EC 607: Bayesian Econometrics

Models of Parameter Change

Prof. Jeremy M. Piger

Department of Economics

University of Oregon

Last Revised: March 22, 2019

1 Models of Parameter Change

Models that allow for parameter change of unknown timing have become very popular tools. Here we will discuss Bayesian estimation of two such models: 1) Models with structural breaks that occur at unknown timing and 2) Markov-switching models.

2 Structural Break Models

- Suppose we have T observations on a variable y_t, collected in the vector $Y = (y_1, y_2, \ldots, y_T)'$. Suppose we specify a model for Y, with associated parameter vector γ_t. We index γ with a t subscript to allow for the possibility that γ changes over time. We will allow γ to change over time by assuming that there have been M structural changes in this parameter vector that occur at dates $\tau = \left(\tau^1, \tau^2, \ldots, \tau^M\right)'$. That is:

$$\gamma_t = \gamma^0 D_{0t} + \gamma^1 D_{1t} + \cdots + \gamma^M D_{Mt}$$

where:

$$D_{0t} = \begin{cases} 1 & \text{if } t < \tau^1 \\ 0 & \text{otherwise} \end{cases}$$

$$D_{it} = \begin{cases} 1 & \text{if } \tau^i \leq t < \tau^{i+1} \\ 0 & \text{otherwise} \end{cases} \quad i = 1, \ldots, M-1$$

$$D_{Mt} = \begin{cases} 1 & \text{if } t \geq \tau^M \\ 0 & \text{otherwise} \end{cases}$$

We will be interested in the case where τ is unobserved. The parameters of the model are then $\theta = (\gamma_0, \gamma_1, \ldots, \gamma_M)'$ and τ.

- **Likelihood Function**

  The likelihood function is written generically as:

$$p\left(Y|\theta,\tau\right)$$

  We will assume that this function can be calculated.

- **Prior**

  We will assume prior independence between θ and τ:

$$p\left(\theta,\tau\right) = p\left(\theta\right)\Pr\left(\tau\right)$$

  The prior for θ is written generically as:

$$p\left(\theta\right)$$

  For τ we need a prior distribution function that describes a random variable that can take on M values ranging from 2 to T. Here we will use a "flat" prior for this purpose:

$$\Pr\left(\tau\right) = \frac{1}{C}$$

  where C is the number of possible ways to choose M ordered values ranging from 2 to T:

$$C = \binom{T-1}{M}$$

  One could place additional prior information. For example you might want to restrict the break to not occur in the first or last $r\%$ of the sample. Or, you might want to enforce there is at least $p\%$ of the sample between breaks.

- **Posterior Density**

  The posterior density is:

  $$p\left(\theta,\tau|Y\right)$$

  This can be sampled using a Gibbs Sampler:

  1. Draw from $p\left(\theta|\tau,Y\right)$
  2. Draw from $\Pr\left(\tau|\theta,Y\right)$

- The first step is a draw from a model with the addition of dummy variables that capture structural breaks at known break dates. If you knew how to sample θ without structural breaks it will usually be simple to extend this sampler to include structural breaks at known break dates.

- Here we will focus on the second step.

  Drawing from $\Pr\left(\tau|\theta,Y\right)$

  Using Bayes Rule:

  $$\Pr\left(\tau|\theta,Y\right)\propto p\left(Y|\tau,\theta\right)Pr\left(\tau|\theta\right)$$

  Inserting our prior for τ, this can be written as:

  $$\Pr\left(\tau|\theta,Y\right)\propto p\left(Y|\tau,\theta\right)\frac{1}{\binom{T-1}{M}}$$

  So:

  $$\Pr\left(\tau|\theta,Y\right)\propto p\left(Y|\tau,\theta\right)$$

This is very simple - it says that the posterior probability that the break dates are at τ, conditional on θ, is proportional to the likelihood function evaluated at that value of τ.

