

# Chase Joyner

882 Homework 3

October 17, 2016

## Problem 1:

- (a) The idea of a power prior is to incorporate historical data into the choice of the prior distribution of  $\theta$  in the current study. To begin, let  $\pi_0(\theta|\cdot)$  denote the initial choice of the prior distribution for  $\theta$  before any historical data, denoted  $D_0$ , is observed. Then, the power prior distribution for  $\theta$  is an 'update' of the initial prior in the following manner:

$$\pi(\theta|D_0, \alpha_0) \propto L(\theta|D_0)^{\alpha_0} \pi_0(\theta|c_0),$$

where  $\alpha_0$  is a scalar prior parameter that controls the effect of the historical data and  $c_0$  is a specified hyperparameter for the initial prior. As a result of the meaning of  $\alpha_0$ , it is reasonable to take  $\alpha_0 \in [0, 1]$ . Notice that if  $\alpha_0 = 0$ , then the historical data has no effect and the initial prior becomes the power prior. On the other end, if  $\alpha_0 = 1$ , then the power prior corresponds to the posterior distribution from the previous study involving  $D_0$ . To complete the power prior specification, Ibrahim, J. and Chen, M. propose the joint power prior distribution

$$\pi(\theta, \alpha_0|D_0) \propto L(\theta|D_0)^{\alpha_0} \pi_0(\theta|c_0) \pi(\alpha_0|\gamma_0).$$

They also claim a natural choice for  $\pi(\alpha_0|\gamma_0)$  is a Beta distribution. To illustrate, consider

$$Y_i = \mathbf{X}_i' \boldsymbol{\beta} + \epsilon_i,$$

where  $\mathbf{X}_i = (1, x_{i1}, \dots, x_{ip})'$ ,  $\boldsymbol{\beta} = (\beta_0, \beta_1, \dots, \beta_p)'$ , and  $\epsilon_i \stackrel{iid}{\sim} N(0, \sigma^2)$ . We impose the initial prior distributions for the regressors  $\boldsymbol{\beta}$  and  $\sigma^2$  by

$$\pi_0(\boldsymbol{\beta}, \sigma^2) = \pi_0(\boldsymbol{\beta}|\sigma^2) \pi_0(\sigma^2).$$

Under the modeling assumptions, we have that  $\mathbf{Y}|\mathbf{X}, \boldsymbol{\beta}, \sigma^2 \sim N(\mathbf{X}\boldsymbol{\beta}, \sigma^2 \mathbf{I})$ , which specifies the likelihood function  $L(\boldsymbol{\beta}|D)$ . Next, we research historically observed data, i.e. response vector  $\mathbf{Y}_0$  and a matrix of covariates  $\mathbf{X}_0$ . Then, letting  $\alpha_0$  be some scalar, we have the joint power prior distribution

$$\pi(\boldsymbol{\beta}, \sigma^2|D_0) \propto \left( (\sigma^2)^{-\frac{n}{2}} \exp \left\{ -\frac{1}{2\sigma^2} (\mathbf{Y}_0 - \mathbf{X}_0 \boldsymbol{\beta})' (\mathbf{Y}_0 - \mathbf{X}_0 \boldsymbol{\beta}) \right\} \right)^{\alpha_0} \times \pi_0(\boldsymbol{\beta}|\sigma^2) \pi_0(\sigma^2).$$

Suppose that we specify the initial prior for  $\boldsymbol{\beta}|\sigma^2 \sim N(\boldsymbol{\mu}_0, \sigma^2 \mathbf{R})$  and the initial prior for  $\pi(\sigma^2) \propto \sigma^{-2}$ , i.e. Jeffrey's prior. Then, we find that the power prior under this full model

specification to be

$$\begin{aligned}\pi(\boldsymbol{\beta}, \sigma^2 | D_0) &\propto \left( (\sigma^2)^{-\frac{n}{2}} \exp \left\{ -\frac{1}{2\sigma^2} (\mathbf{Y}_0 - \mathbf{X}_0 \boldsymbol{\beta})' (\mathbf{Y}_0 - \mathbf{X}_0 \boldsymbol{\beta}) \right\} \right)^{\alpha_0} \\ &\quad \times (\sigma^2)^{-\frac{p}{2}} \exp \left\{ -\frac{1}{2\sigma^2} (\boldsymbol{\beta} - \boldsymbol{\mu}_0)' \mathbf{R}^{-1} (\boldsymbol{\beta} - \boldsymbol{\mu}_0) \right\} \times (\sigma^2)^{-1} \\ &= (\sigma^2)^{-\frac{n\alpha_0+p}{2}-1} \exp \left\{ -\frac{\alpha_0}{2\sigma^2} \left[ (\mathbf{Y}_0 - \mathbf{X}_0 \boldsymbol{\beta})' (\mathbf{Y}_0 - \mathbf{X}_0 \boldsymbol{\beta}) + (\boldsymbol{\beta} - \boldsymbol{\mu}_0)' (\alpha_0 \mathbf{R})^{-1} (\boldsymbol{\beta} - \boldsymbol{\mu}_0) \right] \right\}.\end{aligned}$$

