

# Gibbs' Sampler: Dynamic Linear Regression

Elliot Shannon

shann125@msu.edu

Michigan State University

## Setting

---

The setting of interest involves data generated from a dynamic linear model. Here, dynamic refers to the evolution of regression coefficients ( $\beta$ 's) over time. Rather than standard multiple linear regression, we have different regression coefficients for each time step, where coefficients at time  $t$  are centered on the coefficient values at time  $t - 1$ . Moreover, the covariates ( $\mathbf{X}$ ), may also change over time, hence we have a different design matrix for each time step. The proposed model is described in detail below.

## Model

---

For time points  $t = 1, \dots, T$ , let

- $\mathbf{Y}_t = \begin{pmatrix} y_{1,t} \\ \vdots \\ y_{N,t} \end{pmatrix} \in \mathbb{R}^N$  be the vector of responses.
- $\mathbf{X}_t = \begin{pmatrix} 1 & x_{1,1,t} & \dots & x_{1,P-1,t} \\ & & \ddots & \\ 1 & x_{N,1,t} & \dots & x_{N,P-1,t} \end{pmatrix} \in \mathbb{R}^{N \times P}$  be the design matrix of covariates.
- $\boldsymbol{\beta}_t = \begin{pmatrix} \beta_{1,t} \\ \vdots \\ \beta_{P,t} \end{pmatrix} \in \mathbb{R}^P$  be the vector of regression coefficients.
- $\boldsymbol{\varepsilon}_t = \begin{pmatrix} \varepsilon_{1,t} \\ \vdots \\ \varepsilon_{N,t} \end{pmatrix} \in \mathbb{R}^N$  be the vector of residual terms.

The proposed model is then

$$\begin{aligned} \mathbf{Y}_t &= \mathbf{X}_t \boldsymbol{\beta}_t + \boldsymbol{\varepsilon}_t, \quad \boldsymbol{\varepsilon}_t \sim MVN(\mathbf{0}, \sigma^2 \mathbf{I}_N), \quad t = 1, \dots, T \\ \boldsymbol{\beta}_t &= \boldsymbol{\beta}_{t-1} + \boldsymbol{\eta}_t, \quad \boldsymbol{\beta}_0 \sim MVN(\boldsymbol{\mu}_\beta, \boldsymbol{\Sigma}_\beta), \\ \boldsymbol{\eta}_t &\sim MVN(\mathbf{0}, \boldsymbol{\Sigma}_\eta), \quad \boldsymbol{\Sigma}_\eta \sim IW(\mathbf{H}, v), \quad \sigma^2 \sim IG(a, b) \end{aligned} \tag{1}$$

where  $\mathbf{0}$  is a vector of 0's.

Further,  $MVN(\boldsymbol{\mu}, \boldsymbol{\Sigma})$  is the multivariate normal distribution with mean vector  $\boldsymbol{\mu} \in \mathbb{R}^N$ , covariance matrix  $\boldsymbol{\Sigma} \in \mathbb{R}^{N \times N}$ , and pdf

$$p(\mathbf{x} \mid \boldsymbol{\mu}, \boldsymbol{\Sigma}) = (2\pi)^{-N/2} |\boldsymbol{\Sigma}|^{-1/2} \exp\left(\frac{-1}{2} (\mathbf{x} - \boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1} (\mathbf{x} - \boldsymbol{\mu})\right) \tag{2}$$

$IW(\mathbf{H}, v)$  is the Inverse Wishart distribution with (positive definite) scale matrix  $\mathbf{H} \in \mathbb{R}^{N \times N}$ , degrees of freedom  $v > N - 1$ , and pdf

$$p(\Sigma | \mathbf{H}, v) = \frac{|\mathbf{H}|^{v/2}}{2^{vN/2} \Gamma_N(\frac{v}{2})} |\Sigma|^{-(v+N+1)/2} \exp\left(-\frac{1}{2} \text{tr}(\mathbf{H}\Sigma^{-1})\right) \quad (3)$$

and  $IG(a, b)$  is the Inverse Gamma distribution with shape parameter  $a > 0$ , scale parameter  $b > 0$ , and pdf

$$p(\sigma^2 | a, b) = \frac{b^a}{\Gamma(a)} (\sigma^2)^{-a-1} \exp\left(\frac{-b}{\sigma^2}\right). \quad (4)$$

Given this completed model specification, we would like to conduct inference on the model parameters given data  $\mathbf{Y}_t$  and  $\mathbf{X}_t$ ,  $t = 1, \dots, T$ . These parameters include  $\beta_0, \beta_1, \dots, \beta_T, \Sigma_\eta, \sigma^2$ . We assume that prior parameters  $\mu_\beta, \Sigma_\beta, \mathbf{H}, v, a, b$  are given (specified by the user).

