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 θ_i be the underlying mean test score for school i, and let y_{ij} be the score for the jth 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:

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

Decide upon sensible priors for the unknown model parameters (μ, σ^2, τ^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

$$\begin{aligned} i &= 1, ..., p \text{ indexes the } p \text{ groups.} \\ n_i &= \text{ sample size in each group.} \\ j &= 1, ..., n_i \text{ indexes observations in a group.} \\ n &= \text{ total number of observations.} \end{aligned}$$

The likelihood is

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

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 $\bar{y}_i \sim N\left(\theta_i, \frac{\sigma^2}{n}\right)$.

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

This is the normal-normal model, therefore

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

So full conditional is

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

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

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

We recognize this as a Normal kernel, therefore

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

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

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

We recognize this as an Inverse-Gamma kernel, therefore

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

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

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

We recognize this as an Inverse Gamma kernel, therefore

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

Part 3

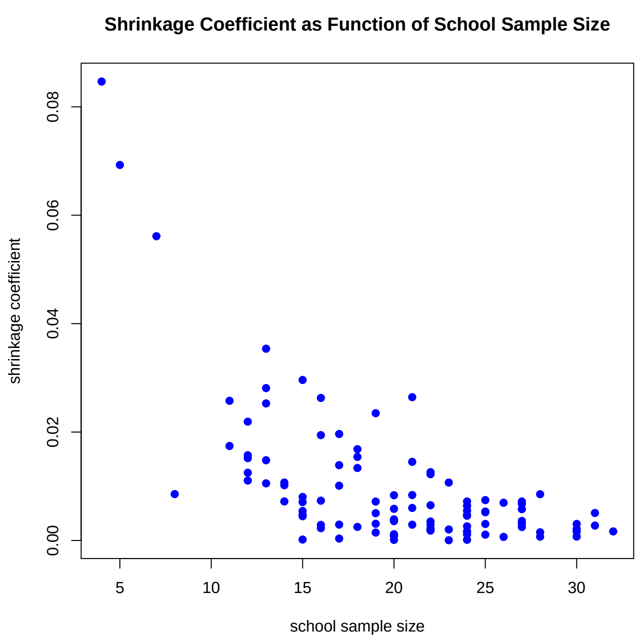

Figure 1: Shrinkage estimator by school sample size

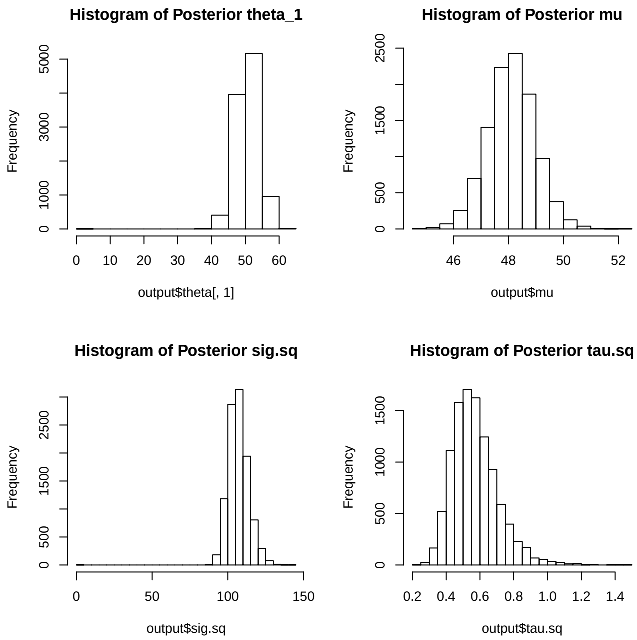

Figure 2: Histograms of Posteriors

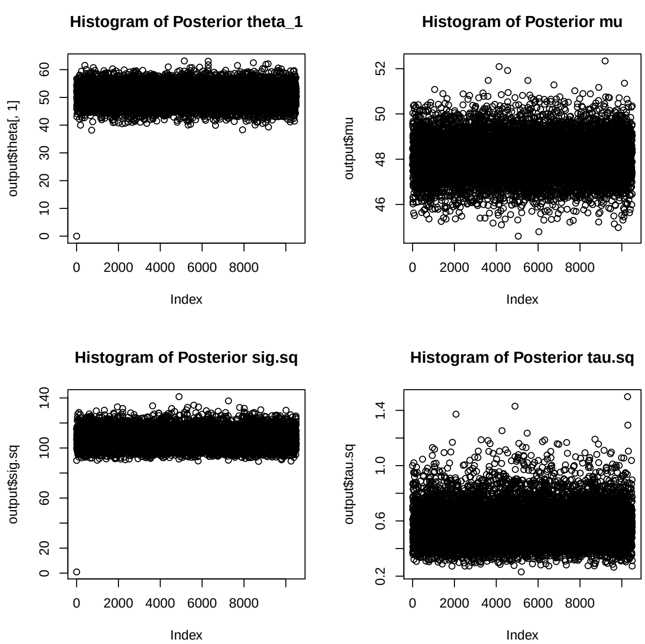

Figure 3: Traces for Gibbs Sampler

Price Elasticity of Demand

Linear Hierarchical Model using Empirical Bayes

Model is specified as

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

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

$$
\begin{aligned}
log(Q_{it}) &= \text{Response; log-volume for store } i \text{ at week } t \\
log(P_{it}) &= \text{Log-price for store } i \text{ at week } t \\
log(\alpha_i) &= \text{Intercept for each store} \\