Define  $\mathbf{R}_\alpha = \alpha_0 \mathbf{R}$  and working with the term inside the exponential, we find

$$\begin{aligned}&(\mathbf{Y}_0 - \mathbf{X}_0 \boldsymbol{\beta})' (\mathbf{Y}_0 - \mathbf{X}_0 \boldsymbol{\beta}) + (\boldsymbol{\beta} - \boldsymbol{\mu}_0)' \mathbf{R}_\alpha^{-1} (\boldsymbol{\beta} - \boldsymbol{\mu}_0) \\ &= \mathbf{Y}_0' \mathbf{Y}_0 - 2\boldsymbol{\beta}' \mathbf{X}_0' \mathbf{Y}_0 + \boldsymbol{\beta}' \mathbf{X}_0' \mathbf{X}_0 \boldsymbol{\beta} + \boldsymbol{\beta}' \mathbf{R}_\alpha^{-1} \boldsymbol{\beta} - 2\boldsymbol{\beta}' \mathbf{R}_\alpha^{-1} \boldsymbol{\mu}_0 + \boldsymbol{\mu}_0' \mathbf{R}_\alpha^{-1} \boldsymbol{\mu}_0 \\ &= \boldsymbol{\beta}' (\mathbf{X}_0' \mathbf{X}_0 + \mathbf{R}_\alpha^{-1}) \boldsymbol{\beta} - 2\boldsymbol{\beta}' (\mathbf{X}_0' \mathbf{Y}_0 + \mathbf{R}_\alpha^{-1} \boldsymbol{\mu}_0) + \mathbf{Y}_0' \mathbf{Y}_0 + \boldsymbol{\mu}_0' \mathbf{R}_\alpha^{-1} \boldsymbol{\mu}_0 \\ &= (\boldsymbol{\beta} - \boldsymbol{\mu})' \mathbf{C}^{-1} (\boldsymbol{\beta} - \boldsymbol{\mu}) - \boldsymbol{\mu}' \mathbf{C}^{-1} \boldsymbol{\mu} + \mathbf{Y}_0' \mathbf{Y}_0 + \boldsymbol{\mu}_0' \mathbf{R}_\alpha^{-1} \boldsymbol{\mu}_0\end{aligned}$$

where  $\boldsymbol{\mu} = (\mathbf{X}_0' \mathbf{X}_0 + \mathbf{R}_\alpha^{-1})^{-1} (\mathbf{X}_0' \mathbf{Y}_0 + \mathbf{R}_\alpha^{-1} \boldsymbol{\mu}_0)$  and  $\mathbf{C} = (\mathbf{X}_0' \mathbf{X}_0 + \mathbf{R}_\alpha^{-1})^{-1}$ . Therefore,

$$\begin{aligned}\pi(\boldsymbol{\beta}, \sigma^2 | D_0) &\propto (\sigma^2)^{-\frac{n\alpha_0+p}{2}-1} \exp \left\{ -\frac{\alpha_0}{2\sigma^2} \left[ (\mathbf{Y}_0 - \mathbf{X}_0 \boldsymbol{\beta})' (\mathbf{Y}_0 - \mathbf{X}_0 \boldsymbol{\beta}) + (\boldsymbol{\beta} - \boldsymbol{\mu}_0)' \mathbf{R}_\alpha^{-1} (\boldsymbol{\beta} - \boldsymbol{\mu}_0) \right] \right\} \\ &= (\sigma^2)^{-\frac{n\alpha_0+p}{2}-1} \exp \left\{ -\frac{\alpha_0}{2\sigma^2} \left[ (\boldsymbol{\beta} - \boldsymbol{\mu})' \mathbf{C}^{-1} (\boldsymbol{\beta} - \boldsymbol{\mu}) - \boldsymbol{\mu}' \mathbf{C}^{-1} \boldsymbol{\mu} + \mathbf{Y}_0' \mathbf{Y}_0 + \boldsymbol{\mu}_0' \mathbf{R}_\alpha^{-1} \boldsymbol{\mu}_0 \right] \right\} \\ &= (\sigma^2)^{-\frac{n\alpha_0+p}{2}-1} \exp \left\{ -\frac{\alpha_0}{2\sigma^2} (\boldsymbol{\beta} - \boldsymbol{\mu})' \mathbf{C}^{-1} (\boldsymbol{\beta} - \boldsymbol{\mu}) \right\} \\ &\quad \times \exp \left\{ -\frac{\alpha_0}{2\sigma^2} \left[ \mathbf{Y}_0' \mathbf{Y}_0 - \boldsymbol{\mu}' \mathbf{C}^{-1} \boldsymbol{\mu} + \boldsymbol{\mu}_0' \mathbf{R}_\alpha^{-1} \boldsymbol{\mu}_0 \right] \right\} \\ &= (\sigma^2)^{-\frac{p}{2}} \exp \left\{ -\frac{\alpha_0}{2\sigma^2} (\boldsymbol{\beta} - \boldsymbol{\mu})' \mathbf{C}^{-1} (\boldsymbol{\beta} - \boldsymbol{\mu}) \right\} \\ &\quad \times (\sigma^2)^{-\frac{n\alpha_0}{2}-1} \exp \left\{ -\frac{\alpha_0}{2\sigma^2} \left[ \mathbf{Y}_0' \mathbf{Y}_0 - \boldsymbol{\mu}' \mathbf{C}^{-1} \boldsymbol{\mu} + \boldsymbol{\mu}_0' \mathbf{R}_\alpha^{-1} \boldsymbol{\mu}_0 \right] \right\}\end{aligned}$$

From here, we see that the power prior distributions are

$$\boldsymbol{\beta} | \sigma^2, D_0 \sim N \left( \boldsymbol{\mu}, \frac{\sigma^2}{\alpha_0} \mathbf{C} \right) \quad \text{and} \quad \sigma^2 | D_0 \sim IG \left( \frac{n\alpha_0}{2}, \frac{\alpha_0}{2} \left[ \mathbf{Y}_0' \mathbf{Y}_0 - \boldsymbol{\mu}' \mathbf{C}^{-1} \boldsymbol{\mu} + \boldsymbol{\mu}_0' \mathbf{R}_\alpha^{-1} \boldsymbol{\mu}_0 \right] \right).$$

- (b) Another prior formulation to consider is what is called a G-prior, which is for the regression coefficients in multiple linear regression models. The G-prior for  $\boldsymbol{\beta}$  is a multivariate normal distribution, whose covariance matrix is proportional to the Fisher information matrix for  $\boldsymbol{\beta}$ . Another factor attached to this covariance matrix is a positive scalar constant  $g$ , which controls the amount of correlation structure put into the prior distribution through the Fisher information matrix. The Fisher information matrix for the multiple linear regression above is given by

$$\mathcal{I}(\boldsymbol{\beta}) = \frac{\mathbf{X}' \mathbf{X}}{\sigma^2}.$$

Therefore, the G-prior for  $\boldsymbol{\beta}$  is  $\boldsymbol{\beta} | \sigma^2 \sim MVN(\boldsymbol{\beta}_0, g\sigma^2(\mathbf{X}' \mathbf{X})^{-1})$ , where  $\boldsymbol{\beta}_0$  is a hyperparameter. To complete the prior specification for this model, one should place a prior distribution on  $\sigma^2$ , perhaps impose that, *a priori*,  $\sigma^2 \sim IG(a/2, b/2)$ .

