Introduction to Bayesian Econometrics

May 2020

Introduction to Bayesian Econometrics

- Consider the following regression Model

$$Y = X\beta + \varepsilon, \qquad \varepsilon \tilde{\ } N(0, \sigma^2 I)$$

Within the Bayesian framework the parameters $\theta = \{\beta, \sigma^2\}$ are treated as random variables. This parameters have probability distributions which reflect the knowledge of the researcher, before observing the sample on Y and X, about the parameters of the model.

- This probability distribution, denoted as $g(\theta)$ is called a prior distribution.

Once Y is observed, the researcher revises the distribution of the parameters by combining the prior distribution with the information obtained in the sample using Bayes Theorem.

We will define the following concepts:

- $f(Y|\theta)$ denotes the distribution of Y (from where we draw the data) given the parameters
- $h(\theta, Y)$ denotes the joint distribution of Y and θ.
- $f(Y)$ denotes the marginal distribution of Y
- $p(\theta|Y)$ denotes the posterior distribution of θ given Y

- Then we can write the joint distribution as

$$h(\theta, Y) = f(Y|\theta)g(\theta) = p(\theta|Y)f(Y)$$

which allows to write the posterior as

$$p(\theta|Y) = \frac{f(Y|\theta)g(\theta)}{f(Y)}$$

or equivalently

$$p(\theta|Y) \propto f(Y|\theta)g(\theta)$$

- Using the functional equivalence between $f(Y|\theta)$ and the likelihood $L(\theta|Y)$ we can express the posterior as

$$p(\theta|Y) \propto L(\theta|Y)g(\theta).$$

As we will see later on, the classical and the Bayesian approach are the same when the prior information is not available, that is, when the prior is diffuse or flat.

Which prior distributions should be used?

- There are groups of densities that may be easier to combine with the information of the likelihood. The natural conjugate priors are priors that once they are combined with the likelihood, they produce a posterior which has the same distribution as the prior.

- Example: Distribution of β assuming that σ^2 is known

Assume $\beta|\sigma^2 \tilde{} N(\beta_0, \Sigma_0)$ (a multivariate normal distribution) where β_0 and Σ_0 are known. Then the distribution can be written as

$$\begin{aligned} g(\beta|\sigma^2) &= (2\pi)^{-\frac{K}{2}} \, |\Sigma_0|^{.5} exp\left\{-\frac{1}{2}(\beta-\beta_0)'\Sigma_0^{-1}(\beta-\beta_0)\right\} \\ &\propto exp\left\{-\frac{1}{2}(\beta-\beta_0)'\Sigma_0^{-1}(\beta-\beta_0)\right\} \end{aligned}$$

where $(2\pi)^{-\frac{K}{2}} \, |\Sigma_0|^{.5}$ is a known constant.

The log likelihood

$$
\begin{aligned}
L(\beta|\sigma^2, Y) &= (2\pi\sigma^2)^{-\frac{T}{2}} exp\left\{-\frac{1}{2\sigma^2}(Y - X\beta)'(Y - X\beta)\right\} \\
&\propto exp\left\{-\frac{1}{2\sigma^2}(Y - X\beta)'(Y - X\beta)\right\}.
\end{aligned}
$$

where $(2\pi\sigma^2)^{-\frac{T}{2}}$ is a known constant.
Then the posterior is

$$
\begin{aligned}
p(\beta|\sigma^2, Y) &\propto g(\beta|\sigma^2)L(\beta|\sigma^2, Y) \\
&\propto exp\left\{\begin{array}{c} -\frac{1}{2}(\beta - \beta_0)'\Sigma_0^{-1}(\beta - \beta_0) \\ -\frac{1}{2\sigma^2}(Y - X\beta)'(Y - X\beta) \end{array}\right\}.
\end{aligned}
$$

Rearranging terms it can be shown that the posterior is also normal, therefore the normal density is the natural conjugate prior for β.

- Posterior distribution of β

It can be shown that $\beta|\sigma^2, Y \tilde{} N(\beta_1, \Sigma_1)$ where

$$
\begin{aligned}
\beta_1 &= (\Sigma_0^{-1} + \sigma^{-2}X'X)^{-1}(\Sigma_0^{-1}\beta_0 + \sigma^{-2}X'Y) \\
&= (\Sigma_0^{-1} + \sigma^{-2}X'X)^{-1}(\Sigma_0^{-1}\beta_0 + \sigma^{-2}X'X\widehat{\beta}) \\
\Sigma_1 &= (\Sigma_0^{-1} + \sigma^{-2}X'X)^{-1}
\end{aligned}
$$

From the previous equation we can see that the posterior mean of β is an average of β_0 and $\widehat{\beta}$.