## Posterior

---

We know that the posterior distribution in Bayesian inference is proportional to the likelihood (data distribution) times the prior. Here, our likelihood is the distribution of the data  $\mathbf{Y}_t$ ,  $t = 1, \dots, T$ . Since the distribution of the residuals is  $MVN$ , we will have a  $MVN$  likelihood, written as

$$\prod_{t=1}^T MVN(\mathbf{Y}_t | \mathbf{X}_t \beta_t, \sigma^2 \mathbf{I}_N) \quad (5)$$

where  $\mathbf{I}_N$  is an  $N \times N$  identity matrix (diagonal matrix with diagonal elements equal to 1).

We will also have priors (some of which are induced) for each parameter for which we seek inference. Specifically, we have prior distributions of the form

$$\prod_{t=1}^T MVN(\beta_t | \beta_{t-1}, \Sigma_\eta) \times MVN(\beta_0 | \mu_\beta, \Sigma_\beta) \times IG(\sigma^2 | a, b) \times IW(\Sigma_\eta | \mathbf{H}, v) \quad (6)$$

Given (5) and (6), their product will be proportional to the posterior distribution, and is given as

$$\begin{aligned} \prod_{t=1}^T MVN(\mathbf{Y}_t | \mathbf{X}_t \beta_t, \sigma^2 \mathbf{I}_N) \times \prod_{t=1}^T MVN(\beta_t | \beta_{t-1}, \Sigma_\eta) \times MVN(\beta_0 | \mu_\beta, \Sigma_\beta) \\ \times IG(\sigma^2 | a, b) \times IW(\Sigma_\eta | \mathbf{H}, v) \end{aligned} \quad (7)$$

Importantly, (7) is only *proportional* to our desired posterior. Why is this okay? We are essentially ignoring the complex integral (normalizing constant) which typically appears in the denominator of Bayes theorem, which represents the marginal distribution of our complete data  $\mathbf{Y}$ . This normalizing constant is necessary to ensure that the posterior we derive is actually a proper probability distribution (integrates to 1). It is important, but not necessary for our purposes of deriving full conditionals for Gibbs sampling. Moving forward, we only work with this proportional posterior (7).

## Conditional Intuition

---

Recall that in the current model (1), we are interested in learning about parameters  $\beta_0, \beta_1, \dots, \beta_T, \Sigma_\eta, \sigma^2$ . These parameters each appear in our posterior distribution given in (7); some appear in several terms, others appear in fewer. The beauty of Gibbs sampling is that for each parameter of interest, we can derive its *full conditional distribution*, meaning we don't need to work with the entirety of (7), but rather just the specific terms in which the parameter of interest is found. This simplifies our lives greatly. Moreover, if we have been clever enough to use conjugate priors for each parameter, we can expect that our full conditional distributions will take nice forms. So nice in fact, that we can skip many tedious steps in our derivation to simply identify the defining features (parameters) of each distribution. The general steps are outlined below.

### General steps for deriving full conditional distributions

1. Identify each term in the posterior distribution in which the parameter of interest is found.
2. Isolate these terms and write them as a product of pdfs.
3. Combine like terms and ignore terms that do not involve the parameter of interest.
4. Recognize the simplified expression as a pdf of the parameter of interest and identify the distributional parameters. This is the full conditional distribution for the parameter of interest.
5. Move to the next parameter of interest and go back to step 1.

### A note on conjugacy

In Bayesian inference, conjugacy refers to the pairing of prior and likelihood distributions to achieve a specific posterior form. When a prior is conjugate, the resulting posterior distribution will have the same distributional form. This is immensely helpful when deriving full conditional distributions, especially in step 4 of the general steps given above. Some important examples of conjugacy are given below:

- Normal Likelihood  $\times$  Normal Prior  $\rightarrow$  Normal Posterior
- Normal Likelihood  $\times$  IG Prior  $\rightarrow$  IG Posterior
- Normal Likelihood  $\times$  IW Prior  $\rightarrow$  IW Posterior

## Incompleting the Square

---

In the case of Normal-Normal conjugacy (normal prior with normal likelihood), we expect the posterior distribution to also follow a normal distribution. The same is true with multivariate normal distributions, which we work with here. A handy trick for identifying this posterior distribution is coined “*incompleting the square*”, which comes from James Clark's book ‘Models for Ecological Data’. See Section 4.2 of his book for more details.

