1 Hierarchical models: data-analysis problems

1.1 Price elasticity of demand

The data in "cheese.csv" are about sales volume, price, and advertisting display activity for packages of Borden sliced "cheese." The data are taken from Rossi, Allenby, and McCulloch's textbook on *Bayesian Statistics and Marketing.* For each of 88 stores (store) in different US cities, we have repeated observations of the weekly sales volume (vol, in terms of packages sold), unit price (price), and whether the product was advertised with an in-store display during that week (disp = 1 for display).

Your goal is to estimate, on a store-by-store basis, the effect of display ads on the demand curve for cheese. A standard form of a demand curve in economics is of the form $Q = \alpha P^{\beta}$, where Q is quantity demanded (i.e. sales volume), P is price, and α and β are parameters to be estimated. You'll notice that this is linear on a log-log scale,

$$\log P = \log \alpha + \beta \log Q$$

which you should feel free to assume here. Economists would refer to β as the price elasticity of demand (PED). Notice that on a log-log scale, the errors enter multiplicatively.

As you can see in figure 1, there appears to be an effect caused by the variable disp, which indicates whether the store was displaying an add for the cheese that week. Observations where there was a display, indicated in pink, are shifted up and to the left of observations where there was no display, indicated in blue. This suggests that more cheese is sold when there is a display, and at a lower price.

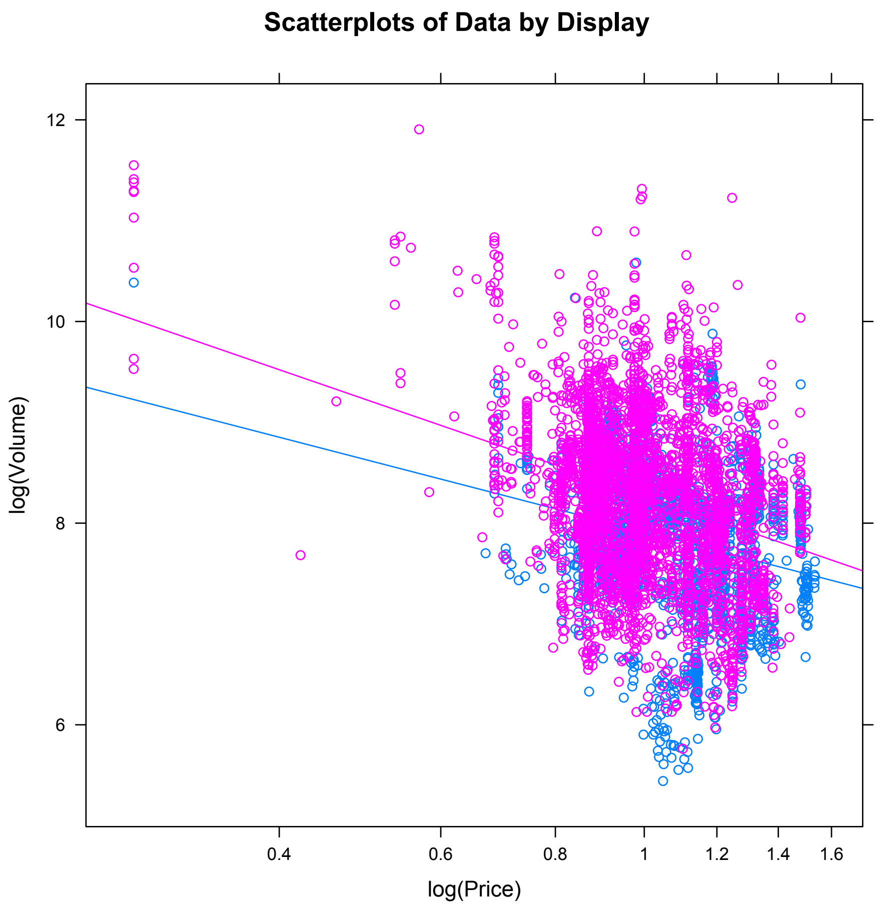

Figure 1: Scatterplot of log(Price) by log(Volume) for all data, stores displaying an advertisement are in pink while those without a display are in blue

Looking at figure 2, which looks at total sales, sales appear higher on average than if there is no display.