- Example: Distribution of σ^2 assuming that β is known

The natural conjugate prior for σ^2 is the inverted Gamma distribution (the natural conjugate prior for $\frac{1}{\sigma^2}$ is the Gamma distribution)[1].
Prior of $\frac{1}{\sigma^2}|\beta \tilde{} \Gamma(\frac{v_0}{2}, \frac{\delta_0}{2})$ where v_0 and δ_0 are known.
Then

$$\begin{aligned} g(\frac{1}{\sigma^2}|\beta) &\propto (\frac{1}{\sigma^2})^{\frac{v_0}{2}-1}\exp(-\frac{\delta_0}{2\sigma^2}) \\ & \textit{and} \\ L(\frac{1}{\sigma^2}|\beta, Y) &= (2\pi\sigma^2)^{-\frac{T}{2}} exp\left\{-\frac{1}{2\sigma^2}(Y-X\beta)'(Y-X\beta)\right\} \\ &\propto (\sigma^2)^{-\frac{T}{2}} exp\left\{-\frac{1}{2\sigma^2}(Y-X\beta)'(Y-X\beta)\right\}. \end{aligned}$$

---

[1]
$$g(W) \propto W^{\frac{v}{2}-1}\exp(-\frac{W\delta}{2})$$

with $E(W) = \frac{v}{\delta}$ and $V(W) = 2\frac{v}{\delta^2}$.

- The posterior, $\frac{1}{\sigma^2}|\beta, Y\tilde{}\Gamma(\frac{v_1}{2}, \frac{\delta 1}{2})$, is therefore

$$\begin{aligned}
p(\frac{1}{\sigma^2}|\beta, Y) &\propto g(\frac{1}{\sigma^2}|\beta)\ L(\frac{1}{\sigma^2}|\beta, Y) \\
&\propto (\frac{1}{\sigma^2})^{\frac{v_0}{2}-1} \exp(-\frac{\delta_0}{2\sigma^2})(\sigma^2)^{-\frac{T}{2}} exp\left\{-\frac{1}{2\sigma^2}(Y - X\beta)'(Y - X\beta)\right\} \\
&= (\frac{1}{\sigma^2})^{\frac{v_0}{2}+\frac{T}{2}-1} exp\left\{-\frac{1}{2\sigma^2}(\delta_0 + (Y - X\beta)'(Y - X\beta))\right\} \\
&= (\frac{1}{\sigma^2})^{\frac{v_1}{2}-1} exp\left\{-\frac{\delta_1}{2\sigma^2}\right\} \\
where\ \delta_1 &= \delta_0 + (Y - X\beta)'(Y - X\beta) \text{ and } v_1 = v_0 + T.
\end{aligned}$$

GIbbs- Sampling in Econometrics

- Gibbs-sampling is a Markov chain Monte-Carlo method for approximating the joint and marginal distributions by sampling from conditional distributions.
- Consider the following joint density

$$f(z_1, z_2, ..., z_k)$$

and that we are interested in obtaining characteristics of the marginal density

$$f(z_t) = \int ..... \int f(z_1, z_2, ..., z_k) dz_1 dz_2, ..dz_{t-1} dz_{t+1} ...., dz_k$$

such as the mean or the variance. This exercise may be, when possible, very difficult to perform

- Gibbs sampling will allow me, if we are given the complete set of conditional distributions $f(z_t | z_1, z_2, ..z_{t-1}, z_{t+1} ...., z_k)$, to generate a sample of $z_1, z_2, ..., z_k$ without the need of knowing the joint $f(z_1, z_2, ..., z_k)$ or the marginals $f(z_t)$.

Methodology

Given arbitrary starting values $z_2^0, ..z_t^0, z_{t+1}^0 ...., z_k^0$

1) Draw z_1^1 from $f(z_1|z_2^0, ..z_t^0, z_{t+1}^0 ...., z_k^0)$
2) Then draw z_2^1 from $f(z_2|z_1^1, z_3^0, ..z_t^0, z_{t+1}^0 ...., z_k^0)$
3) Then draw z_3^1 from $f(z_3|z_1^1, z_2^1, z_4^0, z_5^0 ...., z_k^0)$

.