x_{it} &= \text{Indicator variable for ad display (displayed ad = 1)} \\
log(P_{it}) * x_{it} &= \text{Interaction; shape may change depending on whether ad in store}
\end{aligned}
$$

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

$$
\begin{aligned}
\hat{\tau}^2_\alpha &= 5.0478 \\
\hat{\tau}^2_\beta &= 4.6658 \\
\hat{\tau}^2_\gamma &= 0.9634 \\
\hat{\tau}^2_\theta &= 0.7004 \\
\hat{\sigma}^2 &= 0.06733
\end{aligned}
$$

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

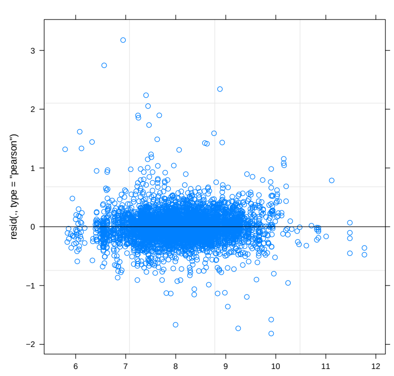

Figure 4: Residual plot for hierarchical model

Model summary:

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

REML criterion at convergence: 1811.9

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

Random effects:
 Groups   Name        Variance Std.Dev. Corr
 store    (Intercept) 5.0478   2.2467
          logP        4.6658   2.1600   -0.94
          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

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

Fully Bayesian Hierarchical Linear Model

Model

Define the following variables.

$$y_i = log(Q), \text{ a } (n_s x1) \text{ vector of log-volume observations for each of the } s \text{ stores.}$$
$$X_i = W_i = \begin{bmatrix} | & | & | & | \\ 1 & logP_{it} & ad_{it} & logP_{it}{*}ad_{it} \\ | & | & | & | \end{bmatrix}, \text{ a } (n_s xp) \text{ matrix of covariates for each store.}$$

where

$$\begin{aligned} s =& \text{ Number of stores. Stores are indexed by } i = \{1, 2, \ldots, s\} \\ n_i =& \text{ Number of observations (weeks) within store } i. \end{aligned}$$

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

$$\begin{aligned} y_i &= X_i\beta + W_i b_i + e_i, \text{ with } e_i \sim N(0, \sigma^2 I_{n_i}) \\ \beta &\sim N_1(\mu_\beta, V_\beta) \\ b_i &\sim N_p(0, \Sigma) \\ \sigma^2 &\sim \frac{1}{\sigma^2} \\ \Sigma &\sim IW(d, C) \end{aligned}$$

Likelihood

$$\begin{aligned} y_i &\sim N_{n_i}(X_i\beta + W_i b_i, \sigma^2 I_{n_i}) \\ y_i &\propto \left(\sigma^2\right)^{-\frac{n_i}{2}} \exp\left[-\frac{1}{2\sigma^2}\left(y_i - X_i\beta - W_i b_i\right)^T \left(y_i - X_i\beta - W_i b_i\right)\right] \\ y_1, \ldots, y_s &\propto \left(\sigma^2\right)^{-\frac{n}{2}} \exp\left[-\frac{1}{2\sigma^2}\sum_{i=1}^{s}\left(y_i - X_i\beta - W_i b_i\right)^T \left(y_i - X_i\beta - W_i b_i\right)\right] \end{aligned}$$

