

**SDS 383D Ex 04:**  
**Hierarchical Models**

March 22, 2016

**Jennifer Starling**

## Math Tests

The data set in “mathtest.csv” shows the scores on a standardized math test from a sample of 10th-grade students at 100 different U.S. urban high schools, all having enrollment of at least 400 10th-grade students. (A lot of educational research involves “survey tests” of this sort, with tests administered to all students being the rare exception.)

Let  $\theta_i$  be the underlying mean test score for school  $i$ , and let  $y_{ij}$  be the score for the  $j$ th student in school  $i$ . Starting with the “mathtest.R” script, you’ll notice that the extreme school-level averages  $\bar{y}_i$  (both high and low) tend to be at schools where fewer students were sampled.

### Part 1

Briefly explain why this would be.

The extreme school-level averages occur in the schools with smaller sample sizes because we do not do a very good job of estimating the mean when sample size is small. These schools do not have min and max observation values that are more extreme than the other schools; they just have fewer observations to balance out the calculation of the mean. The smaller the sample size, the more influential an extreme observation is over the group mean.

### Part 2

Fit a normal hierarchical model to these data via Gibbs sampling:

$$\begin{aligned} y_{ij} &\sim N(\theta_i, \sigma^2) \\ \theta_i &\sim N(\mu, \tau^2 \sigma^2) \end{aligned}$$

Decide upon sensible priors for the unknown model parameters  $(\mu, \sigma^2, \tau^2)$ .

The model is as follows.

$$\begin{aligned} (y_{ij} | \theta_i, \sigma^2) &\sim N(\theta_i, \sigma^2) \\ (\theta_i | \mu, \sigma^2, \tau^2) &\sim N(\mu, \sigma^2 \tau^2) \\ \mu &\sim I_{\mathbb{R}}(\mu), \text{ a flat prior on the real line} \\ \tau^2 &\sim I_{\mathbb{R}^+}(\tau^2), \text{ a flat prior on the positive real line} \\ \sigma^2 &\sim \left( \frac{1}{\sigma^2} \right) I_{\mathbb{R}^+}(\sigma^2), \text{ Jeffreys prior} \end{aligned}$$

where

- $i = 1, \dots, p$  indexes the  $p$  groups.
- $n_i$  = sample size in each group.
- $j = 1, \dots, n_i$  indexes observations in a group.
- $n$  = total number of observations.

The likelihood is

$$L(y | \theta_1, \dots, \theta_p, \sigma^2) \sim \prod_{i=1}^p \prod_{j=1}^{n_i} \left( \frac{1}{\sigma^2} \right)^{\frac{1}{2}} \exp \left[ -\frac{1}{\sigma^2} (y_{ij} - \theta_i)^2 \right] = (\sigma^2)^{-\frac{n}{2}} \exp \left[ -\frac{1}{2\sigma^2} \sum_{i=1}^p \sum_{j=1}^{n_i} (y_{ij} - \theta_i)^2 \right]$$

The full conditionals are as follows.

---


$$(\theta_i | y, \mu, \sigma^2, \tau^2)$$

Note that  $\bar{y}_i$  is a sufficient statistic for the  $y$ 's, with  $\bar{y}_i \sim N\left(\theta_i, \frac{\sigma^2}{n}\right)$ .

$$(\theta_i | y, \mu, \sigma^2, \tau^2) \propto (\sigma^2)^{-\frac{1}{2}} \exp\left[-\frac{1}{2\sigma^2/n}(\bar{y}_i - \theta_i)^2\right] (\tau^2 \sigma^2)^{-\frac{1}{2}} \exp\left[-\frac{1}{2\sigma^2 \tau^2}(\theta_i - \mu)^2\right]$$

This is the normal-normal model, therefore

$$(\theta_i | y, \mu, \sigma^2, \tau^2) \sim N(m^*, v^*)$$

with

$$\begin{aligned} v^* &= \left[ \frac{n_i}{\sigma^2} + \frac{1}{\sigma^2 \tau^2} \right]^{-1} = \left[ \frac{n_i \tau^2 + 1}{\sigma^2 \tau^2} \right]^{-1} = \sigma^2 \left[ \frac{\tau^2}{n_i \tau^2 + 1} \right] \\ m^* &= v^* \left[ \left( \frac{n_i}{\sigma^2} \right) \bar{y}_i + \left( \frac{1}{\sigma^2 \tau^2} \right) \mu \right] \\ &= \sigma^2 \left[ \frac{\tau^2}{n_i \tau^2 + 1} \right] \left[ \left( \frac{n_i}{\sigma^2} \right) \bar{y}_i + \left( \frac{1}{\sigma^2 \tau^2} \right) \mu \right] \\ &= \left[ \frac{n_i \tau^2}{n_i \tau^2 + 1} \right] \bar{y}_i + \left[ \frac{1}{n_i \tau^2 + 1} \right] \mu \\ &= w \bar{y}_i + (1 - w) \mu \end{aligned}$$

So full conditional is

$$(\theta_i | y, \mu, \sigma^2, \tau^2) \sim N\left(\left[ \frac{n_i \tau^2}{n_i \tau^2 + 1} \right] \bar{y}_i + \left[ \frac{1}{n_i \tau^2 + 1} \right] \mu, \sigma^2 \left[ \frac{\tau^2}{n_i \tau^2 + 1} \right]\right) \quad (1)$$

---


$$(\mu | \theta, y, \sigma^2, \tau^2)$$

$$\begin{aligned} (\mu | \theta, y, \sigma^2, \tau^2) &\propto \exp\left[-\frac{1}{2\sigma^2 \tau^2} \sum_{i=1}^p (\theta_i - \mu)^2\right] \cdot 1 \\ &= \exp\left[-\frac{1}{2\sigma^2 \tau^2} \{(\theta_1 - \mu)(\theta_1 - \mu) + \dots + (\theta_p - \mu)(\theta_p - \mu)\}\right] \\ &= \exp\left[-\frac{1}{2\sigma^2 \tau^2} \left\{ p\mu^2 - 2\mu \sum_{i=1}^p \theta_i + \sum_{i=1}^p \theta_i^2 \right\}\right] \\ &= \exp\left[-\frac{p}{2\sigma^2 \tau^2} \left\{ \mu^2 - 2\mu \left( \frac{\sum_{i=1}^p \theta_i}{p} \right) + \frac{\sum_{i=1}^p \theta_i^2}{p} \right\}\right] \\ &\propto \exp\left[-\frac{p}{2\sigma^2 \tau^2} \left\{ \mu^2 - 2\mu \bar{\theta}_i \right\}\right] \end{aligned}$$

We recognize this as a Normal kernel, therefore

$$(\mu | \theta, y, \sigma^2, \tau^2) \sim N\left(\bar{\theta}_i, \frac{\sigma^2 \tau^2}{p}\right) \quad (2)$$