.

.

k) Finally draw z_k^1 from $f(z_k|z_1^1, z_2^1, z_3^1, z_4^1, z_5^1 ...., z_{k-1}^1)$

Then steps 1 to k can be iterated J times to get $z_1^j, z_2^j, ..z_t^j, z_{t+1}^j ...., z_k^j$, *for* $j = 1, 2, ..., J$.

- A crucial result in the literature is that of Geman and Geman (1984) that shows that the joint and marginal distributions of $z_1^j, z_2^j, ..z_t^j, z_{t+1}^j ...., z_k^j$ converge to the joint and marginal distributions of $z_1, z_2, ..z_t, z_{t+1} ...., z_k$ as $J \to \infty$.
- Consider $J = L + M$, then typically what is done is to use the first L simulations until the Gibbs sampler has converged and then use the remaining M simulations to approximate the empirical distribution.
- Convergence of the Gibbs Sampling

The Convergence of the Gibbs sampler is a very important issue which is somehow difficult to handle. For example it is usual to plot the posterior densities over the Gibbs iterations and look for little variation in the generated distribution over the replications.

Example: A univariate Autoregression

- Consider the following autoregressive model

$$y_t = \mu + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \phi_3 y_{t-3} + \phi_4 y_{t-4} + e_t, e_t \tilde{}\, i.i.d.N(0, \sigma^2),$$

where we assume that the roots of $(1 - \phi_1 L - \phi_2 L^2 - \phi_3 L^3 - \phi_4 L^4) = 0$ lie outside the unit circle.

- We can write the autoregressive model in matrix notation as

$$Y = X\beta + \varepsilon, \qquad \varepsilon \tilde{}\, N(0, \sigma^2 I)$$

where $\beta = [\mu,\ \phi_1,\ \phi_2,\ \phi_3,\ \phi_4]'$ and $X = [1,\ y_{t-1},\ y_{t-2},\ y_{t-3},\ y_{t-4}]$.

- Conditional Distributions of β given σ^2

The *prior distribution* of β is $\beta|\sigma^2 \tilde{} N(\beta_0, \Sigma_0)_{I(s(\phi))}$, where $I(s(\phi))$ is an indicator to denote that all the roots of $(1-\phi_1 L - \phi_2 L^2 - \phi_3 L^3 - \phi_4 L^4) = 0$ lie outside the unit circle.
The *posterior distribution* of β is $\beta|\sigma^2, Y \tilde{} N(\beta_1, \Sigma_1)_{I(s(\phi))}$,where

$$\begin{aligned}
\beta_1 &= (\Sigma_0^{-1} + \sigma^{-2} X'X)^{-1}(\Sigma_0^{-1}\beta_0 + \sigma^{-2}X'Y) \\
\Sigma_1 &= (\Sigma_0^{-1} + \sigma^{-2} X'X)^{-1}
\end{aligned}$$

- Conditional Distributions of σ^2 given β

The *Prior* distribution of $\sigma^2|\beta \tilde{} I\Gamma(\frac{v_0}{2}, \frac{\delta_0}{2})$ where v_0 and δ_0 are known and $I\Gamma$ denotes inverted Gamma distribution.
The *posterior distribution* of $\sigma^2|\beta \tilde{} I\Gamma(\frac{v_1}{2}, \frac{\delta_1}{2})$ where
$\delta_1 = \delta_0 + (Y - X\beta)'(Y - X\beta)$ and $v_1 = v_0 + T$.

The Gibbs Sampling procedure consists of the following steps

- Start the iteration of the Gibbs Sampling

-To start the iteration we use an arbitrary starting value $\sigma^2 = \left\{\sigma^2\right\}^0$
-Iterate the following steps $j = L + M$ times

- Conditional on $\sigma^2 = \left\{\sigma^2\right\}^{j-1}$, that is, the value generated in the previous iteration, we generate β^j from the posterior distribution of β.
- Conditional on $\beta = \beta^j$,that is, the value β generated in 1), we generate $\left\{\sigma^2\right\}^j$ from the posterior distribution of σ^2.
- Set $j = j + 1$.