Because this is a discrete distribution, it is simple to recover the summing constant as:

$$\frac{1}{\sum\limits_{\tau \in \Xi} p\left(Y|\tau, \theta\right)}$$

where Ξ represents the set of C possible values for τ. So:

$$Pr\left(\tau|\theta, Y\right) = \frac{p\left(Y|\tau, \theta\right)}{\sum\limits_{\tau \in \Xi} p\left(Y|\tau, \theta\right)}$$

- We can then sample τ from this discrete distribution
- The Bayesian approach to structural breaks is nice for a number of reasons. One in particular is that it provides a posterior distribution for τ, which gives us not only a point estimate but a measure of uncertainty about τ. The other is that we do marginal likelihood analysis to evaluate evidence for the number and type of structural breaks, even when competing models are non-nested. Classical measures of uncertainty about structural break dates and tests of structural breaks are often hard to compute.
- One potential problem with the above sampler is that C can be quite large. For example, if $T = 200$ and $M = 2$ then we have over 19,000 different combinations of break dates to check. The likelihood function must be computed for each of these every time a new τ is drawn. This can be very slow, although efficient coding can help tremendously. For example, using matrix algebra to avoid looping over the C combinations is very important.
- An alternative sampler was suggested by Wang and Zivot (2000, *JBES*). Here, the

Gibbs sampler is expanded so that each element of τ becomes a block of the sampler. So, the Gibbs Sampler becomes:

1. Draw from $p(\theta|\tau, Y)$
2. Draw from $\Pr\left(\tau^1|\tau^{\neq 1}, \theta, Y\right)$
3. Draw from $\Pr\left(\tau^2|\tau^{\neq 2}, \theta, Y\right)$
4. Draw from $\Pr\left(\tau^3|\tau^{\neq 3}, \theta, Y\right)$
5. .
6. .
7. Draw from $\Pr\left(\tau^M|\tau^{\neq M}, \theta, Y\right)$

- What is $\Pr\left(\tau^i|\tau^{\neq i}, \theta, Y\right)$? Using calculations similar to what we used above, we see:

$$\Pr\left(\tau^i|\tau^{\neq i}, \theta, Y\right) = \frac{p(Y|\tau, \theta)}{\sum\limits_{\tau\in\chi} p\left(Y|\tau^i, \tau^{\neq i}, \theta\right)}$$

  where χ is the set of J possible places to put τ^i, conditional on the placement of all the other break dates $\tau^{\neq i}$. J will be less than T. Thus, a draw of τ using the above procedure will involve less than $M * T$ evaluations of the likelihood function, which could be substantially less than C.

- The drawback of this approach is that by drawing each break date conditional on the others, this sampler will be more likely to get stuck on one set of break dates. Thus it will be less efficient. I would recommend using the approach of drawing τ jointly unless C becomes too large to handle in a reasonable amount of time.

- As a specific example of a structural break model, we will consider a linear regression model with structural breaks in intercept, slope and disturbance variance. This is written as:

$$Y = \mathrm{X}\beta_t + \epsilon,$$

where $Y = (y_1, y_2, \cdots, y_T)'$ represents T observations of a dependent variable, $\mathrm{X} = [X_1, X_2, \cdots, X_k]$ is an $N \times k$ matrix holding the T observations of the k dependent variables, and $\epsilon = (\epsilon_1, \epsilon_2, \ldots, \epsilon_T)'$, where:

$$\epsilon_t \sim N\left(0, h_t^{-1}\right)$$

Finally, for the structural break model:

$$\beta_t = \beta^0 D_{0t} + \beta^1 D_{1t} + \cdots + \beta^M D_{Mt}$$
$$h_t = h^0 D_{0t} + h^1 D_{1t} + \cdots + h^M D_{Mt}$$

where D_{it} is a function of the break dates $\tau = \left(\tau^1, \tau^2, \ldots, \tau^M\right)'$ as defined above.