---


$$(\sigma^2 | \theta, y, \mu, \tau^2)$$

$$\begin{aligned}
(\sigma^2 | \theta, y, \mu, \tau^2) &\propto (\sigma^2)^{-\frac{n}{2}} \exp \left[ -\frac{1}{2\sigma^2} \sum_{i=1}^p \sum_{j=1}^{n_i} (y_{ij} - \theta_i)^2 \right] (\sigma^2)^{-\frac{p}{2}} \exp \left[ -\frac{1}{2\sigma^2 \tau^2} \sum_{i=1}^p (\theta_i - \mu)^2 \right] \left( \frac{1}{\sigma^2} \right) \\
&= (\sigma^2)^{-\frac{(n+p)}{2}-1} \exp \left[ -\left( \frac{1}{\sigma^2} \right) \cdot \left\{ \frac{1}{2} \sum_{i=1}^p \sum_{j=1}^{n_i} (y_{ij} - \theta_i)^2 + \frac{1}{2\tau^2} \sum_{i=1}^p (\theta_i - \mu)^2 \right\} \right]
\end{aligned}$$

We recognize this as an Inverse-Gamma kernel, therefore

$$(\sigma^2 | \theta, y, \mu, \tau^2) \sim IG \left( \frac{(n+p)}{2}, \left\{ \frac{1}{2} \sum_{i=1}^p \sum_{j=1}^{n_i} (y_{ij} - \theta_i)^2 + \frac{1}{2\tau^2} \sum_{i=1}^p (\theta_i - \mu)^2 \right\} \right) \quad (3)$$

$$(\tau^2 | \theta, y, \mu, \sigma^2)$$

$$(\tau^2 | \theta, y, \mu, \sigma^2) \propto (\tau^2)^{-\frac{p}{2}} \exp \left[ -\frac{1}{2\sigma^2 \tau^2} \sum_{i=1}^p (\theta_i - \mu)^2 \right] \cdot 1$$

We recognize this as an Inverse Gamma kernel, therefore

$$(\tau^2 | \theta, y, \mu, \sigma^2) \sim IG \left( \frac{p}{2} - 1, \frac{1}{2\sigma^2} \sum_{i=1}^p (\theta_i - \mu)^2 \right) \quad (4)$$

### Part 3

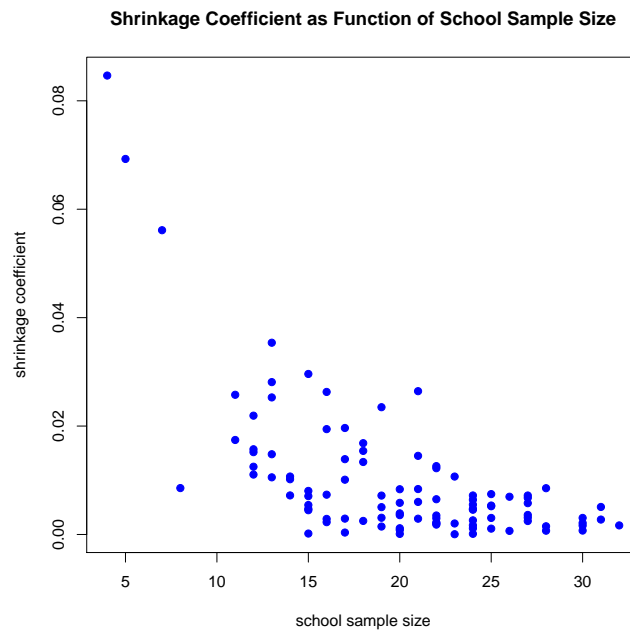


Figure 1: Shrinkage estimator by school sample size

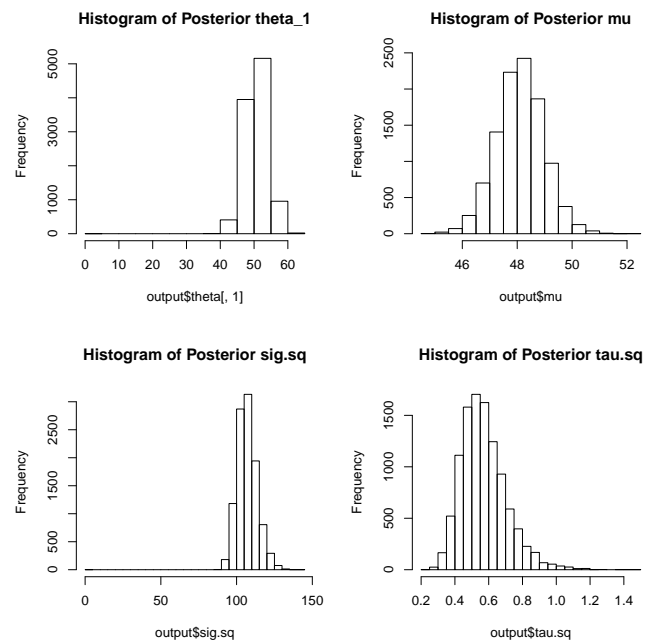


Figure 2: Histograms of Posteriors

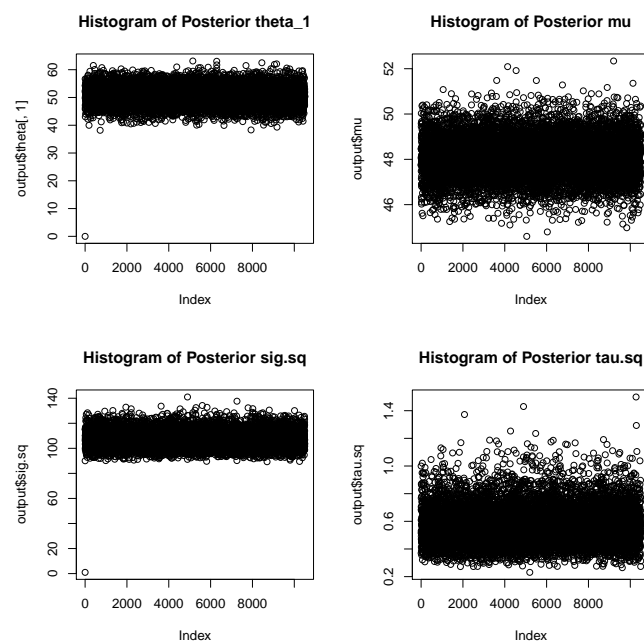


Figure 3: Traces for Gibbs Sampler

## Price Elasticity of Demand

### Linear Hierarchical Model using Empirical Bayes

Model is specified as

$$\begin{aligned}
 \log(Q_{it}) &= \log(\alpha_i) + \beta_i \log(P_{it}) + \gamma_i x_{it} + \theta_i [\log(P_{it}) * x_{it}] + e_{it} \\
 \alpha_i &\sim N(\mu_\alpha, \tau_\alpha^2) \\
 \beta_i &\sim N(\mu_\beta, \tau_\beta^2) \\
 \gamma_i &\sim N(\mu_\gamma, \tau_\gamma^2) \\
 \theta_i &\sim N(\mu_\theta, \tau_\theta^2) \\
 e_{it} &\sim N(0, \sigma^2)
 \end{aligned}$$

Where  $i = \{1, 2, \dots, 88\}$  indexes stores, and  $t = \{1, 2, \dots, 68\}$  indexes week (repeated obs on each store).