- In generating β we employ rejection sampling to ensure that all the roots are outside the unit circle (we discard the draws that do not satisfy this condition).

As a result of this procedure we generate the following sets of values

$$\beta^1, \beta^2, \ldots\ldots\ldots\ldots, \beta^{L+M},$$
$$\left\{\sigma^2\right\}^1, \left\{\sigma^2\right\}^2, \ldots, \left\{\sigma^2\right\}^{L+M}.$$

- We discard the first L generated values to ensure convergence of the Gibbs-Sampler and then use the following M values to make inferences about β and σ^2. The remaining M values provide us with the Joint and the Marginal distribution.

Markov- Switching and Gibbs Sampling

- We have shown that when estimationg M-S models we treat parameters of the model depending on an unobserved sate variable. We typically estimate the models and make inferences on the unobserved variables conditional on the parameters (estimates) of the model.
- The Bayesian approach treats both the parameters of the model and the Markov switching variables as random variables. Then the inference about the states of the economy (denoted as $\widetilde{S}_T = S_1, S_2, ...S_T$) is based on the joint distribution of the sates and the parameters of the model.

- Consider the following model

$$
\begin{aligned}
y_t &= \mu_{S_t} + \varepsilon_t \qquad \varepsilon_t \tilde{} N(0, \sigma^2_{S_t}) \\
\mu_{S_t} &= \mu_0 + \mu_1 S_t, \qquad \mu_1 > 0 \\
\sigma^2_{S_t} &= \sigma^2_0(1 - S_t) + \sigma^2_1 S_t \\
&= \sigma^2_0(1 + h_1 S_t), \qquad h_1 > 0
\end{aligned}
$$

$$
\begin{aligned}
P(S_t &= 0 | S_{t-1} = 0) = q \\
P(S_t &= 1 | S_{t-1} = 1) = p
\end{aligned}
$$

- The Bayesian approach will entail the inference about $T+6$ random variables: $\{S_1, S_2, ..., S_T, \mu_0, \mu_1, \sigma_0^2, \sigma_1^2, p, q\}$.

We need to derive the joint posterior distribution

$$\begin{aligned} g(\widetilde{S}_T, \mu_0, \mu_1, \sigma_0^2, \sigma_1^2, p, q|\widetilde{y}_T) &= g(\mu_0, \mu_1, \sigma_0^2, \sigma_1^2, p, q|\widetilde{y}_T, \widetilde{S}_T).g(\widetilde{S}_T|\widetilde{y}_T) \\ &= g(\mu_0, \mu_1, \sigma_0^2, \sigma_1^2|\widetilde{y}_T, \widetilde{S}_T) \\ &\quad g(p, q|\widetilde{y}_T, \widetilde{S}_T).g(\widetilde{S}_T|\widetilde{y}_T) \\ &= g(\mu_0, \mu_1, \sigma_0^2, \sigma_1^2|\widetilde{y}_T, \widetilde{S}_T) \\ &\quad g(p, q|\widetilde{S}_T).g(\widetilde{S}_T|\widetilde{y}_T) \end{aligned}$$

We assume that conditional on $\widetilde{S}_T$, p and q are independent of both other parameters of the model and of the data. Notice that conditional on $\widetilde{S}_T$, the expression $y_t = \mu_{S_t} + \varepsilon_t$, $\varepsilon_t \tilde{} N(0, \sigma_{S_t}^2)$ is simply a regression model with a known dummy.

The Gibbs Sampling estimation procedure

Using arbitrary starting values we repeat the following steps.

1. Generate S_t from $g(S_t|\widetilde{S}_{\neq t}, \mu_0, \mu_1, \sigma_0^2, \sigma_1^2, p, q, \widetilde{y}_T)$, or generate the whole block $\widetilde{S}_T$ from $g(\widetilde{S}_T|\mu_0, \mu_1, \sigma_0^2, \sigma_1^2, p, q, \widetilde{y}_T)$
2. Generate the transition probabilities p and q from $g(p, q|\widetilde{S}_T)$.
3. Generate $\mu_0, \mu_1, \sigma_0^2, \sigma_1^2$, from $g(\mu_0, \mu_1, \sigma_0^2, \sigma_1^2|\widetilde{y}_T, \widetilde{S}_T)$.