- (c) In the code, there is a function called *normal.invgamma*. This function takes inputs of the observed data  $\mathbf{Y}$ , the covariates matrix  $\mathbf{X}$ , the prior parameters of  $\beta|\sigma^2$ , and the prior parameters of  $\sigma^2$ . Therefore, for any normal-inverse gamma model, one can use this function to model fit. It should be noted that the prior parameters for  $\sigma^2$  are assumed to be halved, i.e. if the prior distribution for  $\sigma^2$  is  $IG(a/2, b/2)$ , the function is looking for  $a, b$  as the prior parameters for  $\sigma^2$ .
- (d) The simulation study consisted of only 100 data sets for the sake of time. We generated the true data under a simulated set of covariates matrix. Then, for the historical data set, more data was randomly generated from the same method as the simulated observed data, just with a different sample size. We considered different values of  $\alpha_0$  and  $g$ . The results of the simulations can be found below for 3 different values of  $\alpha_0$  and  $g$ .

			Estimates		
$\alpha_0, g$	Parameter	True	Power Prior	G-prior	MLR
$\alpha_0 = 0.01$ $g = 5$	$\beta_0$	0.5	0.483	0.709	0.452
	$\beta_1$	1	0.969	1.144	0.973
	$\beta_2$	3	3.015	2.858	3.029
	$\sigma^2$	1	1.035	1.035	0.989
$\alpha_0 = 0.3$ $g = 1$	$\beta_0$	0.5	1.332	1.246	0.493
	$\beta_1$	1	0.874	1.498	0.996
	$\beta_2$	3	2.607	2.499	2.999
	$\sigma^2$	1	2.354	2.354	0.978
$\alpha_0 = 0.5$ $g = 0.1$	$\beta_0$	0.5	1.816	1.864	0.515
	$\beta_1$	1	0.789	1.906	0.973
	$\beta_2$	3	2.427	2.091	3.002
	$\sigma^2$	1	3.101	3.101	0.997

The results show that as  $\alpha_0$  decreases, the estimates get better. This is because we are generating historical data from the same model, but is different than the observed data, and decreasing  $\alpha_0$  results in the historical data set having less of an effect on the inference. However, if the historical data set is similar to the observed data, then increasing  $\alpha_0$  results in inference that is similar to the multiple linear regression inference. Conversely, as  $g$  increases in value, the results get better to a certain extent. The reasoning is that under the G-prior, we are injecting the information explained by the covariates into the model. Lastly, to obtain better results than in the chart above, one should place a prior on  $\alpha_0$  and  $g$ , and perform updates on these as well. Consequently, placing informative priors can be beneficial if one knows where the true parameters live. However, noninformative priors or even flat priors can achieve very similar results with a lot less work.

## Problem 2:

- (a) Suppose that  $\mathbf{b} \sim \text{CAR}(\sigma^2\tau^2, \rho)$ ; that is

$$\mathbf{b} \sim N(0, \sigma^2\tau^2(\mathbf{D} - \rho\mathbf{W})^{-1}).$$

This density takes the form of

$$\pi(\mathbf{b}) \propto (\sigma^2\tau^2)^{-\frac{n}{2}} \exp \left\{ -\frac{1}{2\sigma^2\tau^2} \mathbf{b}'(\mathbf{D} - \rho\mathbf{W})\mathbf{b} \right\}.$$

We are interested in the conditional distribution  $b_i|\mathbf{b}_{(-i)}$ . Notice that

$$\begin{aligned} \mathbf{b}'(\mathbf{D} - \rho\mathbf{W})\mathbf{b} &= \mathbf{b}' \begin{bmatrix} D_{11} - \rho W_{11} & -\rho W_{12} & \dots & -\rho W_{1n} \\ -\rho W_{21} & D_{22} - \rho W_{22} & \dots & -\rho W_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ -\rho W_{n1} & -\rho W_{n2} & \dots & D_{nn} - \rho W_{nn} \end{bmatrix} \mathbf{b} \\ &= \mathbf{b}' \begin{bmatrix} b_1(D_{11} - \rho W_{11}) - \rho \sum_{k \neq 1} b_k W_{1k} \\ b_2(D_{22} - \rho W_{22}) - \rho \sum_{k \neq 2} b_k W_{2k} \\ \vdots \\ b_n(D_{nn} - \rho W_{nn}) - \rho \sum_{k \neq n} b_k W_{nk} \end{bmatrix} \\ &= \sum_{i=1}^n b_i^2(D_{ii} - \rho W_{ii}) - \rho \sum_{i=1}^n b_i \sum_{k \neq i} b_k W_{ik}. \end{aligned}$$

Then, removing all the terms that do not depend on  $b_i$ , and noting that  $W_{ii} = 0$  for all  $i$ ,

$$b_i^2(D_{ii} - \rho W_{ii}) - \rho b_i \sum_{k \neq i} b_k W_{ik} - \rho \sum_{k \neq i} b_k \sum_{j \neq k} b_j W_{jk} \equiv b_i^2 \sum_{k=1}^n W_{ik} - 2\rho b_i \sum_{k=1}^n b_k W_{ik}$$

where we get the equivalency by dropping all terms not dealing with  $b_i$  and noting that in the summation where  $j \neq k$ , we will capture more  $b_i$  terms to give double the first summation and the last summation is because  $W_{ii} = 0$ . Thus, the conditional distribution  $b_i|\mathbf{b}_{(-i)}$  is

$$b_i|\mathbf{b}_{(-i)} \sim N \left( \frac{\rho \sum_{k=1}^n b_k W_{ik}}{\sum_{k=1}^n W_{ik}}, \frac{\sigma^2\tau^2}{\sum_{k=1}^n W_{ik}} \right).$$