$\log(Q_{it})$  = Response; log-volume for store  $i$  at week  $t$

$\log(P_{it})$  = Log-price for store  $i$  at week  $t$

$\log(\alpha_i)$  = Intercept for each store

$x_{it}$  = Indicator variable for ad display (displayed ad = 1)

$\log(P_{it}) * x_{it}$  = Interaction; shape may change depending on whether ad in store

Variance estimates using lmer to fit the model were as follows.

$$\hat{\tau}_\alpha^2 = 5.0478$$

$$\hat{\tau}_\beta^2 = 4.6658$$

$$\hat{\tau}_\gamma^2 = 0.9634$$

$$\hat{\tau}_\theta^2 = 0.7004$$

$$\hat{\sigma}^2 = 0.06733$$

Residual plot does not show evidence of major model mis-fit.

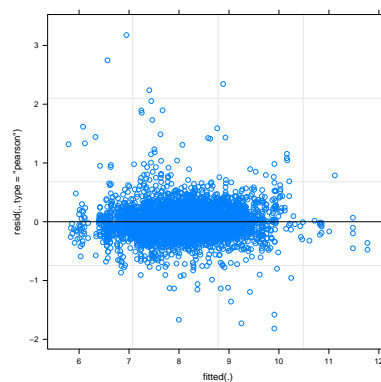


Figure 4: Residual plot for hierarchical model

Model summary:

```

Linear mixed model fit by REML ['lmerMod']
Formula: logQ ~ (logP + disp + disp:logP | store)
Data: data

5 REML criterion at convergence: 1811.9

Scaled residuals:
  Min       1Q   Median       3Q      Max
-7.0245 -0.4898 -0.0317  0.4348 12.2358

10 Random effects:
   Groups   Name                Variance Std.Dev. Corr
   store    (Intercept)  5.0478     2.2467
            logP          4.6658     2.1600   -0.94
15            disp         0.9634     0.9816    0.47  -0.55
            logP:disp     0.7004     0.8369   -0.32  0.38  -0.97
   Residual              0.0675     0.2598
Number of obs: 5555, groups:  store, 88

20 Fixed effects:
              Estimate Std. Error t value
(Intercept)  8.18711     0.07794    105

```

## Fully Bayesian Hierarchical Linear Model

### Model

Define the following variables.

$y_i = \log(Q)$ , a  $(n_s \times 1)$  vector of log-volume observations for each of the  $s$  stores.

$X_i = W_i = \begin{bmatrix} 1 & \log P_{it} & ad_{it} & \log P_{it} * ad_{it} \end{bmatrix}$ , a  $(n_s \times p)$  matrix of covariates for each store.

where

$s$  = Number of stores. Stores are indexed by  $i = \{1, 2, \dots, s\}$

$n_i$  = Number of observations (weeks) within store  $i$ .

We can write the model as follows. This model includes an overall mean of each covariate  $\beta_j$ , plus store-varying offsets.

$$y_i = X_i \beta + W_i b_i + e_i, \text{ with } e_i \sim N(0, \sigma^2 I_{n_i})$$

$$\beta \sim N_p(\mu_\beta, V_\beta)$$

$$b_i \sim N_p(0, \Sigma)$$

$$\sigma^2 \sim \frac{1}{\sigma^2}$$

$$\Sigma \sim IW(d, C)$$

### Likelihood

$$y_i \sim N_{n_i}(X_i \beta + W_i b_i, \sigma^2 I_{n_i})$$

$$y_i \propto (\sigma^2)^{-\frac{n_i}{2}} \exp \left[ -\frac{1}{2\sigma^2} (y_i - X_i \beta - W_i b_i)^T (y_i - X_i \beta - W_i b_i) \right]$$

$$y_1, \dots, y_s \propto (\sigma^2)^{-\frac{n}{2}} \exp \left[ -\frac{1}{2\sigma^2} \sum_{i=1}^s (y_i - X_i \beta - W_i b_i)^T (y_i - X_i \beta - W_i b_i) \right]$$

### Full Conditionals

$(b_i | \dots)$

$$\begin{aligned} (b_i | \dots) &\propto \exp \left[ -\frac{1}{2} b_i^T \Sigma^{-1} b_i \right] \cdot \exp \left[ -\frac{1}{2\sigma^2} (y_i - X_i \beta - W_i b_i)^T (y_i - X_i \beta - W_i b_i) \right] \\ &\propto \exp \left[ -\frac{1}{2} b_i^T \Sigma^{-1} b_i \right] \cdot \exp \left[ -\frac{1}{2\sigma^2} (b_i^T W_i^T W_i b_i - 2b_i^T W_i^T y_i - 2b_i^T W_i^T X_i \beta) \right] \\ &\propto \exp \left[ -\frac{1}{2} b_i^T \Sigma^{-1} b_i \right] \cdot \exp \left[ -\frac{1}{2\sigma^2} (b_i^T W_i^T W_i b_i - 2b_i^T W_i^T (y_i - X_i \beta)) \right] \end{aligned}$$



We recognize this as the multivariate normal kernel.

$$(b_i | \dots) \sim N(m^*, V^*), \text{ with} \quad (5)$$

$$V^* = \left[ \Sigma^{-1} + \frac{1}{\sigma^2} W_i^T W_i \right]^{-1} \quad (6)$$

$$m^* = V^* \left[ \frac{1}{\sigma^2} W_i^T (y_i - X_i \beta) \right] \quad (7)$$

$(\beta | \dots)$

$$\begin{aligned} (\beta | \dots) &\propto \exp \left[ -\frac{1}{2} (\beta - \mu_\beta)^T V_\beta^{-1} (\beta - \mu_\beta) \right] \cdot \exp \left[ -\frac{1}{2\sigma^2} \sum_{i=1}^s (y_i - X_i \beta - W_i b_i)^T (y_i - X_i \beta - W_i b_i) \right] \\ &\propto \exp \left[ -\frac{1}{2} (\beta^T V_\beta^{-1} \beta - 2\beta^T V_\beta^{-1} \mu_\beta) - \frac{s}{2\sigma^2} \left( \beta^T \left( \sum_{i=1}^s X_i^T X_i \right) \beta - 2\beta^T \left( \sum_{i=1}^s X_i^T y_i - \sum_{i=1}^s X_i^T W_i b_i \right) \right) \right] \\ &= \exp \left[ -\frac{1}{2} (\beta^T V_\beta^{-1} \beta - 2\beta^T V_\beta^{-1} \mu_\beta) - \frac{s}{2\sigma^2} \left( \beta^T \left( \sum_{i=1}^s X_i^T X_i \right) \beta - 2\beta^T \left( \sum_{i=1}^s X_i^T (y_i - W_i b_i) \right) \right) \right] \end{aligned}$$