- Step 1:Single move Gibbs Sampling, Generate S_t from $g(S_t|\widetilde{S}_{\neq t}, \mu_0, \mu_1, \sigma_0^2, \sigma_1^2, p, q, \widetilde{y}_T)$

Suppressing the conditioning in the parameters we can write

$$
\begin{aligned}
g(S_t|\widetilde{S}_{\neq t}, \widetilde{y}_T) &= g(\overbrace{S_t}^{A}|\overbrace{\widetilde{S}_{\neq t}, \widetilde{y}_t}^{C}, \overbrace{y_{t+1}, y_{t+2}, ..., y_T}^{B}) \\
&= \frac{g(\overbrace{S_t}^{A}, \overbrace{y_{t+1}, y_{t+2}, ..., y_T}^{B}|\overbrace{\widetilde{S}_{\neq t}, \widetilde{y}_t}^{C})}{g(\overbrace{y_{t+1}, y_{t+2}, ..., y_T}^{B}|\overbrace{\widetilde{S}_{\neq t}, \widetilde{y}_t}^{C})} \\
\text{since } g(AB|C) &\equiv g(A|BC)g(B|C) \\
&= \frac{g(S_t|\widetilde{S}_{\neq t}, \widetilde{y}_t)g(y_{t+1}, y_{t+2}, ..., y_T|\widetilde{S}_{\neq t}, \widetilde{y}_t)}{g(y_{t+1}, y_{t+2}, ..., y_T|\widetilde{S}_{\neq t}, \widetilde{y}_t)} \\
& \quad \text{Since conditional in the state, } S_t \textit{ and } y_{t+1}, .., y_T \\
& \quad \text{are independent.} \\
&= g(S_t|\widetilde{S}_{\neq t}, \widetilde{y}_t) \\
&= g(S_t|\widetilde{S}_{t-1}, S_{t+1}, ..., S_T, \widetilde{y}_{t-1}, y_t)
\end{aligned}
$$

$$
\begin{aligned}
g(\overbrace{S_t}^{A}|\overbrace{\widetilde{S}_{t-1},\widetilde{y}_{t-1}}^{C},\overbrace{S_{t+1},...,S_T,y_t}^{B}) &= \frac{g(\overbrace{S_t}^{A},\overbrace{S_{t+1},...,S_T,y_t}^{B}|\overbrace{\widetilde{S}_{t-1},\widetilde{y}_{t-1}}^{C})}{g(\overbrace{S_{t+1},...,S_T,y_t}^{B}|\overbrace{\widetilde{S}_{t-1},\widetilde{y}_{t-1}}^{C})} \\
&\propto g(S_t,S_{t+1},...,S_T,y_t|\widetilde{S}_{t-1},\widetilde{y}_{t-1})
\end{aligned}
$$

$$
\begin{aligned}
g(S_t,.,S_T,y_t|\widetilde{S}_{t-1},\widetilde{y}_{t-1}) &= g(S_t|\widetilde{S}_{t-1},\widetilde{y}_{t-1})g(S_{t+1},.,S_T,y_t|S_t,\widetilde{S}_{t-1},\widetilde{y}_{t-1}) \\
&= g(S_t|S_{t-1})g(S_{t+1},.,S_T,y_t|S_t,\widetilde{S}_{t-1},\widetilde{y}_{t-1}) \\
&\quad \text{Since is homogeneous Markov.}
\end{aligned}
$$

Notice that

$$
\begin{aligned}
g(\overbrace{y_t}^{A}, \overbrace{S_{t+1}, ., S_T}^{B} | \overbrace{S_t, \widetilde{S}_{t-1}, \widetilde{y}_{t-1}}^{C}) &= g(\overbrace{y_t}^{A} | \overbrace{S_t, \widetilde{S}_{t-1}, \widetilde{y}_{t-1}}^{C}) \\
&\quad g(\overbrace{S_{t+1}, ., S_T}^{B} | \overbrace{S_t, \widetilde{S}_{t-1}, \widetilde{y}_{t-1}}^{C} \overbrace{y_t}^{A}) \\
&= g(y_t | S_t) g(S_{t+1} | S_t, \widetilde{S}_{t-1}, \widetilde{y}_{t-1}, y_t) \\
&\quad g(S_{t+2}, ., S_T, | S_{t+1}, S_t, \widetilde{S}_{t-1}, \widetilde{y}_{t-1}, y_t) \\
&= g(y_t | S_t) g(S_{t+1} | S_t) g(S_{t+2}, ., S_T | S_{t+1}) \\
&\propto g(y_t | S_t) g(S_{t+1} | S_t)
\end{aligned}
$$

