Ch 4: Cheese data analysis

Friday, April 5, 2019 8:49 PM

The cheese dataset contains information about the sales volume, price, and advertising display activity. The goal is to estimate, on a store by store basis, the effect of display ads on the demand curve for cheese. The standard form for the demand curve is

$$Q = \alpha P^{\beta} \qquad \begin{array}{l} Q = \text{quantity demanded} \\ \alpha, \beta = \text{parameters to estimate} \end{array}$$

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

Notice that on a log-log scale, the errors enter multiplicatively. Things to consider:

1.) The demand curve might shift (different α) & change shape (different β), depending on whether there is a display ad or not.

2.) Different stores will have very different typical volumes & the model should account for this.

3.) Do different stores have different PEDs?

4.) If there is an effect on the demand curve due to showing a display ad, does this effect differ by store?

To address all the above issues, I propose the following model:

$$y_{ij} = \beta_{i0} + \beta_{i1} \log(\text{price}_{ij}) + \beta_{i2}\, \mathbb{1}(\text{display}_{ij}=1) + \beta_{i3} \log(\text{price}_{ij}) \times \mathbb{1}(\text{display}_{ij}=1) + \varepsilon_i \quad , \quad \varepsilon_i \sim N(0, \sigma^2)$$

$$= \log(\alpha_i) + \beta_{i1} \log P_{ij} + \beta_{i2}\, \mathbb{1}(D_{ij}=1) + \beta_{i3} \log P_{ij} \times \mathbb{1}(D_{ij}=1) + \varepsilon_i$$

$$y_i \sim N(\log \alpha_i + \beta_{i1} \log P_i + \beta_{i2}\, \mathbb{1}(D_i=1) + \beta_{i3} \log P_i \times \mathbb{1}(D_i=1), \sigma^2)$$

This model gives each store its own default intercept ($\log \alpha_i$) and default slope (β_{i1}). The other two parameters account for the presence of a display and whether having a display changes the price elasticity of the model. When there is a display, both the intercept and the slope of this log linear fit will change:

$$y_{ij} = (\log(\alpha_i) + \beta_{i2}) + (\beta_{i1} + \beta_{i3}) \log(\text{price}_{ij}).$$

The hierarchy of this model is

$$\begin{tikzcd}[column sep=small]
 & & \tilde{\mu}, \Sigma \arrow[dll] \arrow[dl] \arrow[d] \arrow[dr] & \\
\tilde{\beta}_1 \arrow[d] & \tilde{\beta}_2 \arrow[d] \;\cdots & \tilde{\beta}_{n-1} \arrow[d] & \tilde{\beta}_n \arrow[d] \\
\tilde{y}_1 & y_2 \;\cdots & \tilde{y}_{n-1} & y
\end{tikzcd}$$

where $\tilde{\mu}$ is the overall mean between groups and Σ is the covariance matrix.

To set up my MCMC, I used the following parameters and full conditionals.

let $\tilde{\beta}_i = (\beta_{i0}, \beta_{i1}, \beta_{i2}, \beta_{i3})^T$

$\tilde{y}_i$ = log quantity observations for store i

$$X_i = \begin{bmatrix} 1 & \log(P_{i1}) & \mathbb{1}(D_{i1}=1) & \log(P_{i1}) \times \mathbb{1}(D_{i1}=1) \\ \vdots & \vdots & \vdots & \vdots \\ 1 & \log(P_{im}) & \mathbb{1}(D_{im}=1) & \log(P_{im}) \times \mathbb{1}(D_{im}=1) \end{bmatrix}$$

Then $\tilde{y}_i = X_i \tilde{\beta}_i + \varepsilon_i$, $\varepsilon_i \sim N(0, \sigma^2 I) \Rightarrow \tilde{y}_i \mid \tilde{\beta}_i \sim N(X_i \tilde{\beta}_i, \frac{1}{\lambda} I)$ where $\lambda = \frac{1}{\sigma^2}$

Let $\tilde{\beta}_i \sim MVN(\mu, \Sigma)$ with priors $p(\mu) \propto [1, \ldots, 1]^T$, an improper noninformative prior

$$p(\Sigma) \propto \text{Inverse Wishart}(2, I_4)$$

$$p(\lambda) \propto Ga(\tfrac{1}{2}, \tfrac{1}{2})$$

The full posterior is:

$$P(\mu, \Sigma, \sigma^2, \tilde{\beta}_{1:n} \mid \tilde{y}_{1:n}) \propto p(\tilde{\beta}_{1:n} \mid \mu, \Sigma, \sigma^2)\, p(\tilde{y}_{1:n} \mid \tilde{\beta}_{1:n}, \mu, \Sigma, \sigma^2)\, p(\mu) p(\Sigma) p(\sigma^2)$$

$$\propto \left[\prod_{i=1}^{n} \frac{1}{|\Sigma|^{1/2}} \exp\left(-\tfrac{1}{2} (\tilde{\beta}_i - \mu)^T \Sigma^{-1} (\tilde{\beta}_i - \mu)\right)\right] \left[\prod_{i=1}^{n} \prod_{j=1}^{m} \lambda^{1/2} \exp\left(-\tfrac{\lambda}{2} (y_{ij} - \underset{\text{jth row of } X_i}{x_{ij}} \tilde{\beta}_i)^2\right)\right] |\Sigma|^{-\frac{(v+p+1)}{2}} \exp(-\tfrac{1}{2} \operatorname{tr}(\Psi \Sigma^{-1})) \cdot \lambda^{1-1} \exp(-\lambda)$$