Following (2), we know that the kernel of the  $MVN$  distribution is of the form

$$p(\mathbf{x} \mid \boldsymbol{\mu}, \boldsymbol{\Sigma}) \propto \exp \left( \frac{-1}{2} (\mathbf{x} - \boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1} (\mathbf{x} - \boldsymbol{\mu}) \right)$$

where  $\propto$  mean “proportional to”. This structure gives us some insight into the placement of the parameters in the pdf. To see this further, we can focus on the terms inside the exponential, ignoring the  $\frac{-1}{2}$  coefficient. Expanding this term gives us:

$$(\mathbf{x} - \boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1} (\mathbf{x} - \boldsymbol{\mu}) = \mathbf{x}^\top \boldsymbol{\Sigma}^{-1} \mathbf{x} - \mathbf{x}^\top \boldsymbol{\Sigma}^{-1} \boldsymbol{\mu} - \boldsymbol{\mu}^\top \boldsymbol{\Sigma}^{-1} \mathbf{x} + \boldsymbol{\mu}^\top \boldsymbol{\Sigma}^{-1} \boldsymbol{\mu} \quad (8)$$

Now comes the trick. Writing the mean as a function of the covariance matrix, we can define

$$\begin{aligned} \boldsymbol{\mu} &= \mathbf{V}\mathbf{v} \\ \boldsymbol{\Sigma} &= \mathbf{V} \end{aligned}$$

This allows us to rewrite (8) as

$$\begin{aligned} \mathbf{x}^\top \mathbf{V}^{-1} \mathbf{x} - \mathbf{x}^\top \mathbf{V}^{-1} \mathbf{V}\mathbf{v} - \mathbf{v}^\top \mathbf{V}^\top \mathbf{V}^{-1} \mathbf{x} + \mathbf{v}^\top \mathbf{V}^\top \mathbf{V}^{-1} \mathbf{V}\mathbf{v} \\ = \mathbf{x}^\top \mathbf{V}^{-1} \mathbf{x} - \mathbf{x}^\top \mathbf{v} - \mathbf{v}^\top \mathbf{x} + \mathbf{v}^\top \mathbf{V}\mathbf{v} \\ = \mathbf{x}^\top \mathbf{V}^{-1} \mathbf{x} - \mathbf{x}^\top \mathbf{v} \dots \end{aligned}$$

This is super helpful, because when we are mashing together normal priors and likelihoods, we can combine like terms and look for  $\mathbf{V}^{-1}$  and  $\mathbf{v}$  to get the mean and variance of the full conditional *MVN*. They will always be in the same place!  $\mathbf{V}^{-1}$  will be sandwiched between the variable of interest (i.e. the  $\mathbf{x}^\top \mathbf{V}^{-1} \mathbf{x}$  term), and  $\mathbf{v}$  will be attached to the negative transpose of the variable of interest (i.e.  $-\mathbf{x}^\top \mathbf{v}$ ) (in this example,  $\mathbf{x}$  is the variable of interest). Once we identify  $\mathbf{V}^{-1}$  and  $\mathbf{v}$ , we can sample from the full conditional distribution as  $MVN(\mathbf{V}\mathbf{v}, \mathbf{V})$ .

## Working it out

---

We will now go through each of the parameters in model (1) and derive their full conditional distributions. Moving forward, we will use different colors to help keep track of different prior and likelihood terms as we work through the derivations. This will help us see where each component of the full conditional distribution is coming from.