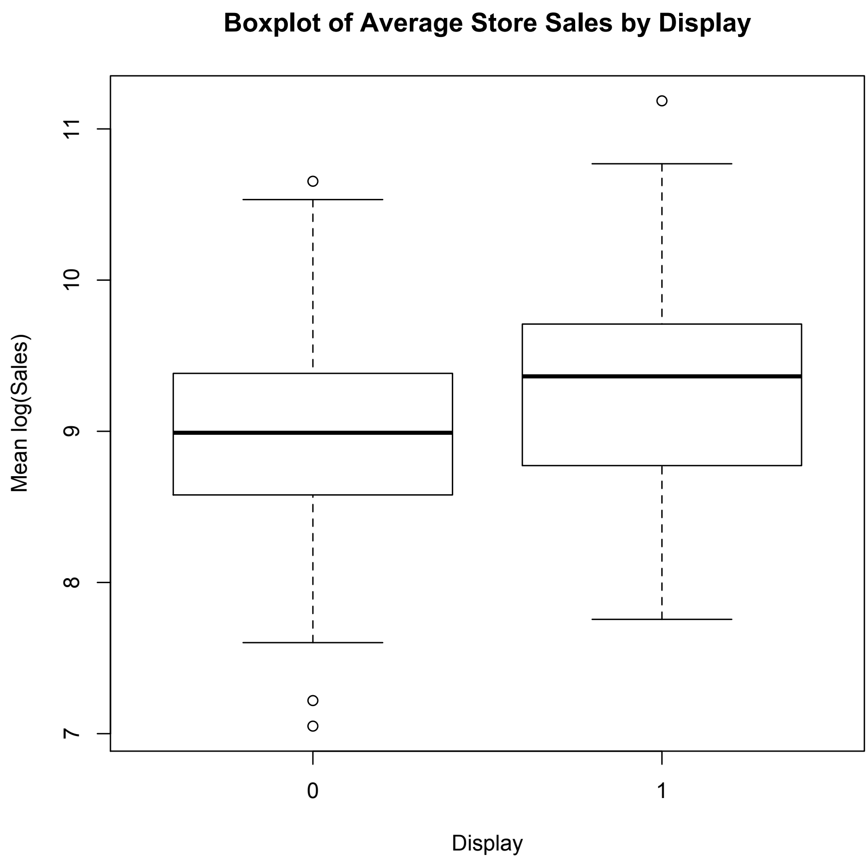

Figure 2: Boxplot of average log(sales) by display. Average log(sales) is calculated by calculating the log sales for each data point, than taking the average by store by display.

Figure 3 looks at the differences by store. It is clear that the data vary by store. It is also apparent that some stores do no have enough data by display to support a model on their own, and would benefit from the stabilizing effect of the other observations on their estimates. These images together indicate that it would be beneficial to fit a model for log(volume) that can vary by store, and includes a display effect in addition to log(price).