The full conditionals are:

$$P(\mu \mid \text{——}) \propto \exp\Big(-\tfrac{1}{2} \underbrace{\sum_{i=1}^{n} (\tilde{\beta}_i - \mu)^T \Sigma^{-1} (\tilde{\beta}_i - \mu)}\Big) \cdot 1$$

$$\sum_{i=1}^{n} \tilde{\beta}_i \Sigma^{-1} \tilde{\beta}_i - 2\mu^T \Sigma^{-1} \tilde{\beta}_i + \mu^T \Sigma^{-1} \mu^T$$

$$= \mu^T n \Sigma^{-1} \mu - 2\mu^T n \Sigma^{-1} \bar{\tilde{\beta}} + \sum_{i=1}^{n} \tilde{\beta}_i \Sigma^{-1} \tilde{\beta}_i$$

$$\propto \exp\left(-\tfrac{1}{2} (\mu - \bar{\tilde{\beta}})^T n \Sigma^{-1} (\mu - \bar{\tilde{\beta}})\right)$$

$$\sim MVN(\bar{\tilde{\beta}}, (n\Sigma^{-1})^{-1})$$

$$\propto \exp\left(-\tfrac{1}{2}(\mu-\bar{\tilde{\beta}})^T n\Sigma^{-1}(\mu-\bar{\tilde{\beta}})\right)$$

$$\sim MVN(\bar{\tilde{\beta}}, (n\Sigma^{-1})^{-1})$$

$$p(\Sigma \mid —) \propto |\Sigma|^{-n/2}\exp\left(-\tfrac{1}{2}\sum_{i=1}^{n}(\tilde{\beta}_i-\mu)^T\Sigma^{-1}(\tilde{\beta}_i-\mu)\right)|\Sigma|^{\frac{-(v+p+1)}{2}}\exp\left(-\tfrac{1}{2}tr(\Psi\Sigma^{-1})\right)$$

$$\propto |\Sigma|^{\frac{-(v+p+n)-1}{2}}\exp\left(-\tfrac{1}{2}\,trace\left(\Sigma^{-1}\left(\Psi+\sum_{i=1}^{n}(\tilde{\beta}_i-\mu)(\tilde{\beta}_i-\mu)^T\right)\right)\right)$$

$$p(\tilde{\beta}_i \mid —) \propto \exp\left(-\tfrac{1}{2}(\tilde{y}_i-X_i\tilde{\beta}_i)^T\lambda I(\tilde{y}_i-X_i\tilde{\beta}_i)\right)\exp\left(-\tfrac{1}{2}(\tilde{\beta}_i-\mu)^T\Sigma^{-1}(\tilde{\beta}_i-\mu)\right)$$

$$\propto \exp\left(-\tfrac{1}{2}(\tilde{\beta}_i^TX_i^T\lambda IX_i\tilde{\beta}_i-2\tilde{\beta}_i^TX_i^T\lambda I\tilde{y}_i)-\tfrac{1}{2}(\tilde{\beta}_i^T\Sigma^{-1}\tilde{\beta}_i-2\tilde{\beta}_i^T\Sigma^{-1}\mu)\right)$$

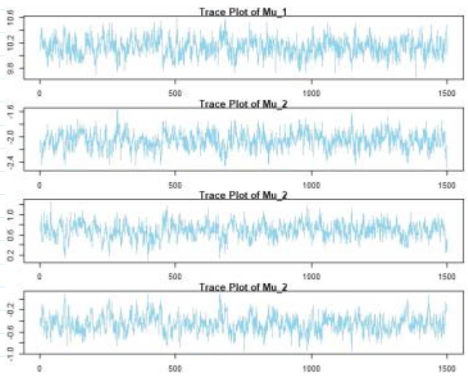
$$\propto \exp\left(-\tfrac{1}{2}\left(\tilde{\beta}_i^T(X_i^T\lambda IX_i+\Sigma^{-1})\tilde{\beta}_i+2\tilde{\beta}_i^T(X_i^T\lambda I\tilde{y}_i+\Sigma^{-1}\mu)\right)\right)$$