Full Conditionals

<u>$(b_i|\ldots)$</u>

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

We recognize this as the multivariate normal kernel.

$$(b_i|\ldots) \sim N\left(m^*, V^*\right), \text{with} \tag{5}$$
$$V^* = \left[\Sigma^{-1} + \frac{1}{\sigma^2} W_i^T W_i\right]^{-1} \tag{6}$$
$$m^* = V^* \left[\frac{1}{\sigma^2} W_i^T \left(y_i - X_i\beta\right)\right] \tag{7}$$

$\underline{(\beta|\ldots)}$

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

We recognize this as the univariate normal kernel.

$$(\beta|\ldots) \sim N_p\left(m^*, V^*\right), \text{with} \tag{8}$$
$$V^* = \left[V_\beta^{-1} + \left(\frac{1}{\sigma^2}\right)\sum_{i=1}^{s} X_i^T X_i\right]^{-1} \tag{9}$$
$$m^* = V^*\left[V_\beta^{-1}\mu_\beta + \left(\frac{1}{\sigma^2}\right)\sum_{i=1}^{s} X_i^T\left(y_i - W_ib_i\right)\right] \tag{10}$$

$\underline{(\sigma^2|\ldots)}$

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

We recognize this as the inverse gamma kernel.

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

$\underline{(\Sigma|\ldots)}$

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

We recognize this as the Inverse Wishart kernel.

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

The demand curves for the 88 stores are as follows. Stores are ordered in decreasing order by average price.