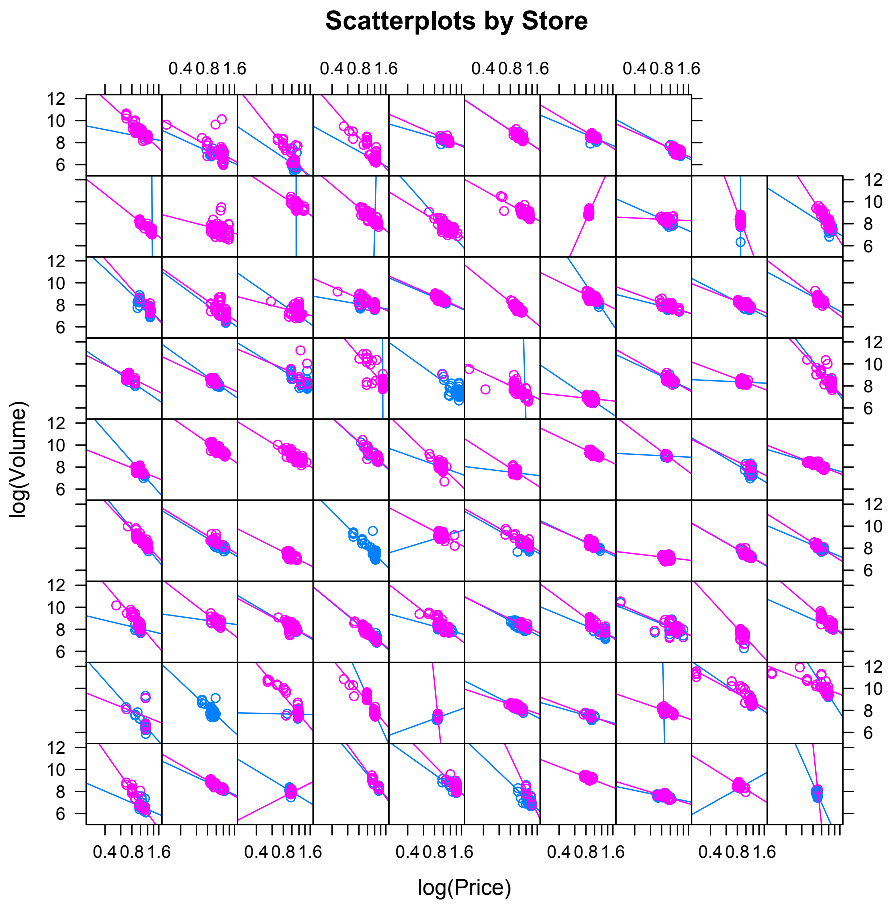

Figure 3: Scatterplot of log(Price) by log(Volume) for each store, stores displaying an advertisement are in pink while those without a display are in blue

We fit the following model:

$$y_{ij} = x_{ij}^T \beta_i z_i + \varepsilon_{ij}$$

where:

- y_{ij}: log volume
- z_i: is 1 if the data comes from store i, 0 otherwise
- x_{ij}: the covariates of the model for observation ij, namely an intercept, log price, display, and the interaction of log price and display
- β_i: store dependent coefficients
- ε: error term
- i: store index, $i = 1, \ldots, 88$ as there are 88 stores
- j: observation index, $j = 1, \ldots, n_i$ where n_i is the number of observations for store i, $\sum_{i=1}^{88} n_i = N$

First, a restricted maximum likelihood model was fit in R using the lme4 package. Figure 4 shows the fit for two randomly chosen stores.

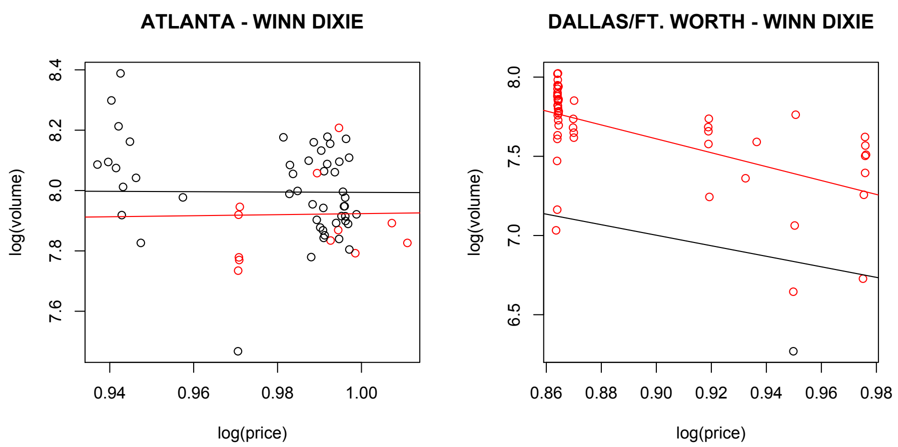

Figure 4: Plot of fit for two stores, points with a display, and there fitted line, are in red, while points without a display, and their fitted line, are in black