## Update $\beta_0$

---

Looking at (7), we see that  $\beta_0$  appears in two terms, namely it's **own prior distribution**, and as the mean value of  $\beta_1$ 's **prior distribution**. The product of these distributions looks like

$$MVN(\beta_0 \mid \boldsymbol{\mu}_\beta, \boldsymbol{\Sigma}_\beta) \times MVN(\beta_1 \mid \beta_0, \boldsymbol{\Sigma}_\eta)$$

Since both distributions are normal, we know their product will also be normal, with mean  $\mathbf{V}\mathbf{v}$  and Covariance matrix  $\mathbf{V}$ . Based on what we know about incompleted the square, we will find  $\mathbf{V}$  and  $\mathbf{v}$  in the following terms of the product

$$\beta_0^\top \mathbf{V}^{-1} \beta_0 - \beta_0^\top \mathbf{v} \dots$$

So, all we have to do is multiply two multivariate normal pdfs (see equation 2), combine like terms in  $\beta_0$ , and identify  $\mathbf{V}^{-1}$  and  $\mathbf{v}$  from where they appear above. Since we are incompleted the square, we don't actually need to multiply the complete pdfs, just the exponential parts where  $\mathbf{V}$  and  $\mathbf{v}$  will appear. Ignoring the other components of the pdfs, we have

$$\begin{aligned}
& MVN(\beta_0 \mid \mu_\beta, \Sigma_\beta) \times MVN(\beta_1 \mid \beta_0, \Sigma_\eta) \\
& \propto \exp\left((\beta_0 - \mu_\beta)^\top \Sigma_\beta^{-1} (\beta_0 - \mu_\beta)\right) \times \exp\left((\beta_1 - \beta_0)^\top \Sigma_\eta^{-1} (\beta_1 - \beta_0)\right) \\
& = \exp\left((\beta_0 - \mu_\beta)^\top \Sigma_\beta^{-1} (\beta_0 - \mu_\beta) + (\beta_1 - \beta_0)^\top \Sigma_\eta^{-1} (\beta_1 - \beta_0)\right) \\
& = \exp\left(\beta_0^\top \Sigma_\beta^{-1} \beta_0 - \beta_0^\top \Sigma_\beta^{-1} \mu_\beta - \mu_\beta^\top \Sigma_\beta^{-1} \beta_0 + \mu_\beta^\top \Sigma_\beta^{-1} \mu_\beta \right. \\
& \quad \left. + \beta_1^\top \Sigma_\eta^{-1} \beta_1 - \beta_1^\top \Sigma_\eta^{-1} \beta_0 - \beta_0^\top \Sigma_\eta^{-1} \beta_1 + \beta_0^\top \Sigma_\eta^{-1} \beta_0\right)
\end{aligned}$$

Now we just need to pay attention to the terms of the form  $\beta_0^\top \mathbf{V}^{-1} \beta_0$  and  $-\beta_0^\top \mathbf{v}$  to identify  $\mathbf{V}^{-1}$  and  $\mathbf{v}$ . Combining like terms gives us

$$\propto \exp\left(\beta_0^\top (\Sigma_\beta^{-1} + \Sigma_\eta^{-1}) \beta_0 - \beta_0^\top (\Sigma_\beta^{-1} \mu_\beta + \Sigma_\eta^{-1} \beta_1) \dots\right)$$

So we have that

$$\begin{aligned}
\mathbf{V}^{-1} &= (\Sigma_\beta^{-1} + \Sigma_\eta^{-1}) \\
\mathbf{v} &= (\Sigma_\beta^{-1} \mu_\beta + \Sigma_\eta^{-1} \beta_1)
\end{aligned}$$

and we can sample  $\beta_0$  from it's full conditional distribution as  $MVN(\mathbf{V}\mathbf{v}, \mathbf{V})$ .

## Update $\beta_t \quad t = 1, \dots, (T - 1)$

---

Looking at (7), we see that  $\beta_t$  appears in 3 terms, namely it's **own prior distribution**, the **prior distribution of  $\beta_{t+1}$** , and in the **likelihood for  $\mathbf{Y}_t$** . The product of these distributions looks like

$$MVN(\mathbf{Y}_t \mid \mathbf{X}_t \beta_t, \sigma^2 \mathbf{I}_N) \times MVN(\beta_t \mid \beta_{t-1}, \Sigma_\eta) \times MVN(\beta_{t+1} \mid \beta_t, \Sigma_\eta)$$

