigma0 later on
% priors for precision (1/sigma_u^2) ~ G(s0,v0):
s0 = 1;              % prior shape parameter
v0 = 0.1;            % prior scale parameter

%% options for Gibbs sampler
R = 5000; % total number of Gibbs iterations
B = 4000; % number of burn-in iterations

%% initialize output matrices
out1 = zeros(3,R-B);  % coefficient draws
out2 = zeros(1,R-B);  % precision draws
sigmau2_j = 1;        % initialize first draw of sigma_u^2

%% Gibbs sampling
count = 1;
for j = 1:R
```

```
    % sample theta conditional on (1/sigma_u^2) from N(theta1,Sigma1)
    Sigma1 = inv(invSigma0 + (1/sigmau2_j)*(X'*X));            % conditional posterior variance of theta
    theta1 = Sigma1*(invSigma0*theta0 + (1/sigmau2_j)*(X'*y)); % conditional posterior mean of theta
    % check stability of draw (to avoid explosive AR process)
    is_stable = 0;
    while is_stable == 0
        theta_j = transpose(mvnrnd(theta1,Sigma1)); % take a draw from multivariate normal
        Acomp = [theta_j(2) theta_j(3); 1 0]; % companion matrix
        if max(abs(eig(Acomp))) < 1
            is_stable = 1; % AR model is stable if all the eigenvalues are less than or equal to 1 in absolute value
        end
    end

    % sample (1/sigma_u^2) conditional on theta from G(s1,v1)
    u = y-X*theta_j; % residuals conditional on theta(j)
    s1 = s0 + T;     % conditional posterior shape parameter
    v1 = v0 + u'*u;  % conditional posterior scale matrix
    sigma2inv_j = gamrnd(s1,1/v1,1,1); % take a draw from gamma distribution
    sigmau2_j = 1/sigma2inv_j; % we'll store the variance instead of the precision

    % save draws for inference if burn-in phase is passed
    if j > B
        out1(:,count) = theta_j;
        out2(:,count)= sigmau2_j;
        count = count+1;
    end
end

%% plot priors, histograms and kernel density estimate of marginal posteriors
x1 = -3:.1:3;
x2 = 0:.01:1;
c_prior = normpdf(x1,theta0(1),Sigma0(1,1));
theta1_prior = normpdf(x1,theta0(2),Sigma0(2,2));
theta2_prior = normpdf(x1,theta0(3),Sigma0(3,3));
sigmau2_prior = 1./gampdf(x2,s0,1/v0);

figure('name','Marginal Posterior Distributions','units','normalized','outerposition',[0 0.1 1 0.9]);

% Constant (c)
subplot(2,2,1)
histogram(out1(1,:),50,'Normalization','pdf', 'FaceColor', '#AEC7E8');
hold on;
[f, xi] = ksdensity(out1(1,:));
plot(xi, f, 'LineWidth', 2, 'Color', '#1F77B4'); % Kernel density estimate
plot(x1, c_prior, 'LineWidth', 2, 'Color', 'r');
axis tight
title('Constant (c)')
```

```
legend('Posterior Histogram', 'Posterior KDE', 'Prior', 'Location', 'Best')

% AR(1) coefficient (\phi_1)
subplot(2,2,2)
histogram(out1(2,:),50,'Normalization','pdf', 'FaceColor', '#AEC7E8');
hold on;
[f, xi] = ksdensity(out1(2,:));
plot(xi, f, 'LineWidth', 2, 'Color', '#1F77B4'); % Kernel density estimate
plot(x1, theta1_prior, 'LineWidth', 2, 'Color', 'r');
axis tight
title('AR(1) coefficient (\phi_1)')
legend('Posterior Histogram', 'Posterior KDE', 'Prior', 'Location', 'Best')

% AR(2) coefficient (\phi_2)
subplot(2,2,3)
histogram(out1(3,:),50,'Normalization','pdf', 'FaceColor', '#AEC7E8');
hold on;
[f, xi] = ksdensity(out1(3,:));
plot(xi, f, 'LineWidth', 2, 'Color', '#1F77B4'); % Kernel density estimate
plot(x1, theta2_prior, 'LineWidth', 2, 'Color', 'r');
axis tight
title('AR(2) coefficient (\phi_2)')
legend('Posterior Histogram', 'Posterior KDE', 'Prior', 'Location', 'Best')

% Error variance (\sigma_u^2)
subplot(2,2,4)
histogram(out2(1,:),50,'Normalization','pdf', 'FaceColor', '#AEC7E8');
hold on;
[f, xi] = ksdensity(out2(1,:), 'Support', 'positive');
plot(xi, f, 'LineWidth', 2, 'Color', '#1F77B4'); % Kernel density estimate
plot(x2, sigmau2_prior, 'LineWidth', 2, 'Color', 'r');
axis tight
title('Error variance (\sigma_u^2)')
legend('Posterior Histogram', 'Posterior KDE', 'Prior', 'Location', 'Best')

% Adjustments for overall figure aesthetics
set(gcf, 'Color', 'w'); % Set figure background to white
```