Next, we fit a hierarchical bayesian model to the data. We specify the mode as follows:

$$\begin{aligned} y_{ij} &\sim N(\mathbf{x_{ij}}^T \boldsymbol{\beta_i}, \sigma^2) \\ \boldsymbol{\beta_i} &\sim N(\boldsymbol{\theta}, \boldsymbol{\Sigma}) \end{aligned}$$

with the following priors:

$$\begin{aligned} p(\sigma^2) &\propto \frac{1}{\sigma^2} \\ \boldsymbol{\theta} &\sim N(\boldsymbol{\theta_0}, \mathbf{I}) \\ \boldsymbol{\Sigma} &\sim InvWish(4, \mathbf{I}) \end{aligned}$$

I take $\boldsymbol{\theta_0}$ to be the coefficients from a normal linear regression on all of the data, ie $y_{ij} = \mathbf{x_{ij}}^T \boldsymbol{\beta} + \varepsilon$.

I constructed a gibbs sampler using the following full conditionals:

$$\sigma^2 | \cdots \sim InvGamma\left(\frac{N}{2} \ , \ \frac{1}{2} \sum_{i=1}^{88} \sum_{j=1}^{n_i} (y_{ij} - \mathbf{x_{ij}}^T \boldsymbol{\beta_i}) \right)$$

$$\boldsymbol{\theta} | \cdots \sim N\left((\mathbf{I} + 88\boldsymbol{\Sigma}^{-1})^{-1} \left(\boldsymbol{\theta_0} + \boldsymbol{\Sigma}^{-1} \sum_{i=1}^{88} \boldsymbol{\beta_i} \right) \ , \ (\mathbf{I} + 88\boldsymbol{\Sigma}^{-1})^{-1} \right)$$

$$\boldsymbol{\Sigma} | \cdots \sim InvWish\left(\mathbf{I} + \sum_{i=1}^{88} (\boldsymbol{\beta_i} - \boldsymbol{\theta})(\boldsymbol{\beta_i} - \boldsymbol{\theta})^T \ , \ 92 \right)$$

$$\boldsymbol{\beta_i} | \cdots \sim N\left(\left(\boldsymbol{\Sigma}^{-1} + \sigma^2 \sum_{j=1}^{n_i} \mathbf{x_{ij}} \mathbf{x_{ij}^T} \right)^{-1} \left(\boldsymbol{\Sigma}^{-1} \boldsymbol{\theta} + \sigma^2 \sum_{j=1}^{n_i} y_{ij} \mathbf{x_{ij}} \right) \ , \ \left(\boldsymbol{\Sigma}^{-1} + \sigma^2 \sum_{j=1}^{n_i} \mathbf{x_{ij}} \mathbf{x_{ij}^T} \right)^{-1} \right)$$

Figure 5 shows the fit of the Bayesian model for the same two stores examined for the restricted maximum likelihood model.

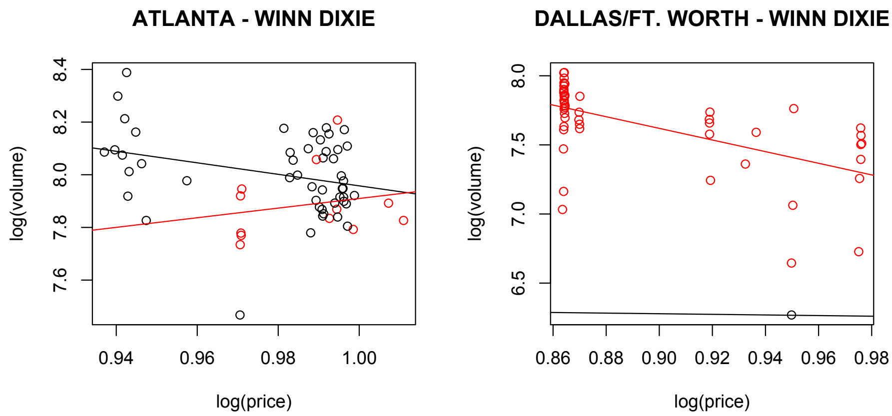

Figure 5: Plot of fit for two stores, points with a display, and there fitted line, are in red, while points without a display, and their fitted line, are in black