- (b) Now we find the full-conditional distributions of  $\beta_0|\text{else}$ ,  $\sigma^{-2}|\text{else}$ ,  $\tau^{-2}|\text{else}$ ,  $\mathbf{b}|\text{else}$ , and  $b_i|\text{else}$ . First, we see that the joint posterior distribution is

$$\begin{aligned} \pi(\mathbf{b}, \beta_0, \sigma^2, \tau^2 | \mathbf{Y}) &\propto (\sigma^2)^{-\frac{n}{2}} \exp \left\{ -\frac{1}{2\sigma^2} (\mathbf{Y} - \beta_0 - \mathbf{b})'(\mathbf{Y} - \beta_0 - \mathbf{b}) \right\} \\ &\times (\sigma^2\tau^2)^{-\frac{n}{2}} \exp \left\{ -\frac{1}{2\sigma^2\tau^2} \mathbf{b}'(\mathbf{D} - \rho\mathbf{W})\mathbf{b} \right\} \\ &\times (\sigma^{-2})^{a_\sigma-1} \exp \{ -\sigma^{-2}b_\sigma \} \times (\tau^{-2})^{a_\tau-1} \exp \{ -\tau^{-2}b_\tau \}. \end{aligned}$$

Therefore, we find that

$$\begin{aligned}
\pi(\beta_0|\text{else}) &\propto \exp \left\{ -\frac{1}{2\sigma^2} (\mathbf{Y} - \beta_0 - \mathbf{b})' (\mathbf{Y} - \beta_0 - \mathbf{b}) \right\} \\
&= \exp \left\{ -\frac{1}{2\sigma^2} \sum_{i=1}^n (Y_i - \beta_0 - b_i)^2 \right\} \\
&= \exp \left\{ -\frac{1}{2\sigma^2} \left( n\beta_0^2 - 2\beta_0 \sum_{i=1}^n (Y_i - b_i) \right) \right\}.
\end{aligned}$$

From here, we have the full-conditional distribution of  $\beta_0$  is

$$\beta_0 | \mathbf{Y}, \mathbf{b}, \sigma^2 \sim N \left( \sum_{i=1}^n (Y_i - b_i) / n, \sigma^2 / n \right).$$

Next, we have

$$\pi(\sigma^{-2}|\text{else}) \propto (\sigma^{-2})^{n+a_\sigma-1} \exp \left\{ -\frac{1}{2\sigma^2} \left[ (\mathbf{Y} - \beta_0 - \mathbf{b})' (\mathbf{Y} - \beta_0 - \mathbf{b}) + \tau^{-2} \mathbf{b}' (\mathbf{D} - \rho \mathbf{W}) \mathbf{b} + b_\sigma \right] \right\}.$$

Therefore, the full-conditional distribution for  $\sigma^{-2}$  is given by

$$\sigma^{-2} | \mathbf{Y}, \mathbf{b}, \beta_0, \tau^{-2} \sim \text{Gamma} \left( n + a_\sigma, b_\sigma + \frac{1}{2} \left[ (\mathbf{Y} - \beta_0 - \mathbf{b})' (\mathbf{Y} - \beta_0 - \mathbf{b}) + \tau^{-2} \mathbf{b}' (\mathbf{D} - \rho \mathbf{W}) \mathbf{b} \right] \right).$$

Very similarly, we have that the full-conditional for  $\tau^{-2}$  is

$$\tau^{-2} | \mathbf{Y}, \mathbf{b}, \beta_0, \sigma^{-2} \sim \text{Gamma} \left( \frac{n}{2} + a_\tau, b_\tau + \frac{\sigma^{-2}}{2} \mathbf{b}' (\mathbf{D} - \rho \mathbf{W}) \mathbf{b} \right).$$

Now, for the full-conditional distribution of  $\mathbf{b}$ , first see that

$$\pi(\mathbf{b}|\text{else}) \propto \exp \left\{ -\frac{1}{2\sigma^2} \left[ (\mathbf{Y} - \beta_0 - \mathbf{b})' (\mathbf{Y} - \beta_0 - \mathbf{b}) + \tau^{-2} \mathbf{b}' (\mathbf{D} - \rho \mathbf{W}) \mathbf{b} \right] \right\}.$$

Manipulating the term in the exponential, we find

$$\begin{aligned}
&(\mathbf{Y} - \beta_0 - \mathbf{b})' (\mathbf{Y} - \beta_0 - \mathbf{b}) + \tau^{-2} \mathbf{b}' (\mathbf{D} - \rho \mathbf{W}) \mathbf{b} \\
&= (\mathbf{Y} - \beta_0)' (\mathbf{Y} - \beta_0) - 2\mathbf{b}' (\mathbf{Y} - \beta_0) + \mathbf{b}' \mathbf{I} \mathbf{b} + \tau^{-2} \mathbf{b}' (\mathbf{D} - \rho \mathbf{W}) \mathbf{b} \\
&\equiv \mathbf{b}' (\mathbf{I} + \tau^{-2} (\mathbf{D} - \rho \mathbf{W})) \mathbf{b} - 2\mathbf{b}' (\mathbf{Y} - \beta_0).
\end{aligned}$$

From here, we have that the full-conditional distribution for  $\mathbf{b}$  is

$$\mathbf{b} | \mathbf{Y}, \beta_0, \sigma^2, \tau^2 \sim N \left( (\mathbf{I} + \tau^{-2} (\mathbf{D} - \rho \mathbf{W}))^{-1} (\mathbf{Y} - \beta_0), \sigma^2 (\mathbf{I} + \tau^{-2} (\mathbf{D} - \rho \mathbf{W}))^{-1} \right).$$