Again, we will expand this product, only paying attention to the terms in the exponent so we can incomplete the square. We have

$$\begin{aligned}
& MVN(\mathbf{Y}_t \mid \mathbf{X}_t \beta_t, \sigma^2 \mathbf{I}_N) \times MVN(\beta_t \mid \beta_{t-1}, \Sigma_\eta) \times MVN(\beta_{t+1} \mid \beta_t, \Sigma_\eta) \\
& \propto \exp\left((\mathbf{Y}_t - \mathbf{X}_t \beta_t)^\top (\sigma^2 \mathbf{I}_N)^{-1} (\mathbf{Y}_t - \mathbf{X}_t \beta_t)\right) \times \exp\left((\beta_t - \beta_{t-1})^\top \Sigma_\eta^{-1} (\beta_t - \beta_{t-1})\right) \\
& \quad \times \exp\left((\beta_{t+1} - \beta_t)^\top \Sigma_\eta^{-1} (\beta_{t+1} - \beta_t)\right) \\
& = \exp\left((\mathbf{Y}_t - \mathbf{X}_t \beta_t)^\top (\sigma^2 \mathbf{I}_N)^{-1} (\mathbf{Y}_t - \mathbf{X}_t \beta_t) + (\beta_t - \beta_{t-1})^\top \Sigma_\eta^{-1} (\beta_t - \beta_{t-1}) \right. \\
& \quad \left. + (\beta_{t+1} - \beta_t)^\top \Sigma_\eta^{-1} (\beta_{t+1} - \beta_t)\right) \\
& = \exp\left(\mathbf{Y}_t^\top (\sigma^2 \mathbf{I}_N)^{-1} \mathbf{Y}_t - \mathbf{Y}_t^\top (\sigma^2 \mathbf{I}_N)^{-1} \mathbf{X}_t \beta_t - \beta_t^\top \mathbf{X}_t^\top (\sigma^2 \mathbf{I}_N)^{-1} \mathbf{Y}_t + \beta_t^\top \mathbf{X}_t^\top (\sigma^2 \mathbf{I}_N)^{-1} \mathbf{X}_t \beta_t \right. \\
& \quad + \beta_t^\top \Sigma_\eta^{-1} \beta_t - \beta_t^\top \Sigma_\eta^{-1} \beta_{t-1} - \beta_{t-1}^\top \Sigma_\eta^{-1} \beta_t + \beta_{t-1}^\top \Sigma_\eta^{-1} \beta_{t-1} \\
& \quad \left. + \beta_{t+1}^\top \Sigma_\eta^{-1} \beta_{t+1} - \beta_{t+1}^\top \Sigma_\eta^{-1} \beta_t - \beta_t^\top \Sigma_\eta^{-1} \beta_{t+1} + \beta_t^\top \Sigma_\eta^{-1} \beta_t\right)
\end{aligned}$$

Now we just need to pay attention to the terms of the form  $\beta_t^\top \mathbf{V}^{-1} \beta_t$  and  $-\beta_t^\top \mathbf{v}$  to identify  $\mathbf{V}^{-1}$  and  $\mathbf{v}$ . Combining like terms gives us

$$\propto \exp \left( \beta_t^\top (\mathbf{X}_t^\top (\sigma^2 \mathbf{I}_N)^{-1} \mathbf{X}_t + \Sigma_\eta^{-1} + \Sigma_\eta^{-1}) \beta_t^\top - \beta_0^\top (\mathbf{X}_t^\top (\sigma^2 \mathbf{I}_N)^{-1} \mathbf{Y}_t + \Sigma_\eta^{-1} \beta_{t-1} + \Sigma_\eta^{-1} \beta_{t+1}) \dots \right)$$

So we have that

$$\begin{aligned} \mathbf{V}^{-1} &= (\mathbf{X}_t^\top (\sigma^2 \mathbf{I}_N)^{-1} \mathbf{X}_t + \Sigma_\eta^{-1} + \Sigma_\eta^{-1}) \\ \mathbf{v} &= (\mathbf{X}_t^\top (\sigma^2 \mathbf{I}_N)^{-1} \mathbf{Y}_t + \Sigma_\eta^{-1} \beta_{t-1} + \Sigma_\eta^{-1} \beta_{t+1}) \end{aligned}$$

and we can sample  $\beta_t$  from it's full conditional distribution as  $MVN(\mathbf{Vv}, \mathbf{V})$ .