We recognize this as the univariate normal kernel.

$$(\beta | \dots) \sim N_p(m^*, V^*), \text{ with} \quad (8)$$

$$V^* = \left[ V_\beta^{-1} + \left( \frac{1}{\sigma^2} \right) \sum_{i=1}^s X_i^T X_i \right]^{-1} \quad (9)$$

$$m^* = V^* \left[ V_\beta^{-1} \mu_\beta + \left( \frac{1}{\sigma^2} \right) \sum_{i=1}^s X_i^T (y_i - W_i b_i) \right] \quad (10)$$

$(\sigma^2 | \dots)$

$$(\sigma^2 | \dots) \propto \left( \frac{1}{\sigma^2} \right) (\sigma^2)^{-\frac{n}{2}} \exp \left[ -\frac{1}{2\sigma^2} \underbrace{\sum_{i=1}^s (y_i - X_i \beta - W_i b_i)^T (y_i - X_i \beta - W_i b_i)}_{RSS_{\sigma^2}} \right]$$

We recognize this as the inverse gamma kernel.

$$(\sigma^2 | \dots) \sim IG \left( \frac{n}{2}, \frac{RSS_{\sigma^2}}{2} \right) \quad (11)$$

$(\Sigma | \dots)$

$$\begin{aligned} (\sigma^2 | \dots) &\propto |\Sigma|^{-\left(\frac{d+p+1}{2}\right)} \exp \left[ -\frac{1}{2} \text{tr} (C \Sigma^{-1}) \right] \cdot |\Sigma|^{-\left(\frac{s}{2}\right)} \exp \left[ -\frac{1}{2} \sum_{i=1}^s b_i^T \Sigma^{-1} b_i \right] \\ &= |\Sigma|^{-\left(\frac{d+s+p+1}{2}\right)} \exp \left[ -\frac{1}{2} \text{tr} (C \Sigma^{-1}) - \frac{1}{2} \text{tr} \left( \sum_{i=1}^s b_i b_i^T \Sigma^{-1} \right) \right] \end{aligned}$$

We recognize this as the Inverse Wishart kernel.

$$(\sigma^2 | \dots) \sim IW \left( d + s, C + \sum_{i=1}^s b_i b_i^T \right) \quad (12)$$

The demand curves for the 88 stores are as follows. Stores are ordered in decreasing order by average price. Red represents weeks where an ad is displayed; blue represents no ad.