Lastly, we find the full-conditional distribution of  $b_i$ . Denote  $\mu_i = \frac{\rho \sum_{k=1}^n W_{ki} b_k}{\sum_{k=1}^n W_{ki}}$ . Then,

$$\begin{aligned}
\pi(b_i|\text{else}) &\propto \exp \left\{ -\frac{1}{2\sigma^2} (Y_i - \beta_0 - b_i)^2 \right\} \exp \left\{ -\frac{\sum_{k=1}^n W_{ki}}{2\sigma^2 \tau^2} (b_i - \mu_i)^2 \right\} \\
&= \exp \left\{ -\frac{1}{2\sigma^2} \left[ (Y_i - \beta_0 - b_i)^2 + W_\tau (b_i - \mu_i)^2 \right] \right\}
\end{aligned}$$

where  $W_\tau = \frac{1}{\tau^2} \sum_{k=1}^n W_{ki}$ . Working with the term in the exponential, we find that

$$\begin{aligned} (Y_i - \beta_0 - b_i)^2 + W_\tau(b_i - \mu_i)^2 &= (Y_i - \beta_0)^2 - 2b_i(Y_i - \beta_0) + b_i^2 + W_\tau b_i^2 - 2b_i W_\tau \mu_i + W_\tau \mu_i^2 \\ &\equiv b_i^2(1 + W_\tau) - 2b_i(Y_i - \beta_0 + W_\tau \mu_i). \end{aligned}$$

Therefore, the full-conditional distribution of  $b_i$  is

$$b_i | Y_i, \mathbf{b}_{(-i)}, \beta_0, \sigma^2, \tau^2 \sim N \left( \frac{(Y_i - \beta_0 + W_\tau \mu_i)}{1 + W_\tau}, \frac{\sigma^2}{1 + W_\tau} \right).$$

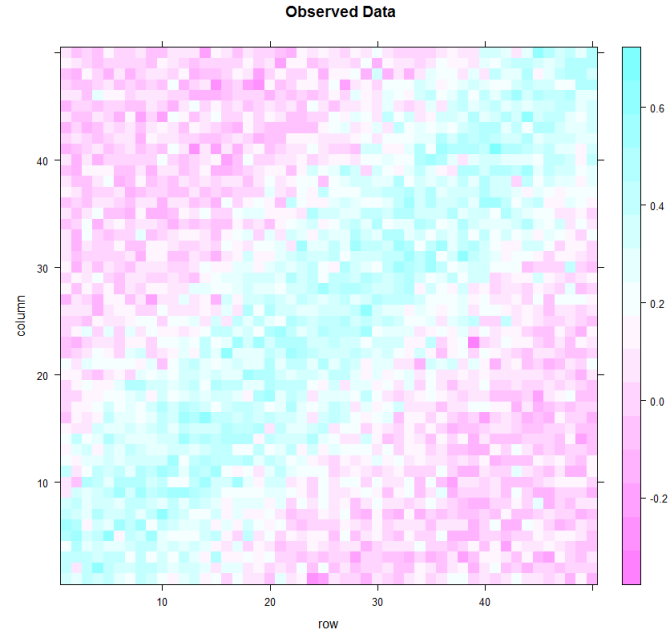
- (c) We will implement a Gibbs sampler to perform inference on the parameters of the model. The first Gibbs sampler will update the spatial random effects simultaneously, and the second will update them one at a time. For the first algorithm, we perform the following algorithm:

1. Initialize  $\mathbf{b}^{(0)}, \beta_0^{(0)}, \sigma^{2(0)}, \tau^{2(0)}$  and set  $t = 1$ .
2. Sample from the full-conditionals as follows:
  - (2.1)  $\mathbf{b}^{(t+1)} \sim \mathbf{b} | \mathbf{Y}, \beta_0^{(t)}, \sigma^{2(t)}, \tau^{2(t)}$
  - (2.2)  $\beta_0^{(t+1)} \sim \beta_0 | \mathbf{Y}, \mathbf{b}^{(t+1)}, \sigma^{2(t)}$
  - (2.3)  $\sigma^{-2(t+1)} \sim \sigma^{-2} | \mathbf{Y}, \mathbf{b}^{(t+1)}, \beta_0^{(t+1)}, \tau^{-2(t)}$
  - (2.4)  $\tau^{-2(t+1)} \sim \tau^{-2} | \mathbf{Y}, \mathbf{b}^{(t+1)}, \beta_0^{(t+1)}, \sigma^{-2(t+1)}$
3. Increment  $t$  by one and repeat step 2 for a large number of iterations.

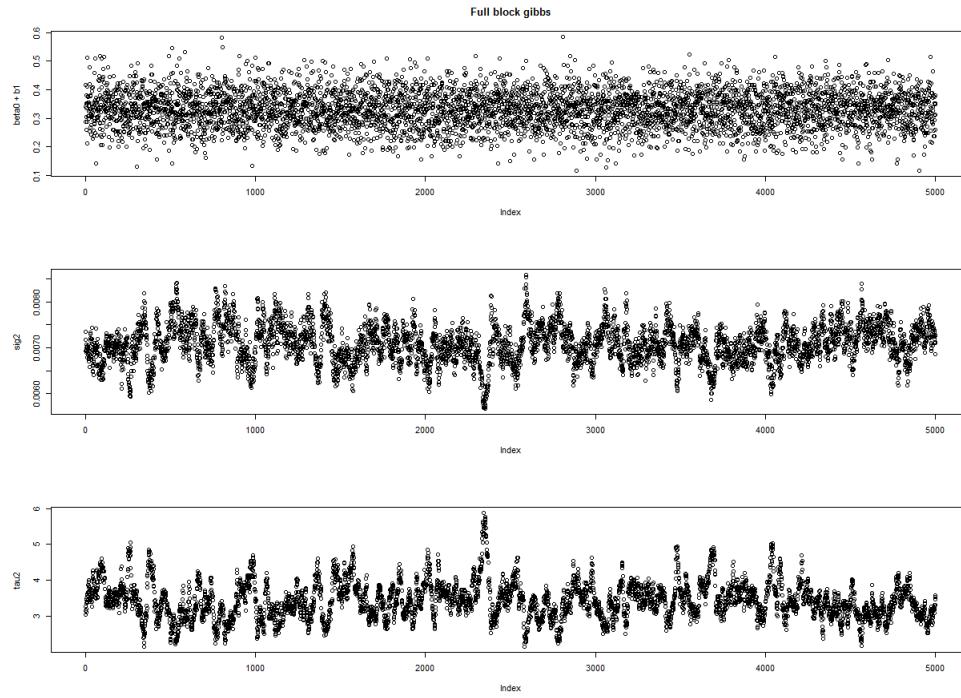
For the second Gibbs sampler, we will update the spatial random effects one at a time. The algorithm for this is as follows:

1. Initialize  $\mathbf{b}^{(0)} = (b_1^{(0)}, \dots, b_n^{(0)})', \beta_0^{(0)}, \sigma^{2(0)}, \tau^{2(0)}$  and set  $t = 1$ .
2. Sample from the full-conditionals as follows:
  - (2.1) For  $i = 1, \dots, n$ , sample  $b_i^{(t+1)} \sim b_i | Y_i, (b_1^{(t+1)}, \dots, b_{i-1}^{(t+1)}, b_{i+1}^{(t)}, \dots, b_n^{(t)})', \beta_0^{(t)}, \sigma^{2(t)}, \tau^{2(t)}$
  - (2.2)  $\beta_0^{(t+1)} \sim \beta_0 | \mathbf{Y}, \mathbf{b}^{(t+1)}, \sigma^{2(t)}$
  - (2.3)  $\sigma^{-2(t+1)} \sim \sigma^{-2} | \mathbf{Y}, \mathbf{b}^{(t+1)}, \beta_0^{(t+1)}, \tau^{-2(t)}$
  - (2.4)  $\tau^{-2(t+1)} \sim \tau^{-2} | \mathbf{Y}, \mathbf{b}^{(t+1)}, \beta_0^{(t+1)}, \sigma^{-2(t+1)}$
3. Increment  $t$  by one and repeat step 2 for a large number of iterations.

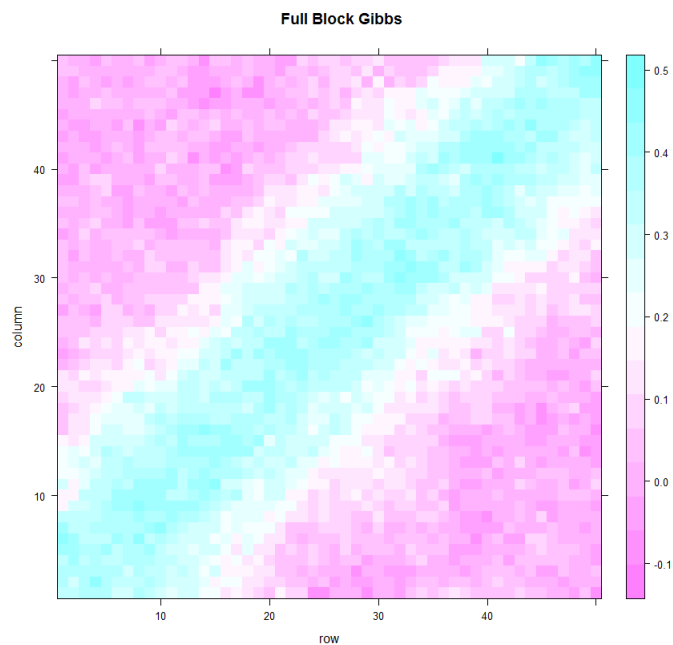
(d) We now implement the algorithms in part (c) to the observed data found below:



For both algorithms, we set  $\rho = 1$ . The first algorithm, which updates the spatial random effects simultaneously has the following trace plots:

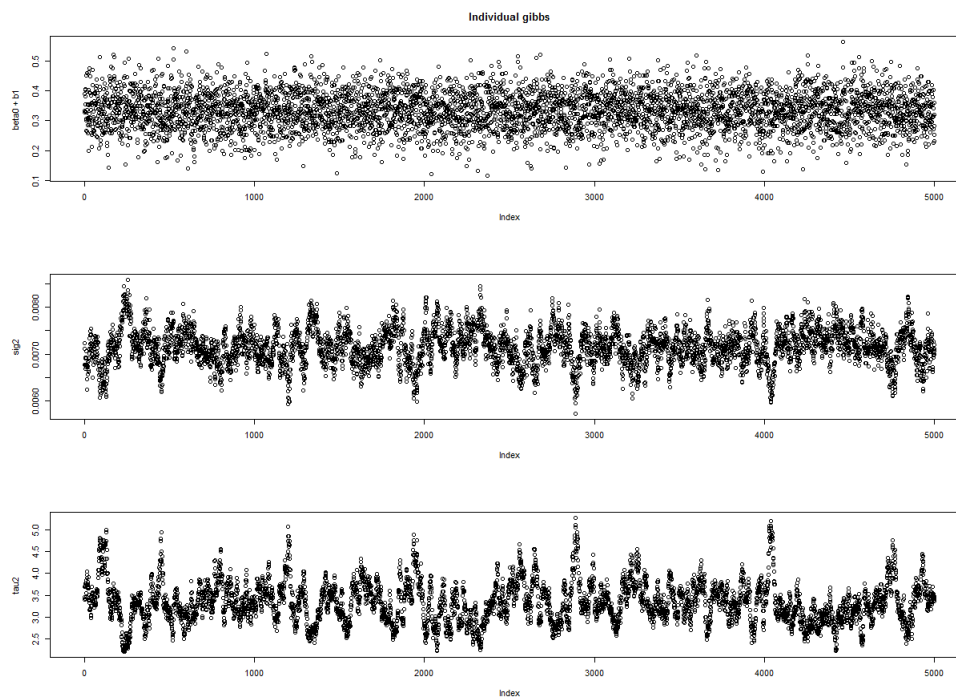


These trace plots were constructed by throwing away the first 5000 iterates and only considered the last 5000 iterates since it appears convergence for  $\sigma^2$  and  $\tau^2$  has taken place by then. Based on the mean estimates of these estimates, we have the estimated values from the model fitted under the full block gibbs sampler to be:

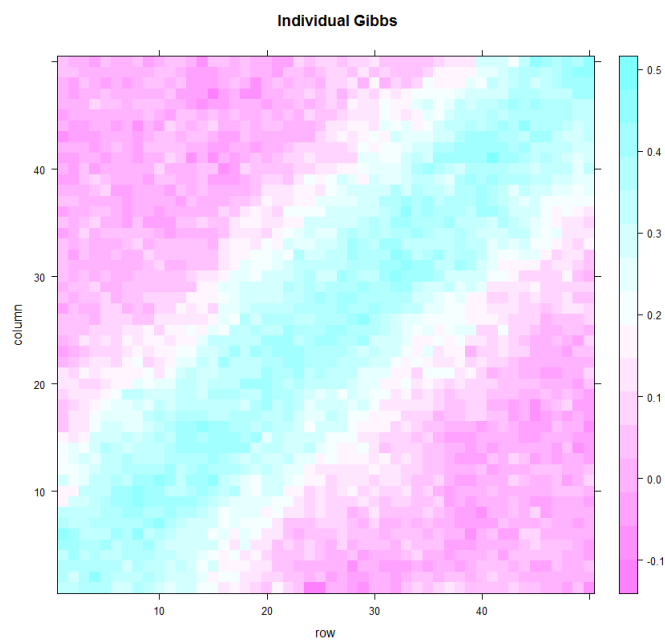




Next, the second algorithm produces the following trace plots:



where the same stories are being told. However, the runtime for this algorithm was much slower than the full block gibbs sampler. The estimated values for the response variable is:



## APPENDIX

```
#####
#####
#####
#####           Chase Joyner
#####           882 Homework 3
#####
#####
#####
```

```
#####
## Problem 1 ##
```

```
## Part c ##
```

```
normal.invgamma = function(Y, X, mu0, R0, a, b){
```

```
  ## Preliminaries ##
```

```
  n = length(Y)
```

```
  p = dim(X)[2]
```

```
  iter = 1e4
```

```
  ## Save Records ##
```

```
  Beta = matrix(-99, ncol = iter, nrow = p)
```

```
  Sigmasq = rep(-99, iter)
```

```
  ## Initial values ##
```

```
  beta = rep(1, p)
```

```
  sigmasq = 1
```

```
  ## Calculations ##
```

```
  IR = solve(t(X) %*% X + solve(R0))
```

```
  mu = IR %*% (t(X) %*% Y + solve(R0) %*% mu0)
```

```
  for(i in 1:iter){
```

```
    ## Update Beta ##
```

```
    beta = t(rmvnorm(1, mu, sigmasq[1] * IR))
```

```
    ## Update Sigmasq ##
```

```
    sigmasq = rinvgamma(1, (n+p+a)/2, (1/2)*(t(Y-X%*%beta)%*%(Y-X%*%beta))
```

```
    ## Save records ##
```

```
    Beta[,i] = beta
```

```
    Sigmasq[i] = sigmasq
```

```
  }
```

```

        return(list(Beta = Beta, Sigmasq = Sigmasq))
    }

## Part d ##
library(MCMCpack)
library(mvtnorm)
library(coda)
beta.true = c(0.5,1,3)
sigmasq.true = 1
n = 1000
eta = 500      ## historical sample size
p = length(beta.true)
X = cbind(rep(1,n), rbinom(n, 1, 0.5), rnorm(n, 2, 1))
alpha0 = 1
g = 0.1
sims = 100
Beta.power = matrix(-99, ncol = sims, nrow = p)
Sigmasq.power = rep(-99, sims)
Beta.g = matrix(-99, ncol = sims, nrow = p)
Sigmasq.g = rep(-99, sims)
Beta.mlr = matrix(-99, ncol = sims, nrow = p)
Sigmasq.mlr = rep(-99, sims)
## Power initial priors ##
mu0 = rep(2,p)
R = diag(100, p)
## G initial priors ##
beta0 = rep(2,p)
a = b = 1
for(i in 1:sims){
    Y = rnorm(n, X %*% beta.true, sqrt(sigmasq.true))

    ## For power prior ##
    #Y0 = rnorm(eta, X %*% beta.true, sqrt(sigmasq.true))
    #X0 = cbind(rep(1,eta), rbinom(eta, 1, 0.5), rnorm(eta, 2, 1))
    Y0 = Y
    X0 = X
    C.power = solve(t(X0)%*%X0 + solve(alpha0*R))
    mu.power = C.power %*% (t(X0)%*%Y0 + solve(alpha0*R)%*%mu0)
    cov.power = sigmasq.true / alpha0 * C.power
    alpha.power = eta*alpha0
    beta.power = alpha0*(t(Y0)%*%Y0-t(mu.power)%*%solve(C.power)%*%mu.power+t(mu.power)%*%mu.power)
    res.power = normal.invgamma(Y,X,mu.power,cov.power,alpha.power,beta.power)
    Beta.power[,i] = apply(res.power$Beta,1,mean)
    Sigmasq.power[i] = mean(res.power$Sigmasq)
}