## Update $\beta_T$

---

To update  $\beta_T$ , we follow the same steps as for  $\beta_t$ ,  $t = 1, \dots, (T-1)$ , except the blue terms are no longer used, since there is no  $T+1$  time point. We will then have

$$\begin{aligned} \mathbf{V}^{-1} &= (\mathbf{X}_T^\top (\sigma^2 \mathbf{I}_N)^{-1} \mathbf{X}_T + \Sigma_\eta^{-1}) \\ \mathbf{v} &= (\mathbf{X}_T^\top (\sigma^2 \mathbf{I}_N)^{-1} \mathbf{Y}_T + \Sigma_\eta^{-1} \beta_{T-1}) \end{aligned}$$

and we can sample  $\beta_T$  from it's full conditional distribution as  $MVN(\mathbf{Vv}, \mathbf{V})$ .

## Update $\sigma^2$

---

Up until now, we have used normal likelihoods and priors to determine the normally distributed full conditional distributions for parameters of interest via completing the square. To update  $\sigma^2$ , we will have to use a different form of conjugacy; the  $IG$ - $MVN$  conjugacy. Here, the  $IG$  prior on  $\sigma^2$  times the  $MVN$  likelihood for  $\mathbf{Y}$  yields another  $IG$  full conditional distribution for  $\sigma^2$ . It is our job then to determine what these new  $IG$  parameters will be.

Looking at (7), we see that  $\sigma^2$  appears in  $T+1$  many terms, namely it's own prior distribution, and the likelihood for  $\mathbf{Y}_t$ ,  $t = 1, \dots, T$ . The product of these distributions looks like

$$\prod_{t=1}^T MVN(\mathbf{Y}_t | \mathbf{X}_t \beta_t, \sigma^2 \mathbf{I}_N) \times IG(\sigma^2 | a, b)$$

From our knowledge of conjugate distributions, we know that the product of the  $IG$  distribution with  $MVN$  distributions should also be  $IG$ . Recall the  $IG$  pdf defined in (4). We will multiply these pdfs together, rearrange terms, and hopefully recover another  $IG$  distribution with different parameter values. We have the following

$$\begin{aligned} & \prod_{t=1}^T MVN(\mathbf{Y}_t | \mathbf{X}_t \beta_t, \sigma^2 \mathbf{I}_N) \times IG(\sigma^2 | a, b) \\ &= \prod_{t=1}^T (2\pi)^{-N/2} |\sigma^2 \mathbf{I}_N|^{-1/2} \exp \left( \frac{-1}{2\sigma^2} (\mathbf{Y}_t - \mathbf{X}_t \beta_t)^\top (\mathbf{Y}_t - \mathbf{X}_t \beta_t) \right) \times \frac{b^a}{\Gamma(a)} (\sigma^2)^{-a-1} \exp \left( \frac{-b}{\sigma^2} \right) \\ &\propto (\sigma^2)^{\frac{-NT}{2} - a - 1} \exp \left( -\frac{\frac{1}{2} \sum_{t=1}^T (\mathbf{Y}_t - \mathbf{X}_t \beta_t)^\top (\mathbf{Y}_t - \mathbf{X}_t \beta_t) + b}{\sigma^2} \right) \end{aligned}$$

Here in the second step, we write the statement as  $\propto$  because we are only interested in the components of the product that will become part of the  $IG$  pdf. We also combined like terms, and we can now recognize the expression as an  $IG$  distribution for  $\sigma^2$ , with the following parameters

$$\begin{aligned}\tilde{a} &= a + \frac{NT}{2} \\ \tilde{b} &= b + \frac{1}{2} \sum_{t=1}^T (\mathbf{Y}_t - \mathbf{X}_t \beta_t)^\top (\mathbf{Y}_t - \mathbf{X}_t \beta_t)\end{aligned}$$

and we can sample from  $\sigma^2$ 's full conditional distribution as  $IG(\tilde{a}, \tilde{b})$ .

## Update $\Sigma_\eta$

---

In this final step, we derive the full conditional distribution for  $\Sigma_\eta$ . One important thing to realize here is that  $\Sigma_\eta$  is a *matrix*, not a scalar or vector as we have worked with before. A conjugate prior for a matrix-valued parameter to the multivariate normal is the Inverse-Wishart  $IW$  prior (3). Here we use this prior, which will yield a full conditional distribution for  $\Sigma_\eta$  that is also distributed as  $IW$ .