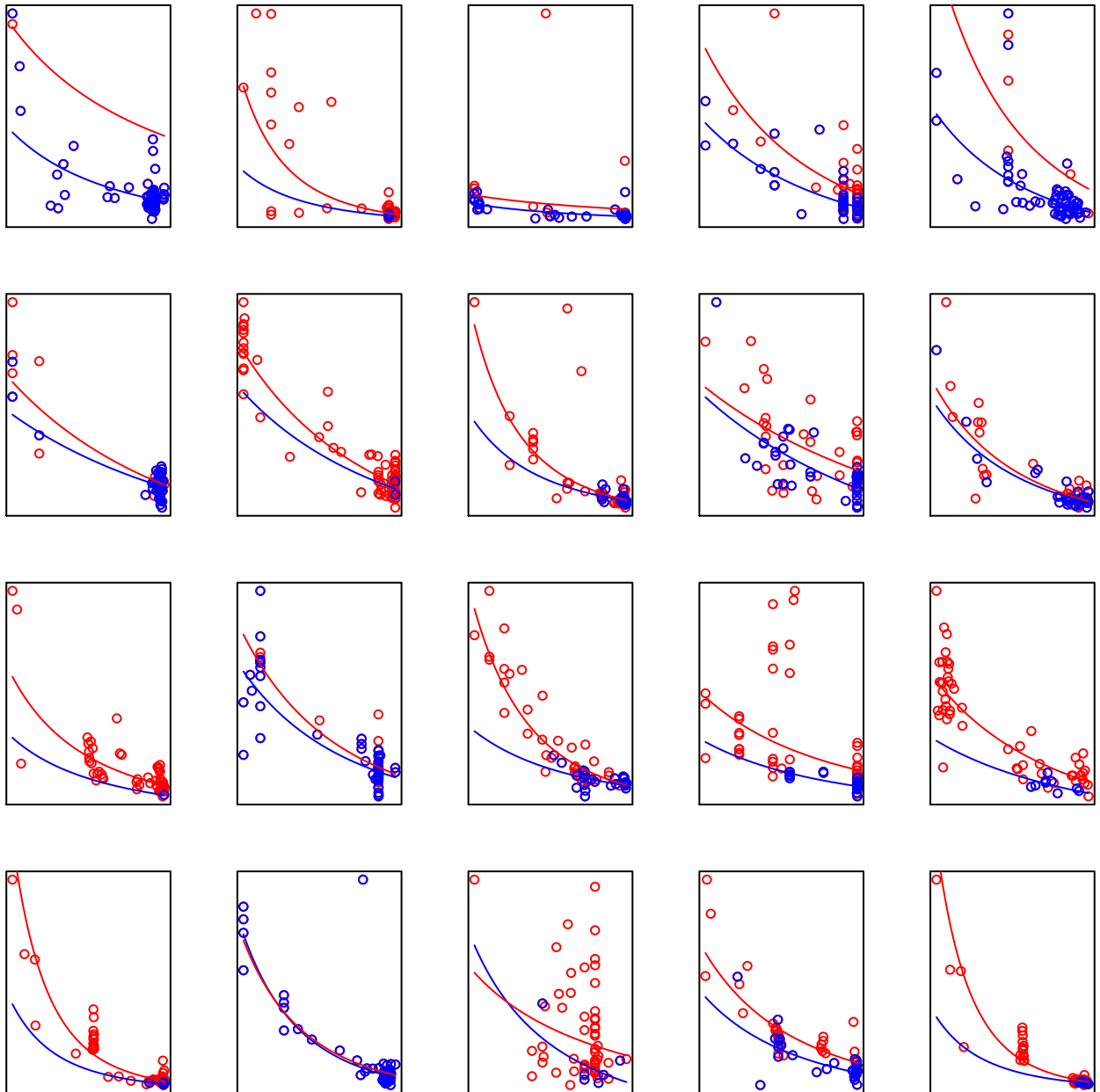


Figure 5: Demand Curves for 88 Stores

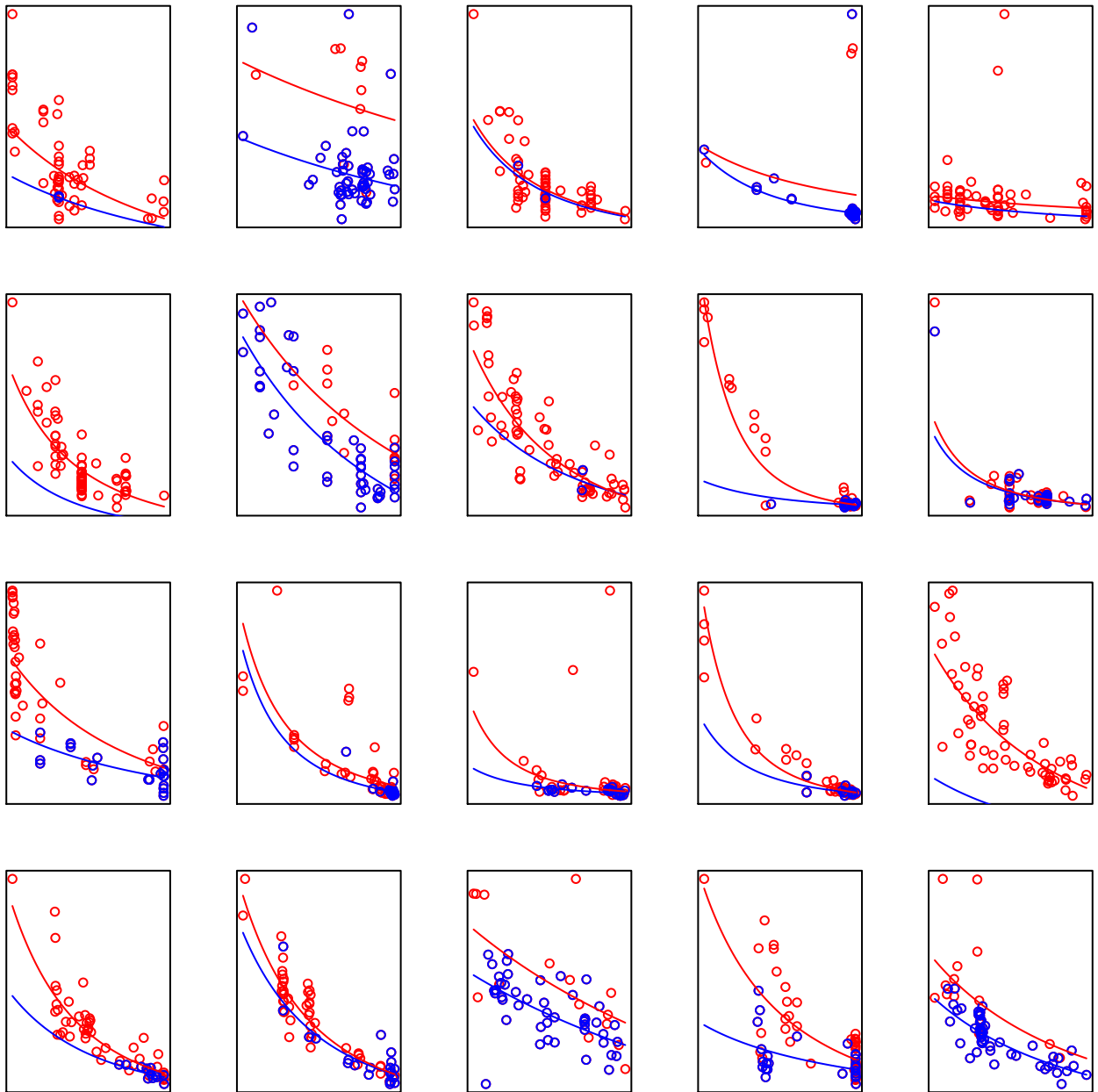


Figure 6: Demand Curves for 88 Stores

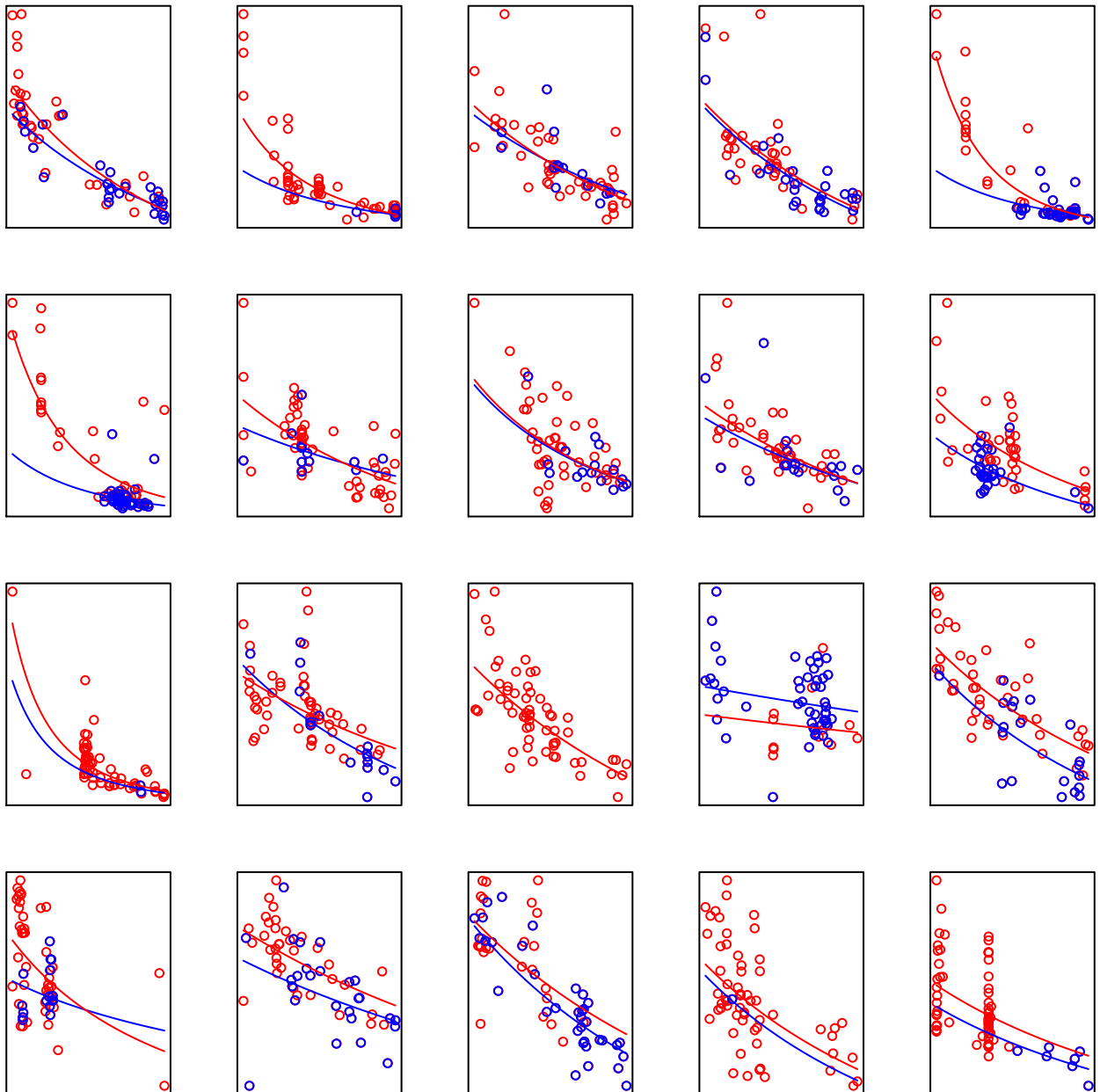


Figure 7: Demand Curves for 88 Stores

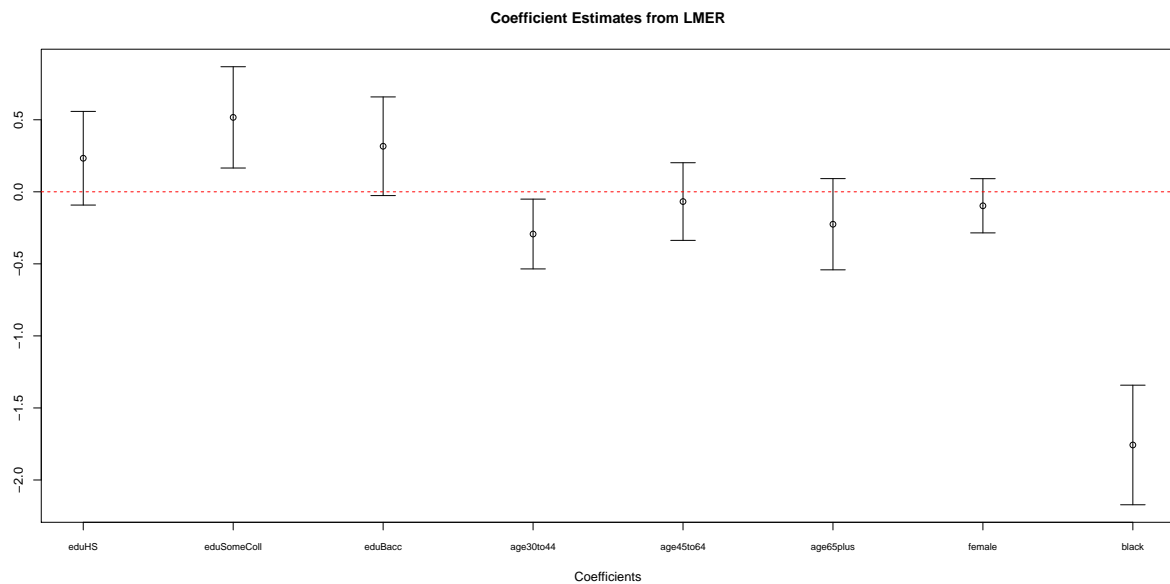
## A Hierarchical Probit Model via Data Augmentation

### Empirical Bayes Analysis using LMER

We can fit a hierarchical model with state-varying intercepts and a fixed  $\beta$  model using `glmer`, with `family = 'binomial'`. This model is fit using the following call. (wt.sc indicates scaled and centered weight.)

```
hlm = glmer(bush ~ edu + age + female + black + wt.sc + (1|state),  
            data=data, family='binomial')
```

The variance estimate for the state-varying intercept term is 0.1732. Confidence intervals for fixed effects are below.



## Fully Bayesian Hierarchical Augmented Model using Gibbs Sampler

### Model:

Original probit model:

$$P(y_{ij} = 1) = \Phi(z_{ij})$$

$$z_{ij} = \mu_i + x_{ij}^T \beta$$

The trick proposed by Albert and Chib (1993) is to introduce a latent variables  $z_{ij}$ , where we observe the  $y_{ij}$  but the underlying  $z_{ij}$  are normally distributed. In this case, the  $z_{ij}$  in the model formulation above can act as our latent variable.

We can define the model as follows.

$$z_i \sim N_{n_i}(W_i \mu_i + X_i \beta, I_{n_i})$$