- The parameters of the model are then $\beta = \left(\beta^0, \beta^1, \ldots, \beta^M\right)'$, $h = \left(h^0, h^1, \ldots, h^M\right)'$ and τ.
- The following matrix definitions will prove useful:

$$D_i = (D_{i1}, D_{i2}, \ldots, D_{iT})', \;\; i = 0, \ldots, M$$

$$\widetilde{X} = [(X_1 \cdot D_0)\;\; (X_2 \cdot D_0) \ldots (X_k \cdot D_0) \ldots (X_1 \cdot D_M)(X_2 \cdot D_M) \ldots (X_k \cdot D_M)]$$

where $\cdot$ indicates the Hadermand product (element by element multiplication).

$$H = (h_1, h_2, \ldots, h_T)'$$

$$\Sigma^{-1} = diag(H)$$

- **Priors**

$$p(\beta, h, \tau) = p(\beta) \prod_{i=1}^{M} p(h_i)\, p(\tau)$$

$$\beta \sim N(\mu, V)$$

$$h_i \sim \text{Gamma}(m, v)$$

$$\Pr(\tau) = \frac{1}{C}$$

$$C = \binom{T-1}{M}$$

where we are using the second formulation of the Gamma density described in the "Review of Important Probability Density Functions" notes.

- **Posterior Density**

The posterior density is:

$$p(\beta, h, \tau | Y, X)$$

This can be sampled using a Gibbs Sampler:

1. Draw from $p(\beta | h, \tau, Y, X)$

2. Draw from $p(h | \beta, \tau, Y, X)$

3. Draw from $\Pr(\tau|\beta, h, Y, X)$

- **Sampling** $p(\beta|h, \tau, Y, X)$:

  Using Bayes Rule we have:

$$
\begin{aligned}
p(\beta|h, \tau, Y, X) &\propto p(Y|\beta, h, \tau, X)\, p(\beta|h, \tau, X) \\
p(\beta|h, \tau, Y, X) &\propto p(Y|\beta, h, \tau, X)\, p(\beta)
\end{aligned}
$$

  where:

$$
p(Y|\beta, h, \tau, X) = (2\pi)^{(-T/2)} \prod_{t=1}^{T} h_t^{1/2} \exp\left[\left(Y - \widetilde{X}\beta\right)' \Sigma^{-1} \left(Y - \widetilde{X}\beta\right)\right]
$$

  Plugging in the equation for the prior and likelihood function, and doing some rearranging, gives us:

$$
\beta|h, Y, X \sim N\left(\bar{\mu}, \bar{V}^{-1}\right)
$$

  where:

$$
\begin{aligned}
\bar{V} &= V^{-1} + \widetilde{X}'\Sigma^{-1}\widetilde{X} \\
\bar{\mu} &= \bar{V}^{-1}\left(V^{-1}\mu + \widetilde{X}'\Sigma^{-1}Y\right)
\end{aligned}
$$

  We can then sample β from this multivariate Normal distribution.

- **Sampling** $p(h|\beta, \tau, Y, X)$:

  Using derivations similar to those in our discussion of the linear regression model, we can show that:

$$
p(h|\beta, \tau, Y, X) = p\left(h^0|\beta, \tau, Y, X\right) p\left(h^1|\beta, \tau, Y, X\right) \ldots p\left(h^M|\beta, \tau, Y, X\right)
$$

where:

$$h^i|\beta,\tau,Y,X \sim \text{Gamma}\left(\overline{m},\overline{v}\right)$$

where:

$$\overline{v} = T^i + v$$
$$\overline{m} = \frac{\overline{v}}{\left(Y^i - \widetilde{X}^i\beta\right)'\left(Y^i - \widetilde{X}^i\beta\right) + \frac{v}{m}}$$

where Y^i and $\widetilde{X}^i$ are the rows of Y and $\widetilde{X}$ corresponding to when $D_{it} = 1$ and T^i is the number of periods for which $D_{it} = 1$.

- We can then obtain a draw from $p\left(h|\beta,\tau,Y,X\right)$ by taking $M+1$ independent draws from $h^i|\beta,\tau,Y,X \sim \text{Gamma}\left(\overline{m},\overline{v}\right)$, $i = 0,\ldots,M$.
- **Sampling** $\Pr\left(\tau|\beta,h,Y,X\right)$:
- Following the discussion of the general structural break model, we have:

$$\Pr\left(\tau|\beta,h,Y,X\right) = \frac{p\left(Y|\tau,\beta,h,X\right)}{\sum_{\tau\in\Xi} p\left(Y|\tau,\beta,h,X\right)}$$

where, from our earlier discussion:

$$p\left(Y|\beta,h,\tau,X\right) = (2\pi)^{(-T/2)}\prod_{t=1}^{T} h_t^{1/2}\exp\left[\left(Y - \widetilde{X}\beta\right)'\Sigma^{-1}\left(Y - \widetilde{X}\beta\right)\right]$$

Thus, we can calculate the C discrete probabilities that make up $\Pr\left(\tau|\beta,h,Y,X\right)$. We can then sample from this distribution.