```

```

## For g prior ##
mu.g = beta0
cov.g = g*sigmasq.true*solve(t(X)%*%X)
alpha.g = a
beta.g = b
res.g = normal.invgamma(Y,X,mu.g,cov.g,alpha.g,beta.g)
Beta.g[,i] = apply(res.g$Beta,1,mean)
Sigmasq.g[i] = mean(res.power$Sigmasq)

## Multiple Linear Regression in class ##
res.mlr = normal.invgamma(Y,X,rep(2,p),sigmasq.true*R,1,1)
Beta.mlr[,i] = apply(res.mlr$Beta,1,mean)
Sigmasq.mlr[i] = mean(res.mlr$Sigmasq)

print(i)
}

apply(Beta.power,1,mean)
mean(Sigmasq.power)
apply(Beta.g,1,mean)
mean(Sigmasq.g)
apply(Beta.mlr,1,mean)
mean(Sigmasq.mlr)

#####
## Problem 2 ##

## Part c ##
full.gibbs = function(Y, D, W, rho, b, sig2, tau2, as2, bs2, at2, bt2, iter = 1e4,
  n = length(Y)
  DW = D - rho * W
  I = sparseMatrix(1:n, 1:n, x = 1)
  #I = diag(n)

  b.save = matrix(-99, nrow = n, ncol = iter)
  sig2.save = rep(-99, n)
  tau2.save = rep(-99, n)
  beta0.save = rep(-99, n)

  for(t in 1:iter){

```

```

## Sample intercept ##
beta0 = rnorm(1, mean(Y - b), sqrt(sig2 / n))

## Sample spatial random effects all at once ##
Prec = I + DW / tau2
CH = chol(Prec)
R1 = solve(CH, rnorm(n, 0, sqrt(sig2)), sparse = TRUE)
R2 = solve(t(CH), Y - beta0, sparse = TRUE)
R3 = solve(CH, R2, sparse = TRUE)
b = as.vector(R1 + R3)

## Sample sig2 ##
as2s = n + as2
bs2s = as.vector(bs2 + (1/2) * (sum((Y - beta0 - b)^2) + b %% DW %
sig2 = 1 / rgamma(1, as2s, bs2s)

## Sample tau2 ##
at2s = n / 2 + at2
bt2s = bt2 + (1/2) * (t(b) %% DW %% b / sig2)
tau2 = 1 / rgamma(1, at2s, bt2s)

## Save ##
beta0.save[t] = beta0
b.save[,t] = b
sig2.save[t] = sig2
tau2.save[t] = tau2

if(verbose == TRUE){
  print(t)
  if(t %% 100 == 0){
    par(mfrow = c(3,1))
    plot(beta0.save[1:t])
    plot(sig2.save[1:t])
    plot(tau2.save[1:t])
  }
}

}
return(list("sig2" = sig2.save, "tau2" = tau2.save, "beta0" = beta0.save, "b" = b.save))
}

one.gibbs = function(Y, D, W, rho, b, sig2, tau2, as2, bs2, at2, bt2, iter = 1e4, v
  n = length(Y)
  DW = D - rho * W
  D = diag(D)

```

```

b.save = matrix(-99, nrow = n, ncol = iter)
sig2.save = rep(-99, n)
tau2.save = rep(-99, n)
beta0.save = rep(-99, n)

for(t in 1:iter){
  ## Sample intercept ##
  beta0 = rnorm(1, mean(Y - b), sqrt(sig2 / n))

  ## Sample spatial random effects one at a time ##
  for(i in 1:n){
    mui = rho * W[,i] %*% b / D[i]
    Wtau = D[i] / tau2
    pm = as.vector((Y[i] - beta0 + Wtau * mui) / (1 + Wtau))
    ps = as.vector(sqrt(sig2 / (1 + Wtau)))
    b[i] = rnorm(1, pm, ps)
  }

  ## Sample sig2 ##
  as2s = n + as2
  bs2s = as.vector(bs2 + (1/2) * (sum((Y - beta0 - b)^2) + t(b) %*% D))
  sig2 = 1 / rgamma(1, as2s, bs2s)

  ## Sample tau2 ##
  at2s = n / 2 + at2
  bt2s = bt2 + (1/2) * (t(b) %*% DW %*% b / sig2)
  tau2 = 1 / rgamma(1, at2s, bt2s)

  ## Save ##
  beta0.save[t] = beta0
  b.save[,t] = b
  sig2.save[t] = sig2
  tau2.save[t] = tau2

  if(verbose == TRUE){
    print(t)
    if(t %% 100 == 0){
      par(mfrow = c(3,1))
      plot(beta0.save[1:t])
      plot(sig2.save[1:t])
      plot(tau2.save[1:t])
    }
  }
}

```

```

    return(list("sig2" = sig2.save, "tau2" = tau2.save, "beta0" = beta0.save, "
}

```

```

## Part d ##
library(lattice)
library(MASS)
library(Matrix)
library(mvtnorm)
library(hierarchicalDS)
P1 = 50
P2 = 50
Y = matrix(-99, P1, P2)
for(p1 in 1:P1){
  for(p2 in 1:P2){
    Y[p1,p2] = 10*dnorm((p1-p2), 0, 10) + rnorm(1, 0, 0.1)
  }
}
levelplot(Y, main = "Observed Data")
W = square_adj(P1)
D = diag(rowSums(W))
Y = as.vector(Y)
n = length(Y)

```

```

load(file.choose())

```

```

res1 = full.gibbs(Y=Y,D=D,W=W,rho=1,b=rep(0,n),sig2=1,tau2=1,as2=0.001,bs2=0.001,at

```

```

res2 = one.gibbs(Y=Y,D=D,W=W,rho=1,b=rep(0,n),sig2=1,tau2=1,as2=0.001,bs2=0.001,at

```

```

Yhat1 = mean(res1$beta0[5000:10000]) + apply(res1$b[,5000:10000], 1, mean)
levelplot(matrix(Yhat1, ncol = P1, nrow = P2), main = "Full Block Gibbs")
par(mfrow = c(4,1))
plot(res1$beta0[5000:10000])
plot(res1$sig2[5000:10000])
plot(res1$tau2[5000:10000])
plot(res1$b[1,5000:10000])
Yhat2 = mean(res2$beta0[5000:10000]) + apply(res2$b[,5000:10000], 1, mean)
levelplot(matrix(Yhat2, ncol = P1, nrow = P2), main = "Individual Gibbs")
par(mfrow = c(4,1))
plot(res2$beta0[5000:10000])
plot(res2$sig2[5000:10000])

```

```
plot(res2$tau2[5000:10000])  
plot(res2$b[1,5000:10000])
```