$$\mu_i \sim N_1(0, \tau^2)$$

$$\beta \sim N(\mu_\beta, \Sigma)$$

This model allows for a state-varying intercept, with other covariates fixed. In this formulation,  $X_i$  is the  $(n_i \times p)$  matrix of demographic covariates for each state's observations, including a column of 1's for the intercept. Then  $W_i$  is a  $(n_i \times 1)$  matrix; a single column of 1's, to hold the intercept offset for each state.

Then we can easily that

$$P(Y_{ij} = 1) = \Phi(W_i \mu_i + X_i \beta)$$

And we can define  $y$  as

$$y_{ij} = \begin{cases} 1 & \text{if } z_{ij} > 0 \\ 0 & \text{if } z_{ij} < 0 \end{cases}$$

The Gibbs Sampler will then proceed as follows:

- (1) Update  $\mu_1, \dots, \mu_s, \beta, \tau^2$  by drawing from full conditionals.
- (2) Update  $z_{ij}$  by drawing from the truncated normal, based on whether each observed  $y_{ij}$  is greater than zero. (\*\*)
- (3) Calculate update  $P(y_{ij} = 1)$  based on all updated parameter values.

### Likelihood

$$(z_1, \dots, z_n) \propto \exp \left[ -\frac{1}{2} \sum_{i=1}^s (z_i - W_i \mu_i - X_i \beta)^T (z_i - W_i \mu_i - X_i \beta) \right]$$

(\*\*) Note: For updating the  $z_{ij}$ , as shown below, this 'likelihood' acts as the prior, and we update  $z_{ij}$  with the observed  $y$  likelihood to obtain its posterior. The latent  $z_{ij}$ s must be included in the sampler.