Then, we can write

$$g(S_t|\widetilde{S}_{\neq t}, \widetilde{y}_T) \propto g(S_t|S_{t-1})g(y_t|S_t)g(S_{t+1}|S_t)$$

where $g(S_t|S_{t-1})$ and $g(S_{t+1}|S_t)$ are given by the transition probabilities and

$$g(y_t|S_t) = \frac{1}{\sqrt{2\pi\sigma_{s_t}^2}} \exp\{-\frac{1}{2\sigma_{s_t}^2}(y_t - \mu_{s_t})^2\}$$

We can then calculate

$$P(S_t = j|\widetilde{S}_{\neq t}, \widetilde{y}_T) = \frac{g(S_t = j|\widetilde{S}_{\neq t}, \widetilde{y}_T)}{\sum_{j=0}^{1} g(S_t = j|\widetilde{S}_{\neq t}, \widetilde{y}_T)}$$

We generate S_t using a uniform distribution between 0 and 1. If the generated number is less or equal than $P(S_t = j|\widetilde{S}_{\neq t}, \widetilde{y}_T)$, we set the value of S_t to zero, otherwise we set the value equal to one.

- Step 1):Multimove Gibbs Sampling - Generate $\widetilde{S}_t$ from $g(\widetilde{S}_T|\mu_0, \mu_1, \sigma_0^2, \sigma_1^2, p, q, \widetilde{y}_T)$

Suppressing the conditioning in the parameters we can write

$$
\begin{aligned}
g(\widetilde{S}_T|\widetilde{y}_T) &= g(S_1, S_2, ..., S_T, |\widetilde{y}_T) \\
&= g(S_T, |\widetilde{y}_T)g(S_1, S_2, ..., S_{T-1}, |S_T, \widetilde{y}_T) \\
&= g(S_T, |\widetilde{y}_T)g(S_{T-1}, |S_T, \widetilde{y}_T)g(S_1, ., S_{T-2}, |S_{T-1}, S_T, \widetilde{y}_T) \\
&= g(S_T, |\widetilde{y}_T)g(S_{T-1}, |S_T, \widetilde{y}_T)g(S_{T-2}, |S_{T-1}, S_T, \widetilde{y}_T).. \\
& \quad ....g(S_1, |S_2, ., S_{T-2}, S_{T-1}, S_T, \widetilde{y}_T)... \\
(NB &: \quad g(S_{T-1}, |S_T, \widetilde{y}_T) = g(S_{T-1}, |S_T, \widetilde{y}_{T-1}) \\
& \quad \text{ since cond on } S_T,\ y_T \text{ adds no info.}) \\
&= g(S_T, |\widetilde{y}_T)g(S_{T-1}, |S_T, \widetilde{y}_{T-1}) \\
& \quad g(S_{T-2}, |S_{T-1}, \widetilde{y}_{T-2})..g(S_1, |S_2, y_1) \\
&= g(S_T, |\widetilde{y}_T) \prod_{t=1}^{T-1} g(S_t, |S_{t+1}, \widetilde{y}_t).
\end{aligned}
$$

- The derivation is based on the Markov property that states that to makes inference about S_t conditional on S_{t+1} the variables S_{t+2},....S_T, y_{t+1},....y_T have no information beyond that contained in S_{t+1}.
- Then we proceed in the following way: we first generate $\widetilde{S}_T$ conditional on $\widetilde{y}_T$ and then, for the other values of $t = T-1, t-2, ..., 1$, we generate S_t conditional on y_t and the generated $t+1$.

We can carry out this using the following steps:

Step 1 Run the Hamilton filter to get $g(S_t|\widetilde{y}_t)$. The last iteration of the filter provides $g(S_T|\widetilde{y}_T)$ that is used to generate S_T.