3 Exercise: Linear Regression Model with Two Structural Breaks

In the dropbox you will find a zipped collection of files called "Linear Regression with Structural Breaks.zip." This implements the Gibbs sampler for a 2-state Markov-switching AR(1) model. It also computes the marginal likelihood using the approach of Chib (1995). Work with these files to make sure you understand what they do. Pay particular attention to the marginal likelihood calculation, which uses a reduced Gibbs run. Suggested exercises include:

1. Add in some MCMC diagnostics to the program and assess the convergence of the sampler.

2. Play around with the true data generating process and see how the Bayesian estimation performs.

3. Put a restriction on the estimated model and use marginal likelihoods to evaluate the posterior probability of the unrestricted vs. restricted model.

4 Markov-Switching Model

- In the scale mixture model, we had a very particular example of a mixture of normals distribution. The disturbance term for a linear regression followed a mixture of normals, where the different normal distributions varied only in the variance of the density, and each disturbance term had its own mixture.

- We could also consider a more general mixture of normals, where both the mean and variance of each observations differs, and which mixture is relevant for each observation is unknown to the econometrician. A popular example of this is the **Markov-Switching Model**. In a (Gaussian) Markov-switching model, each observation comes from one of N different normal densities. We will consider the specific example of a Markov-switching regression model in which there are $N = 2$ regimes:

$$
\begin{aligned}
y_t &= x_t' \beta_{S_t} + \epsilon_t \qquad (1) \\
\epsilon_t &\sim i.i.d.N(0, h_{S_t}^{-1})
\end{aligned}
$$

where y_t is scalar, x_t is a $k \times 1$ vector of observed exogenous or predetermined explanatory variables, which may include lagged values of y_t, and $S_t \in \{0, 1\}$ is an integer valued variable indicating which of 2 regimes is active at time t. Regimes differ in both the intercept and slope parameters collected in β_{S_t}, as well as in the conditional variance parameter $h_{S_t}^{-1}$.

- Another, equivalent, way to write this model is to say that $Y = (y_1, y_2, \ldots, y_T)'$ is a mixture of two different normal densities:

  1. $N\left(x_t'\beta_0, h_0^{-1}\right)$
  2. $N\left(x_t'\beta_1, h_1^{-1}\right)$

- The regime indicator variable, S_t, tells us which of these two mixtures y_t comes from. In a Markov-switching model, we treat S_t as unobserved, but we assume that it follows an N-state Markov process with transition probabilities $p_{ji} = \Pr\left(S_t = j | S_{t-1} = i\right)$. For our specific example of a 2-state process, we have four transition probabilities, p_{00}, $p_{01} = 1 - p_{00}$, p_{11}, $p_{10} = 1 - p_{11}$

- The parameters of the model will be then be $\theta = (\beta, h, P)'$,

  where $\beta = (\beta_0, \beta_1)'$, $h = (h_0, h_1)'$, and $P = (p_{00}, p_{11})'$,

- **Priors**

- Our prior will take the form:

$$p(\beta, h, P) = p(\beta) \prod_{i=0}^{1} p(h_i) \, p(p_{00}) p(p_{11})$$

where:

$$\beta \sim N(\mu, V)$$

$$h_i \sim \text{Gamma}(m_i, v_i)$$

$$p_{00} \sim \text{Beta}\left(\alpha_1^0, \alpha_2^0\right)$$

$$p_{11} \sim \text{Beta}\left(\alpha_1^1, \alpha_2^1\right)$$

- **Posterior Density**

  The posterior density is:

  $p(\beta, h, P|Y, X)$

- where $Y = (y_1, y_2, \ldots, y_T)'$ and X is the $T \times k$ matrix with the k^{th} column holding the T observations for the k^{th} independent variable. This posterior can be sampled via a Gibbs Sampler with three blocks: β, h, P and $S = (S_1, S_2, \ldots, S_T)'$. Incorporating S as part of the sampler is an example of data augmentation. Specifically, we will sample iteratively from:

  1. $p(\beta|h, P, S, Y, X)$

  2. $p(h|\beta, P, S, Y, X)$

3. $p\left(P|\beta,h,S,Y,X\right)$

4. $p\left(S|\beta,h,P,Y,X\right)$

- **Sampling from** $p\left(\beta|h,P,S,Y,X\right)$
- $p\left(\beta|h,P,S,Y,X\right)$ will be similar to the structural break model discussed previously.
- Define:

$$\widetilde{X}=\left[\left(X\cdot\left(\iota_T-S\right)\right)\ \left(X\cdot S\right)\right]$$

where ι_T is a $T\times 1$ vector of ones, and $\cdot$ indicates the Hadermand product (element by element multiplication).

$$H=\left(h_0\cdot(1-S)+h_1\cdot S\right)$$

$$\Sigma^{-1}=diag(H)$$

Using similar calculations to previous models we have considered gives us:

$$\beta|h,P,S,Y,X\sim N\left(\bar{\mu},\bar{V}^{-1}\right)$$

where:

$$\begin{aligned}\bar{V}&=V^{-1}+\widetilde{X}'\Sigma^{-1}\widetilde{X}\\ \bar{\mu}&=\bar{V}^{-1}\left(V^{-1}\mu+\widetilde{X}'\Sigma^{-1}Y\right)\end{aligned}$$

We can then sample β from this multivariate Normal distribution.

- **Sampling** $p\left(h|\beta,P,S,Y,X\right)$:

Using derivations similar to those in our discussion of the linear regression model, we can show that:

$$p\left(h|\beta,P,S,Y,X\right)=p\left(h^{0}|\beta,P,S,Y,X\right)p\left(h^{1}|\beta,P,S,Y,X\right)$$

where:

$$h^{i}|\beta,P,S,Y,X\sim\text{Gamma}\left(\overline{m}_{i},\overline{v}_{i}\right)$$

where:

$$\begin{aligned}\overline{v}_{i}&=T^{i}+v_{i}\\ \overline{m}_{i}&=\frac{\overline{v}_{i}}{\left(Y^{i}-\widetilde{X}^{i}\beta\right)'\left(Y^{i}-\widetilde{X}^{i}\beta\right)+\frac{v_{i}}{m_{i}}}\end{aligned}$$

where Y^{i} and $\widetilde{X}^{i}$ are the rows of Y and $\widetilde{X}$ corresponding to when $S_{t}=i$ and T^{i} is the number of periods for which $S_{t}=i$.

- We can then obtain a draw from $p\left(h|\beta,P,S,Y,X\right)$ by taking 2 sequential independent draws from $h^{i}|\beta,P,S,Y,X\sim\text{Gamma}\left(\overline{m}_{i},\overline{v}_{i}\right),\ i=0,1$.

- **Sampling from** $p\left(P|\beta,h,S,Y,X\right)$

  Applying Bayes Rule:

$$p\left(P|\beta,h,S,Y,X\right)\propto p\left(Y,S|\beta,h,P,X\right)p\left(P|\beta,h,X\right)$$

Because of the independence of the prior density we have:

$$p\left(P|\beta,h,S,Y,X\right)\propto p\left(Y,S|\beta,h,P,X\right)p\left(p_{00}\right)p\left(p_{11}\right)$$

Applying the law of total probability:

$$p\left(P|\beta,h,S,Y,X\right)\propto p\left(Y|\beta,h,P,S,X\right)p\left(S|\beta,h,P,X\right)p\left(p_{00}\right)p\left(p_{11}\right)$$

Now, conditional on β, h, P, S and X, the density for Y does not depend on p. Thus:

$$p\left(P|\beta,h,S,Y,X\right)\propto p\left(S|\beta,h,P,X\right)p\left(p_{00}\right)p\left(p_{00}\right)$$