**Full Conditionals, including  $z_i$**  $(\beta | \dots)$ 

$$(\beta | \dots) \propto \exp \left[ -\frac{1}{2} (\beta - \mu_\beta)^T \Sigma^{-1} (\beta - \mu_\beta) \right] \exp \left[ -\frac{1}{2} \sum_{i=1}^s \left\{ \beta^T X_i^T X_i \beta - 2\beta^T X_i^T (z_i - W_i \mu_i) \right\} \right]$$

We recognize this as a normal-normal update.

$$(\beta | \dots) \sim N_p(m^*, V^*), \text{ with} \quad (13)$$

$$V^* = \left( \Sigma^{-1} + \sum_{i=1}^s X_i^T X_i \right)^{-1} \quad (14)$$

$$m^* = V^* \left[ \Sigma^{-1} \mu_\beta + \sum_{i=1}^s X_i^T (z_i - W_i \mu_i) \right] \quad (15)$$

 $(\mu_i | \dots)$ 

$$\begin{aligned} (\mu_i | \dots) &\propto \exp \left[ -\frac{1}{2} \mu_i^T \mu_i \right] \exp \left[ -\frac{1}{2} \left\{ \mu_i^T W_i^T W_i \mu_i - 2\mu_i^T W_i^T (z_i - X_i \beta) \right\} \right] \\ &= \exp \left[ -\frac{1}{2} \left\{ \mu_i^2 (1 + n_i) - 2\mu_i^T W_i^T (z_i - X_i \beta) \right\} \right] \end{aligned}$$

We recognize this as a normal-normal update.

$$(\mu_i | \dots) \sim N_1(m^*, v^*), \text{ with}$$

$$v^* = \frac{1}{\tau^2 + n_i}$$

$$m^* = v^* \left[ W_i^T (z_i - X_i \beta) \right]$$

 $(\tau^2 | \dots)$ 

$$(\tau^2 | \dots) \propto (1) \left( \tau^2 \right)^{-\frac{s}{2}} \exp \left[ -\frac{1}{2} \sum_{i=1}^s \mu_i^2 \right]$$

We recognize this as the inverse gamma kernel.

$$(\tau^2 | \dots) \sim IG\left(\frac{s}{2}, \frac{1}{2} \sum_{i=1}^s \mu_i^2\right) \quad (16)$$

 $(z_i | \dots)$ 

$$(z_{ij} | y_{ij} = 1) \sim [\mathbf{1}(y_{ij} = 1) \cdot N(\mu_i + X_{ij}\beta, 1)_+]$$

$$(z_{ij} | y_{ij} = 0) \sim [\mathbf{1}(y_{ij} = 0) \cdot N(\mu_i + X_{ij}\beta, 1)_-]$$

Where

$$N(\mu_i + X_{ij}\beta, 1)_+ = \text{Truncated Normal } (0, \infty)$$

$$N(\mu_i + X_{ij}\beta, 1)_- = \text{Truncated Normal } (-\infty, 0)$$

We can confirm mixing of the Gibbs sampler using trace plots for a handful of the predictors. Trace plots are also shown for a few of the state intercepts, where the fixed intercept and the offset are combined.