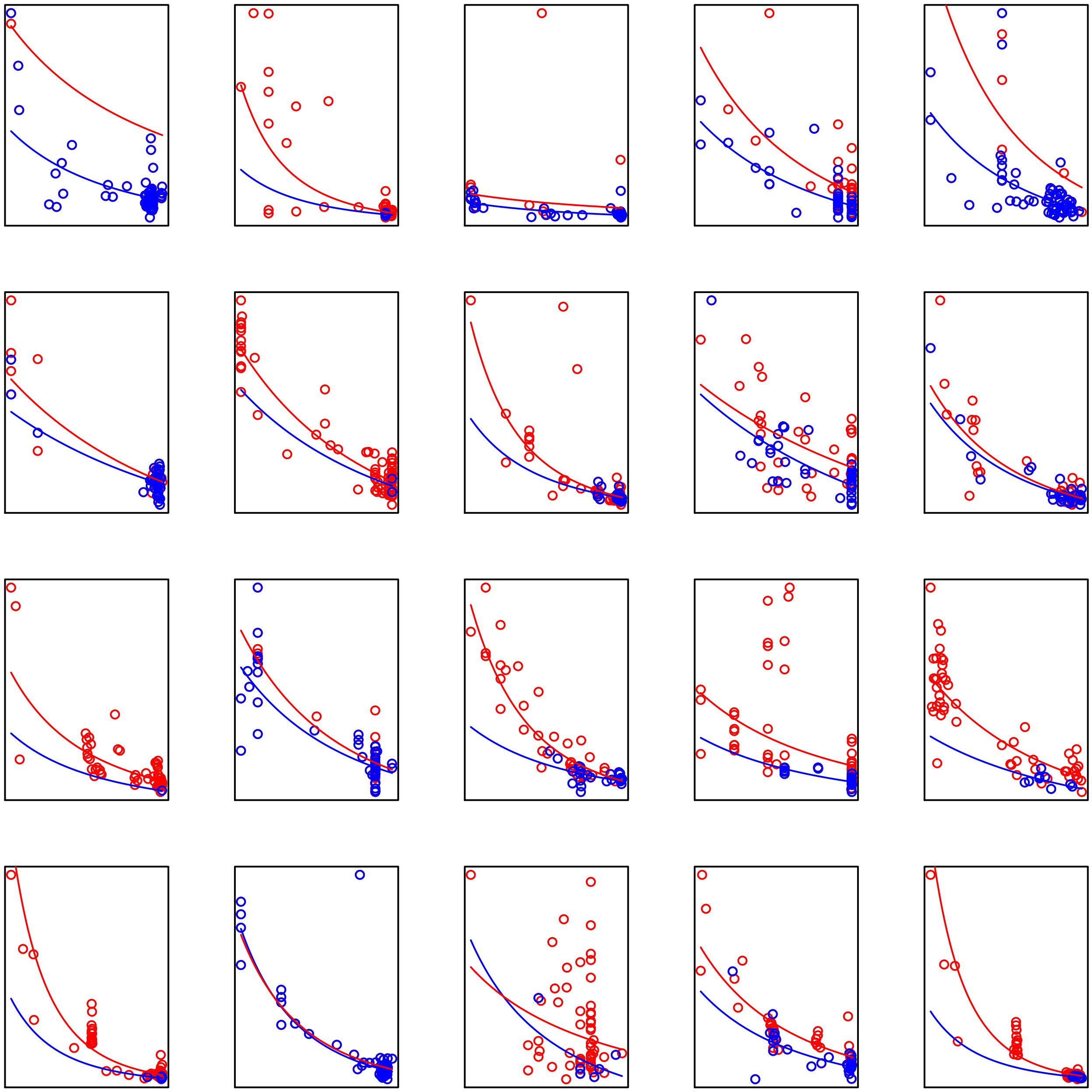

Figure 5: Demand Curves for 88 Stores

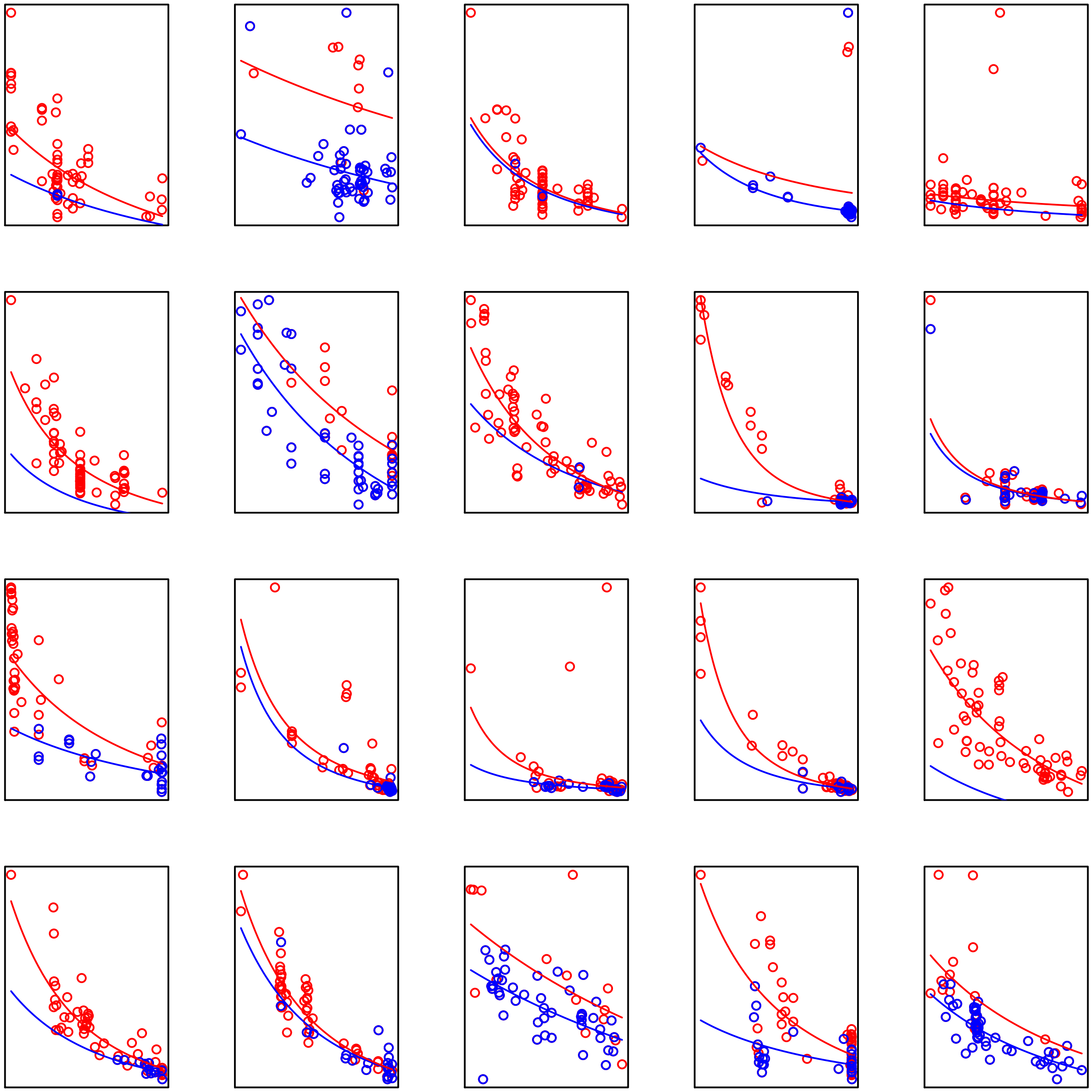

Figure 6: Demand Curves for 88 Stores