Step 2 Generate S_t conditional on S_{t+1} and $\widetilde{y}_t$, for $t = T-1, t-2, ..., 1$, form $g(S_t|S_{t+1}, \widetilde{y}_t)$ using the fact that

$$\begin{aligned} g(S_t|S_{t+1}, \widetilde{y}_t) &= \frac{g(S_t, S_{t+1}|\widetilde{y}_t)}{g(S_{t+1}|\widetilde{y}_t)} \\ &= \frac{g(S_{t+1}|S_t, \widetilde{y}_t).g(S_t|\widetilde{y}_t)}{g(S_{t+1}|\widetilde{y}_t)} \\ &= \frac{g(S_{t+1}|S_t).g(S_t|\widetilde{y}_t)}{g(S_{t+1}|\widetilde{y}_t)} \\ &\propto g(S_{t+1}|S_t).g(S_t|\widetilde{y}_t) \end{aligned}$$

Then we calculate

$$P(S_t = 1|S_{t+1}, \widetilde{y}_t) = \frac{g(S_{t+1}|S_t = 1)g(S_t = 1|\widetilde{y}_t)}{\sum_{j=0}^{1} g(S_{t+1}|S_t = j)g(S_t = j|\widetilde{y}_t)}$$

We generate S_t using a uniform distribution between 0 and 1. If the generated number is less or equal than $P(S_t = 1|S_{t+1}, \widetilde{y}_t)$, we set the value of S_t to zero, otherwise we set the value equal to one.

- Generating Transition Probabilities p and q, conditional on $\widetilde{S}_T$

Conditional on $\widetilde{S}_T$, p and q are independent of the data set $\widetilde{y}_T$, and the other parameters of the models[2].

- Prior Distribution

p˜$beta(u_{11}, u_{10})$,
q˜$beta(u_{00}, u_{01})$,
with $g(p,q) \propto p^{u_{11}-1}(1-p)^{u_{10}-1}q^{u_{00}-1}(1-q)^{u_{01}-1}$, where the $u\prime s$ are known hyper parameters of the priors.

- The likelihood function is given by

$L(p,q|\widetilde{S}_T) = p^{n_{11}}(1-p)^{n_{10}}q^{n_{00}}(1-q)^{n_{01}}$ where n_{ij} refers to the number of transitions from i to j which can be counted from $\widetilde{S}_T$.

$$\begin{aligned} g(z|\alpha_0,\alpha_1) &\propto z^{\alpha_0-1}(1-z)^{\alpha_1-1} for\ 0<z<1 \\ &= 0 \qquad\qquad otherwise, \end{aligned}$$

with $E(z) = \frac{\alpha_0}{\alpha_0+\alpha_1}$ and $Var(z) = \frac{\alpha_0\alpha_1}{(\alpha_0+\alpha_1)^2(\alpha_0+\alpha_1+1)}$

Posterior distribution

$$
\begin{aligned}
g(p,q|\widetilde{S}_T) &= g(p,q)L(p,q|\widetilde{S}_T) \\
&\propto p^{u_{11}-1}(1-p)^{u_{10}-1}q^{u_{00}-1}(1-q)^{u_{01}-1} \\
&\quad p^{n_{11}}(1-p)^{n_{10}}q^{n_{00}}(1-q)^{n_{01}} \\
&= p^{u_{11}+n_{11}-1}(1-p)^{u_{10}+n_{10}-1}q^{u_{00}+n_{00}-1}(1-q)^{u_{01}+n_{01}-1},
\end{aligned}
$$

then,
$p|\widetilde{S}_T\,\tilde{}\,beta(u_{11+}n_{11}, u_{10}+n_{10})$,
$q|\widetilde{S}_T\,\tilde{}\,beta(u_{00}+n_{00}, u_{01}+n_{01})$.

- Generating μ_0, μ_1,conditional on $\sigma_0^2, \sigma_1^2, \widetilde{y}_T$ and $\widetilde{S}_T$

Given

$$
\begin{aligned}
y_t &= \mu_0 + \mu_1 S_t + \varepsilon_t \qquad \varepsilon_t \tilde{} N(0, \sigma_{S_t}^2) \\
&\quad \text{we can write} \\
y_t^* &= \mu_0 x_{0t} + \mu_1 x_{1t} + v_t \qquad v_t \tilde{} N(0, 1) \\
\text{where } y_t^* &= \frac{y_t}{\sigma_{S_t}},\ x_{0t} = \frac{1}{\sigma_{S_t}} \text{ and } x_{1t} = \frac{S_t}{\sigma_{S_t}}
\end{aligned}
$$

- Prior Distribution: We can write the model in matrix notation as

$$Y = X\mu + V, \qquad V\tilde{}N(0, I)$$

then if we assume a normal prior $\mu|\sigma_0^2, \sigma_1^2\tilde{}N(b_0, B_0)$, where b_0, B_0 are given.