Next:

$$p\left(S|\beta,h,P,X\right)=p\left(S|P\right)$$

since S does not depend on the conditional mean or variance parameters of the regression model. The Markov property of S means we can factor this density as follows:

$$p\left(S|\beta,h,P,X\right)=p\left(S_1|P\right)\prod_{t=2}^{T}p\left(S_t|S_{t-1},P\right)$$

where:

$$p\left(S_t|S_{t-1},P\right)=p_{S_{t-1}S_t}$$

Define c_{ij} as the number of times that S_t switches from i to j over the sample period $t=2,\ldots,T$. Then:

$$p\left(S|\beta,h,P,X\right)=p\left(S_1|P\right)p_{00}^{c_{00}}\left(1-p_{00}\right)^{c_{01}}p_{11}^{c_{11}}\left(1-p_{11}\right)^{c_{10}}$$

Plugging in the equations for the Beta prior densities, we then have:

$$p\left(P|\beta,h,S,Y,X\right)\propto$$
$$p\left(S_1|P\right)p_{00}^{c_{00}}\left(1-p_{00}\right)^{c_{01}}p_{11}^{c_{11}}\left(1-p_{11}\right)^{c_{10}}\left(p_{00}^{\alpha_1^0-1}\left(1-p_{00}\right)^{\alpha_2^0-1}\right)\left(p_{11}^{\alpha_1^1-1}\left(1-p_{11}\right)^{\alpha_2^1-1}\right)$$

The term $p\left(S_1|P\right)$ makes this density non-standard, and it would require a Metropolis-Hastings step to sample it directly. However, the usual practice is to ignore this density of the initial observation (set it equal to one) which is in the spirit of conditional estimation often used with autoregressive processes. This will have little effect on results for a reasonably large sample size. If we do this, we have:

$$p\left(P|\beta, h, S, Y, X\right) \propto p_{00}^{c_{00}+\alpha_1^0-1}(1-p_{00})^{c_{01}+\alpha_2^0-1}p_{11}^{c_{11}+\alpha_1^1-1}(1-p_{11})^{c_{10}+\alpha_2^1-1}$$

- This is the product of the kernels for two beta densities, one describing p_{00} and one describing p_{11}. Thus, $p\left(p|\beta, h, S, Y, X\right)$ is given by:

$$p\left(P|\beta, h, S, Y, X\right) = p\left(p_{00}|\beta, h, S, Y, X\right)p\left(p_{11}|\beta, h, S, Y, X\right)$$

where:

$$p_{00}|\beta, h, S, Y, X \sim Beta(c_{00}+\alpha_1^0, c_{01}+\alpha_2^0)$$
$$p_{11}|\beta, h, S, Y, X \sim Beta(c_{11}+\alpha_1^1, c_{10}+\alpha_2^1)$$

- We can then obtain a draw from $p\left(P|\beta, h, S, Y, X\right)$ by taking 2 sequential independent draws from a $Beta(c_{00}+\alpha_1^0, c_{01}+\alpha_2^0)$ and a $Beta(c_{11}+\alpha_1^1, c_{10}+\alpha_2^1)$.

- **Sampling from** $\Pr\left(S|\beta, h, P, Y, X\right)$

Using the law of total probability, this density can be factored as follows:

$$\Pr(S|\beta,h,P,Y,X) = \Pr(S_T|\beta,h,P,Y,X)\prod_{t=1}^{T-1}\Pr(S_t|S_{t+1},S_{t+2},\ldots,S_T,\beta,h,P,Y,X)$$
$$\Pr(S|\beta,h,P,Y,X) = \Pr(S_T|\beta,h,P,Y,X)\prod_{t=1}^{T-1}\Pr(S_t|S_{t+1},\beta,h,P,Y,X)$$
$$\Pr(S|\beta,h,P,Y,X) = \Pr(S_T|\beta,h,P,Y,X)\prod_{t=1}^{T-1}\Pr(S_t|S_{t+1},\beta,h,P,Y_t,X_t)$$

where Y_t and X_t contain the elements of Y and X from through period t. The first equation follows from the law of total probability. The validity of moving to the second and third equation relies on the Markov property of S_t. Conditional on S_{t+1} there is no information about S_t contained in $S_{t+2},\ldots,S_T$ or in $y_{t+1},\ldots,y_T$.