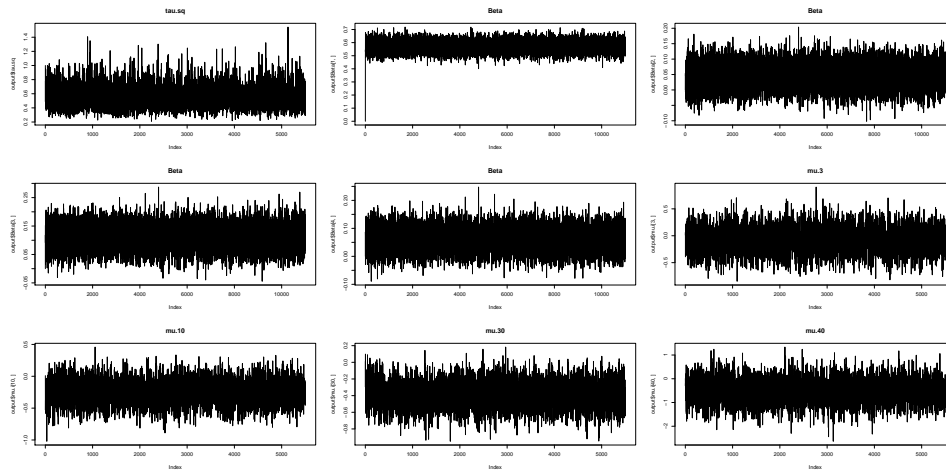


Figure 8: Trace plots for selected predictors

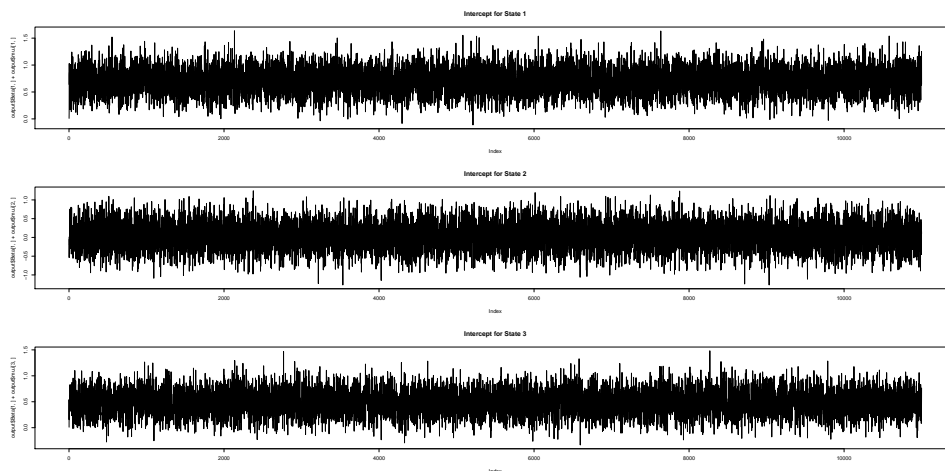


Figure 9: Trace plots for selected intercepts, including fixed and offset terms

## Results

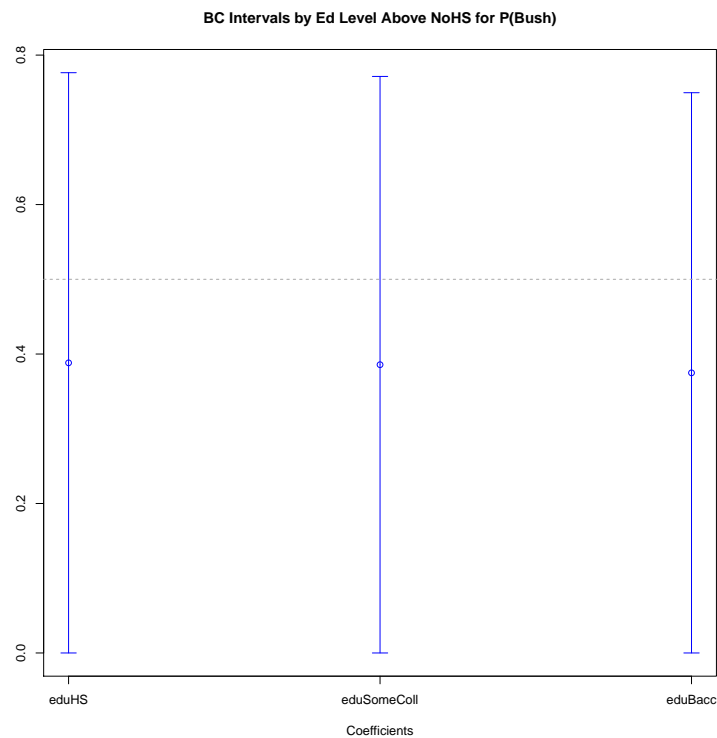
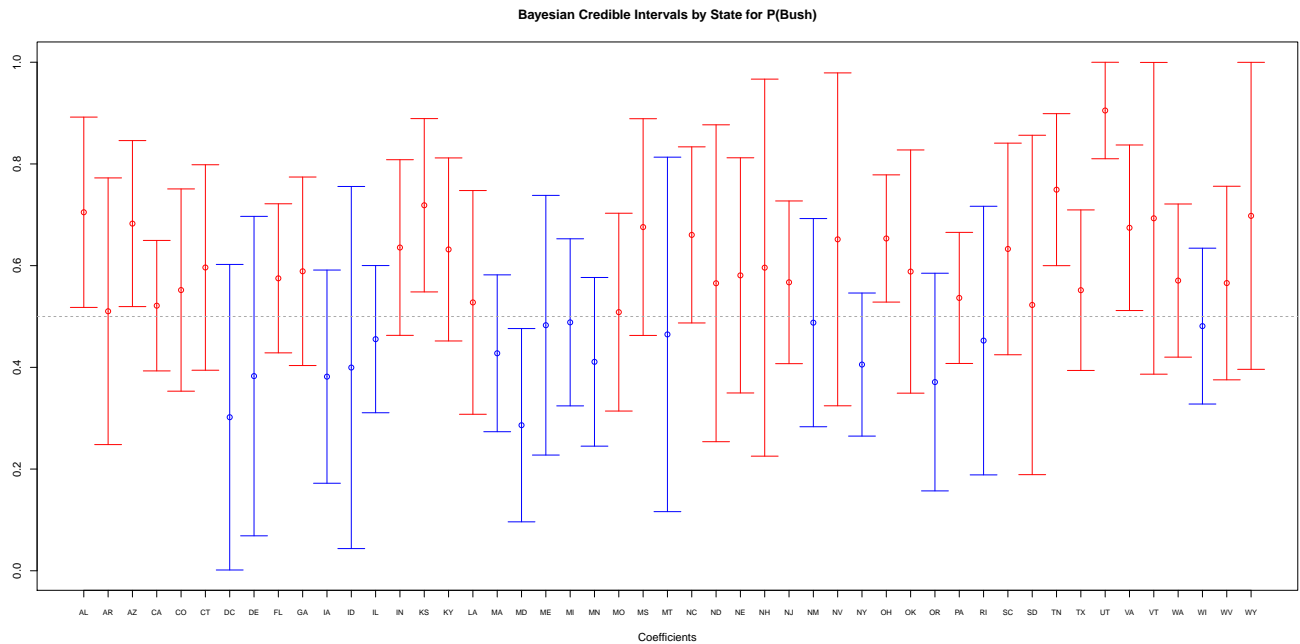
The goal of the analysis is to understand how demographic information, including state, relate to the probability that someone will support Bush in the 1988 presidential election.

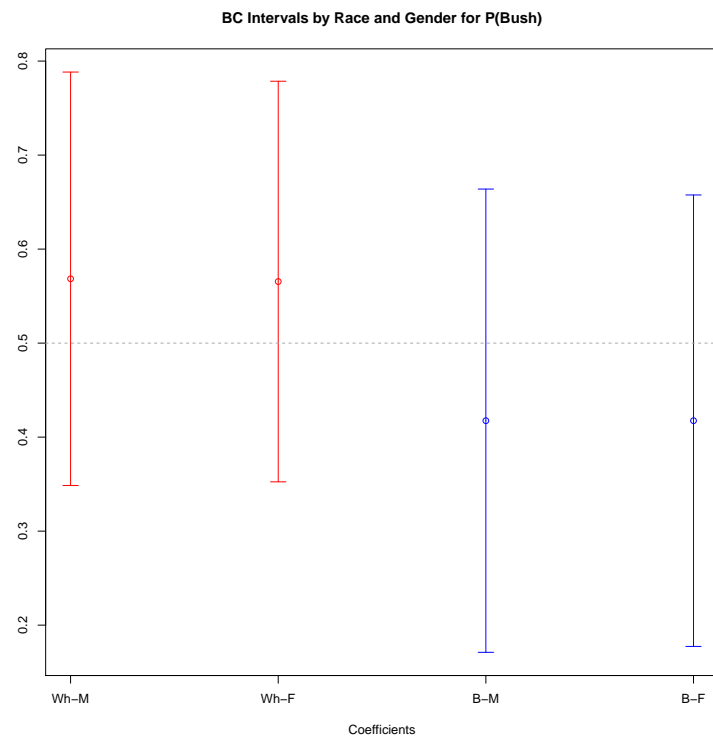
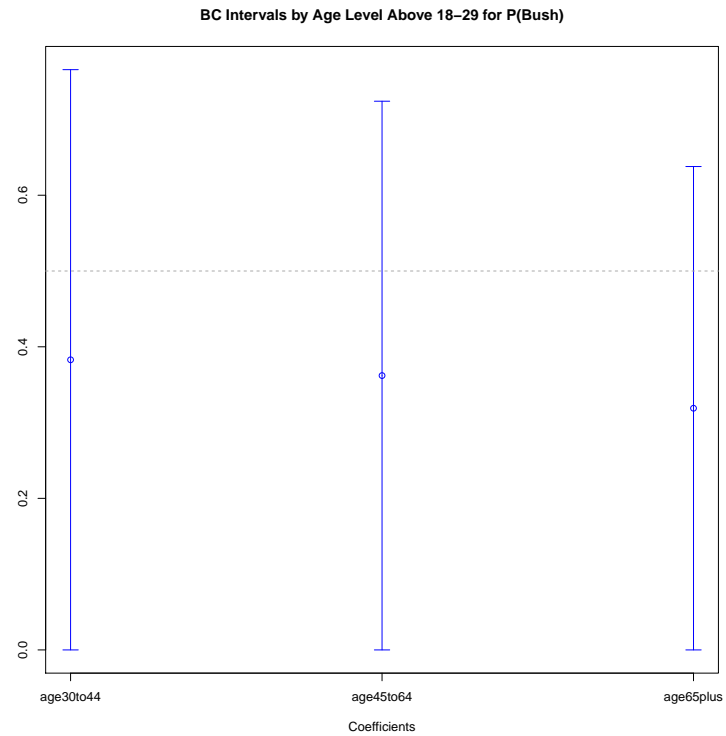
As illustrated in the plots below,

- State appears to vary the most in terms of probability of supporting Bush.
- There does not appear to be any marginal education effects, for adding each consecutive education level in addition to the previous level. (Baseline is no highschool.)
- Age also does not appear to factor significantly into the probability of supporting Bush. This was surprising given today's fairly polarized age demographics.



- There was not a significant race-gender interaction, and there was not a significant difference in genders for either race. However, white voters were far more likely to support Bush than black voters of both genders.





## Gene Expression Over Time