$$\sim MVN\left((\lambda X_i^TX_i+\Sigma^{-1})^{-1}(\lambda X_i^T\tilde{y}_i+\Sigma^{-1}\mu),\ (\lambda X_i^TX_i+\Sigma^{-1})^{-1}\right)\ \lambda$$

$$p(\sigma^2 \mid —) \propto \lambda^{\frac{mn}{2}}\exp\left(-\tfrac{\lambda}{2}\sum_{i=1}^{n}\sum_{j=1}^{m}(y_{ij}-X_{ij}^T\tilde{\beta}_i)^2\right)\lambda^{a-1}\exp(-\lambda b)$$

$$\sim Gamma\left(\tfrac{mn}{2}+a,\ b+\tfrac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{m}(y_{ij}-\underset{88\times4}{X_{ij}^T}\underset{4\times1}{\tilde{\beta}_i})^2\right)$$

After running this chain for 2000 iterations and discarding the first 500 as burn in I have the following results. Below, I have the trace plots for μ, the over all between group means of the βs, and the trace plot of σ^2, the variance of the y_{ij}'s. The mixing is reasonable which suggests that the chain has reached stationarity.



The trace plot of the μ vector suggests that:

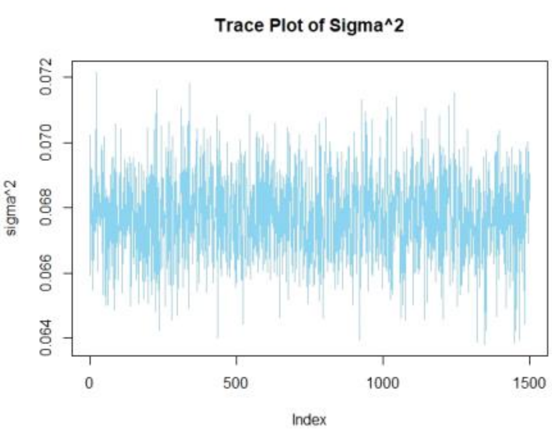
μ_1, the mean for β_0 or $\log\alpha \approx 10$

μ_2, the mean for $\beta_1 \approx -2.2$

μ_3, the mean for $\beta_2 \approx 0.6$

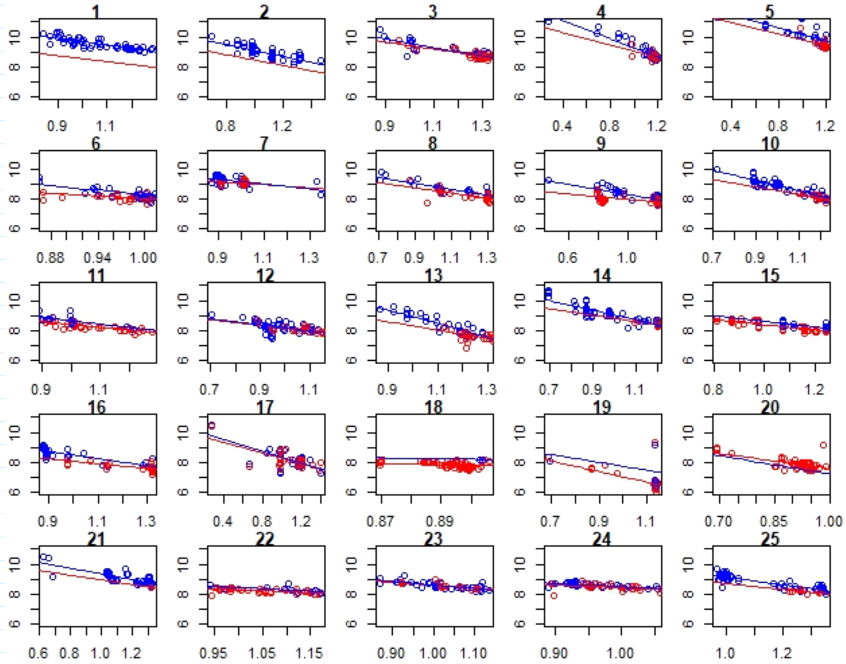
μ_4, the mean for $\beta_3 \approx -0.5$

These values are in line with our expectations. We can infer from these results that as price increases, the volume sold will decrease. Having a display results in more sales but makes people more price sensitive.



The trace of σ^2 suggests there is not much noise in this model since the values do not fluctuate in a wide range.

Below is the predictive plots for the first 25 stores. The red points and lines in each plot represent no display and the blue points and lines are for display. The lines were fitted using the posterior mean of the β's for each store.



The predictive lines are in line with our previous analysis. The blue display lines have higher interceps but steeper downward slopes. This indicates a display results in more volume sold but also more negative price elasticity. Another point to notice is that even when there are few or even no points for either display or no display, the hierachical model still generates reasonable parameters based on information generalized from other stores. In general, when a group does not have enough data, the parameters will be shrunk towards the overall mean for the parameter.