- Posterior distribution, $\mu|\sigma_0^2, \sigma_1^2, \widetilde{S}_T, \widetilde{y}_T\tilde{}N(b_1, B_1)$, where b_1, B_1 are

$$\begin{aligned} b_1 &= (B_0^{-1} + X'X)^{-1}(B_0^{-1}b_0 + X'Y) \\ B_1 &= (B_0^{-1} + X'X)^{-1} \end{aligned}$$

to constrain $\mu_1 > 0$, we discard the draws where this condition is not satisfied.

- Generating σ_0^2, σ_1^2,conditional on $\mu_0, \mu_1, \widetilde{y}_T$ and $\widetilde{S}_T$

Given

$$
\begin{aligned}
\sigma_{S_t}^2 &= \sigma_0^2(1 + h_1 S_t), \qquad h_1 > 0 \\
& \textit{where} \\
\sigma_1^2 &= \sigma_0^2(1 + h_1).
\end{aligned}
$$

we can first generate σ_0^2 conditional on h_1, and then generate $(1 + h_1)$ conditional on σ_0^2.

- Generating σ_0^2 conditional on h_1

We divide both sides of y_t by $\sqrt{(1+h_1 S_t)}$:

$$\begin{aligned} y_t^{**} &= \mu_0 x_{0t}^* + \mu_1 x_{1t}^* + v_t^* \qquad v_t^* \sim N(0, \sigma_0^2) \\ \text{where } y_t^{**} &= \frac{y_t}{\sqrt{(1+h_1 S_t)}},\ x_{0t}^* = \frac{1}{\sqrt{(1+h_1 S_t)}}, \\ x_{1t}^* &= \frac{S_t}{\sqrt{(1+h_1 S_t)}} \text{ and } v_t^* = \frac{\varepsilon_t}{\sqrt{(1+h_1 S_t)}} \end{aligned}$$

then:
The Prior distribution of $\sigma_0^2 | \mu_0, \mu_1, h_1 \sim I\Gamma(\frac{v_0}{2}, \frac{\delta_0}{2})$ where v_0 and δ_0 are known and $I\Gamma$ denotes inverted Gamma distribution.
The posterior distribution of $\sigma_0^2 | \mu_0, \mu_1, h_1, \widetilde{S}_T, \widetilde{y}_T \sim I\Gamma(\frac{v_1}{2}, \frac{\delta_1}{2})$ *where*

$$\delta_1 = \delta_0 + \sum_{t=1}^{T} (y_t^{**} - \mu_0 x_{0t}^* - \mu_1 x_{1t}^*)^2 \text{ and } v_1 = v_0 + T.$$

- Generating h_1 conditional on σ_0^2

We divide both sides of y_t by σ_0 :

$$\begin{aligned} y_t^{***} &= \mu_0 x_{0t}^{**} + \mu_1 x_{1t}^{**} + v_t^{**} \qquad v_t^{**} \tilde{} N(0, (1 + h_1 S_t)) \\ \text{where } y_t^{***} &= \frac{y_t}{\sigma_0},\ x_{0t}^{**} = \frac{1}{\sigma_0},\ x_{1t}^{**} = \frac{S_t}{\sigma_0} \text{ and } v_t^{**} = \frac{\varepsilon_t}{\sigma_0} \end{aligned}$$

then:

- The Prior distribution of $h_1|\mu_0, \mu_1, \sigma_0^2 \tilde{} I\Gamma(\frac{v_2}{2}, \frac{\delta_2}{2})$ where v_2 and δ_2 are known and $I\Gamma$ denotes inverted Gamma distribution.
- The posterior distribution of $h_1|\mu_0, \mu_1, \sigma_0^2, \widetilde{S}_T, \widetilde{y}_T \tilde{} I\Gamma(\frac{v_3}{2}, \frac{\delta_3}{2})$ *where* $\delta_3 = \delta_2 + \sum_{t=1}^{N_1} (y_t^{***} - \mu_0 x_{0t}^{**} - \mu_1 x_{1t}^{**})^2$ and $v_3 = v_2 + T$, where N_1 is the number of times $S_t = 1$.