Looking at (7), we have  $\Sigma_\eta$  appearing in  $T + 1$  many terms, namely it's *own prior distribution*, and the *distribution for each  $\beta_t, t = 1, \dots, T$* . The product of these distributions looks like

$$\prod_{t=1}^T MVN(\beta_t \mid \beta_{t-1}, \Sigma_\eta) \times IW(\Sigma_\eta \mid \mathbf{H}, v)$$

Our goal now is to simplify the product of these pdfs, and get it in the form of a new  $IW$  distribution where we can identify new parameters for the full conditional distribution. The product of these pdfs will be

$$\begin{aligned}& \prod_{t=1}^T MVN(\beta_t \mid \beta_{t-1}, \Sigma_\eta) \times IW(\Sigma_\eta \mid \mathbf{H}, v) \\&= \prod_{t=1}^T (2\pi)^{-P/2} |\Sigma_\eta|^{-1/2} \exp\left(\frac{-1}{2} (\beta_t - \beta_{t-1})^\top \Sigma_\eta^{-1} (\beta_t - \beta_{t-1})\right) \\& \quad \times \frac{|\mathbf{H}|^{v/2}}{2^{vN/2} \Gamma_N\left(\frac{v}{2}\right)} |\Sigma_\eta|^{-(v+N+1)/2} \exp\left(-\frac{1}{2} \text{tr}(\mathbf{H} \Sigma_\eta^{-1})\right) \\& \propto |\Sigma_\eta|^{-(T+v+P+1)/2} \exp\left(\frac{-1}{2} \left[ \underbrace{\sum_{t=1}^T (\beta_t - \beta_{t-1})^\top \Sigma_\eta^{-1} (\beta_t - \beta_{t-1})}_I + \text{tr}(\mathbf{H} \Sigma_\eta^{-1}) \right]\right)\end{aligned}$$

We will now use a trick to massage the expression into the form of the  $IW$  distribution. Since  $I$  is evaluated as a scalar (sum of quadratic forms), we can take the *trace* of it without changing the expression (the trace of a scalar is the scalar itself). So we can write

$$\begin{aligned}
I &= \sum_{t=1}^T (\beta_t - \beta_{t-1})^\top \Sigma_\eta^{-1} (\beta_t - \beta_{t-1}) \\
&= \text{tr} \left( \sum_{t=1}^T (\beta_t - \beta_{t-1})^\top \Sigma_\eta^{-1} (\beta_t - \beta_{t-1}) \right)
\end{aligned}$$

We also know that the trace of a sum is equal to the sum of a trace, so we can move the trace operator inside the sum.

$$= \sum_{t=1}^T \text{tr} \left( (\beta_t - \beta_{t-1})^\top \Sigma_\eta^{-1} (\beta_t - \beta_{t-1}) \right)$$

Another useful identity of the trace operator is the following

$$\text{tr}(\mathbf{ABC}) = \text{tr}(\mathbf{BCA}) = \text{tr}(\mathbf{CAB})$$

Here, we can change the order of the product of matrices in a trace by moving the first matrix to the end of the product. Applying that above yields

$$= \sum_{t=1}^T \text{tr} \left( (\beta_t - \beta_{t-1})(\beta_t - \beta_{t-1})^\top \Sigma_\eta^{-1} \right)$$

This is the form we want for  $I$ , and plugging it back in gives

$$\begin{aligned}
& |\Sigma_\eta|^{-(T+v+P+1)/2} \exp \left( \frac{-1}{2} \left[ \sum_{t=1}^T \text{tr} \left( (\beta_t - \beta_{t-1})(\beta_t - \beta_{t-1})^\top \Sigma_\eta^{-1} \right) + \text{tr}(\mathbf{H} \Sigma_\eta^{-1}) \right] \right) \\
&= |\Sigma_\eta|^{-(T+v+P+1)/2} \exp \left( \frac{-1}{2} \text{tr} \left[ \left( \sum_{t=1}^T (\beta_t - \beta_{t-1})(\beta_t - \beta_{t-1})^\top + \mathbf{H} \right) \Sigma_\eta^{-1} \right] \right)
\end{aligned}$$

which is in the same form as (3), with parameter values

$$\begin{aligned}
\tilde{v} &= v + T \\
\tilde{\mathbf{H}} &= \mathbf{H} + \sum_{t=1}^T (\beta_t - \beta_{t-1})(\beta_t - \beta_{t-1})^\top
\end{aligned}$$

and we can update  $\Sigma_\eta$  from its full conditional distribution as  $IW(\tilde{\mathbf{H}}, \tilde{v})$ .