# BSTS Part 3 Script

**Slide 1**

Welcome back! We hope your journey so far has been rewarding. As we noted in the previous part, this stage of your journey is the most challenging. But worry not. For once you understand how the Bayesian philosophy can be applied to structural time series, you will be ready to apply it to data. (-> Slide 2)

**Slide 3**

In this part of you adventure, we will see how the Bayesian framework can be applied to structural time series. We first discuss Bayesian structural time series when we do not have a regression component. This includes:

(->) the prior distribution of our parameters (->) and how to obtain the posterior distribution.

We only briefly cover this as our main focus is the case where we do have a regression component (->). Again, we discuss

(->) the prior distribution (->) and obtaining the posterior distribution. (-> Slide 4)

**Slide 4**

Assuming we don't have a regression component in our model, for example, the local level model, then our parameters are the variances of the error terms.

(->) So, we would need to specify a distribution for say, $\sigma_\varepsilon^2$. To keep things simple, we won't discuss a particular prior here. Just remember the parameters in structural time series models are the error variances and we would need to specify a prior for each.

(->) The posterior distribution is obtained numerically.

(->) The Kalman filter and Kalman smoother are used,

(->) along with Markov Chain Monte Carlo sampling.

When we do have a regression component, we use what is called a *Spike and Slab Prior* (-> Slide 5). We will consider our regression coefficients fixed through time because otherwise (->) they could be added as an additional state component.

(->) The *Spike* in Spike and Slab regression is a prior that specifies a positive probabilty for each regression coefficient being equal to zero.

(->) The *Slab* is the prior for each of the regression coefficients that are not zero, and also the prior for the variance of the error term.

This should become more clear as we discuss each of these in more detail. (-> Slide 6)

**Slide 6**

To understand the Spike component, we will need a bit of set up notation. Suppose we have a vector $\gamma$ where each component $\gamma_k$ is one if $\beta_k \neq 0$ and is otherwise zero. Here, $\beta$ is the vector of regression coefficients.

(->) Further, we use $\beta_\gamma$ to denote the subset of the vector $\beta$ corresponding to the $\gamma_k$s that are equal to one. For example, if $\gamma_1$ and $\gamma_2$ are equal to one and all other components are zero, then $\beta_\gamma$ will be the vector of $\beta_1$ and $\beta_2$.

The prior distribution we use for $\gamma$ is an independent Bernoulli prior (->). This means that, for each component $\gamma_k$ of $\gamma$, $\pi_k$ is the probability that $\gamma_k$ is equal to one and thus, $\beta_k$ is included in the model.

While we could use a different prior, this prior provides a lot of flexibility. For example, if we are very certain $\beta_1$ should be in the model, we could specify $\pi_1$ to be 0.95. And actually we can specify an inclusion probability $\pi_k$ for each $\beta_k$. If we don't have any previous knowledge as to which regression coefficients should be included, we could set $\pi_k$ to $1/2$ for every $k$, which is the default in the bsts package.

Along with specifying our *spike* prior, we need to specify the *slab* prior. (-> Slide 7).

**Slide 7**

Again, we need a little notation to start with. For a symmetric matrix $\Omega^{-1}$, (->) we let $\Omega_\gamma^{-1}$ denote the sub matrix whose rows and columns correspond to the indices $k$ where $\gamma_k = 1$. This is the same idea as with $\beta_\gamma$.

The *slab* prior is shown here.

(->) It says that $\beta_\gamma$, given $\sigma_\varepsilon^2$ and $\gamma$ follows a normal distribution with mean vector $b_\gamma$ and covariance matrix $\sigma_\varepsilon^2 \Omega_\gamma$. As is often done in Bayesian statistics to ease calculations, we specify a distribution for $\frac{1}{\sigma_\varepsilon^2}$ instead of $\sigma_\varepsilon^2$. This distribution is Gamma with parameters $\frac{v}{2}$ and $\frac{ss}{2}$. The parameterization for the Gamma distribution here is the shape-rate parameterization.

For this prior, we need to specify $b_\gamma, \sigma_\varepsilon^2, \Omega_\gamma^{-1}, v$, and $ss$. But before we do we want to note the full prior. (-> Slide 8)