- This factorization gives us an approach to draw from $\Pr(S|\beta,h,P,Y,X)$:

  1. Draw $S_T^{[g]}$ from $\Pr(S_T|\beta,h,P,Y,X)$. This draw can be generated by drawing $u \sim U(0,1)$ and setting $S_T^{[g]} = 1$ if $U < \Pr(S_T = 1|\beta,h,P,Y,X)$ and $S_T^{[g]} = 0$ otherwise.

  2. Draw $S_{T-1}^{[g]}$ from $\Pr\left(S_{T-1}|S_T^{[g]},\beta,h,P,Y_{T-1},X_{T-1}\right)$.

$\vdots$

  T. Draw $S_1^{[g]}$ from $\Pr\left(S_1|S_2^{[g]},\beta,h,P,Y_1,X_1\right)$.

  $S^{[g]} = \left(S_1^{[g]},S_2^{[g]},\ldots,S_T^{[g]}\right)'$ will then be a draw from $\Pr(S|\beta,h,P,Y,X)$.

- To implement this procedure, we need to compute the various probabilities in the steps above. As a first step to do this, we need to compute the "filtered" state probabilities, $\Pr(S_t = i|\beta,h,P,Y_t,X_t)$, $i = 0,1$, for $t = 1,\ldots,T$. This can be done using the filter given in Hamilton (1989, *Econometrica*). Numerous textbook treatments of this filter exist, with a particularly useful example being Kim and Nelson (1999). My survey paper on Markov-switching models also covers the Hamilton (1989) filter in detail. This is available here:

`http://pages.uoregon.edu/jpiger/research/published-papers/piger_2009_ecss.pdf`

- Once we have these filtered probabilities, we can then compute each of the probabilities above as follows:

  1. The last step of Hamilton's filter gives us $\Pr\left(S_T = i|\beta, h, P, Y, X\right)$, $i = 0, 1$.

  2. For t = T-1 through 1: Compute $\Pr\left(S_t = i|S_{t+1}^{[g]}, \beta, h, P, Y_t, X_t\right)$ as follows. Apply Bayes Rule:

$$\Pr\left(S_t = i|S_{t+1}^{[g]}, \beta, h, P, Y_t, X_t\right) \propto \Pr\left(S_{t+1}^{[g]}|S_t = i, \beta, h, P, Y_t, X_t\right) \Pr\left(S_t = i|\beta, h, P, Y_t, X_t\right)$$

  The first term on the right hand side is the transition probability, $p_{iS_{t+1}^{[g]}}$ while the second is the filtered state probability, which we have from the Hamilton (1989) filter. As this is a discrete distribution, we can then recover the probability as follows:

$$\Pr\left(S_t = i|S_{t+1}^{[g]}, \beta, h, P, Y_t, X_t\right) = \frac{p_{iS_{t+1}^{[g]}} \Pr\left(S_t = i|\beta, h, p, Y_t, X_t\right)}{\sum_{j=0}^{1} p_{jS_{t+1}^{[g]}} \Pr\left(S_t = j|\beta, h, p, Y_t, X_t\right)}$$

- **Normalization**: A final note about Markov-switching models. Like all models with mixture distributions where the assignment of mixtures to observations is unknown, there is a "labeling" problem in that we could switch the names of the regimes (0 to 1 and 1 to 0) as well as the parameters (e.g. β_0 to β_1) and the likelihood function would be unchanged. A normalization restriction is required to nail down the labeling of regimes. For example, you might restrict $h_1 < h_0$, which labels regime 1 as the "high variance" regime.

5 Exercise: Two State Markov Switching AR(1) Model

In the dropbox you will find a zipped collection of files called "Markov Switching Autoregression.zip." This implements the Gibbs sampler for a 2-state Markov-switching AR(1) model. Work with these files to make sure you understand what they do. Suggested exercises include:

1. Add in some MCMC diagnostics to the program and assess the convergence of the sampler.

2. Play around with the true data generating process and see how the Bayesian estimation performs.

3. Extra, Extra Credit: Instead of generated data, fit the model to some actual macroeconomic data (e.g. the growth rate of real GDP or payroll employment).