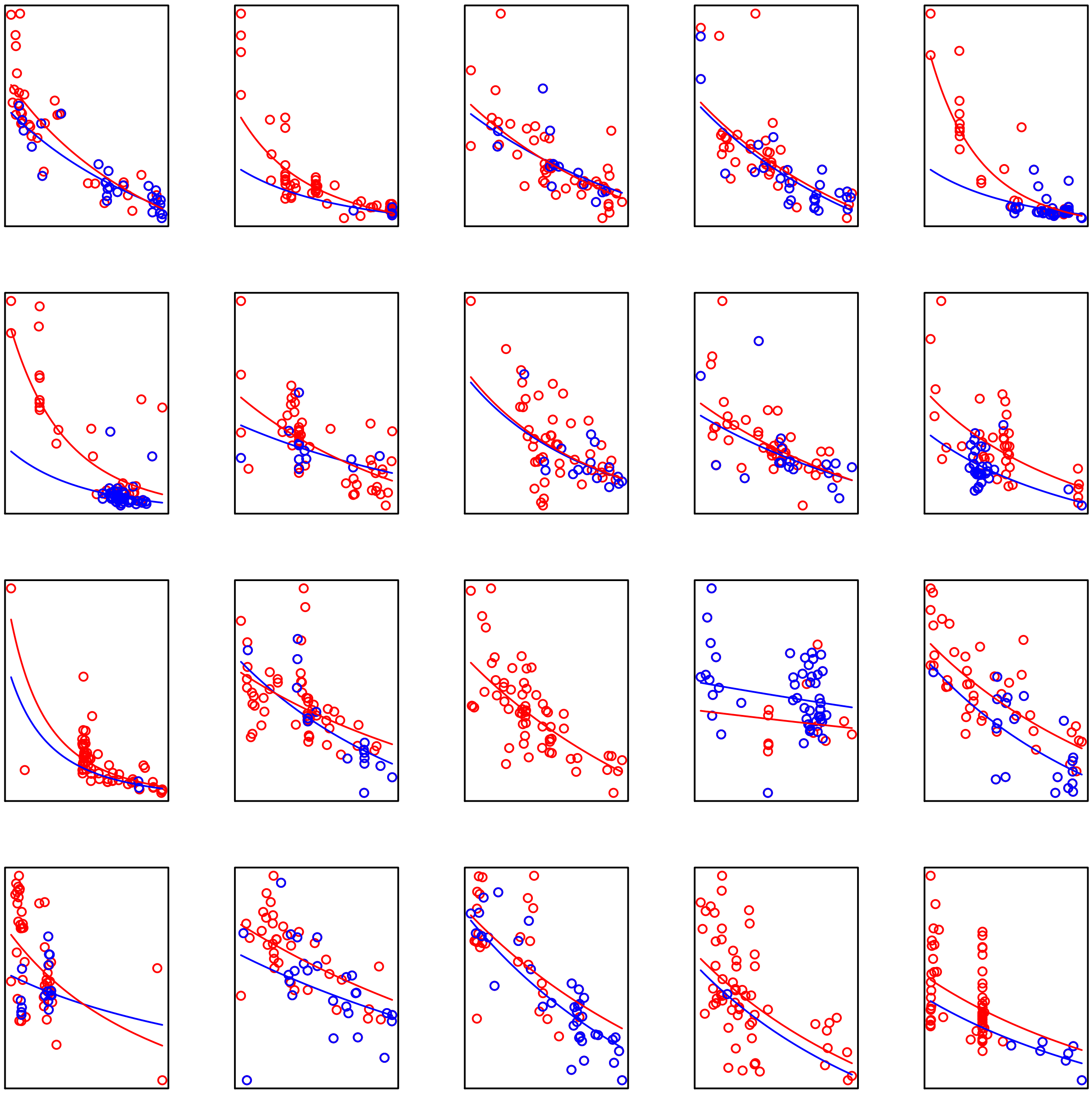

Figure 7: Demand Curves for 88 Stores

A Hierarchical Probit Model via Data Augmentation

Gene Expression Over Time