# Hierarchical Bayesian Model

1

## Content

- Review of Heterogeneity
- Random effects with regressors

2

## Panel Data in Marketing

Large m (number of cross-sectional units)

stores, consumers, markets, physicians, accounts

small T (short interval of observation)

discrete dep vars

interest in decentralized decision-making

e.g. store-sku pricing, account-level promo strategies

unit-level parameters of interest! (contrast with econ.)

unit-level parms measured with high degree of uncertainty! Prior will matter!

Heterogeneity

3

## Heterogeneous logit model

MNL model likelihood

$$\Pr(y_{hti} = 1) = \frac{\exp(X_{hti}\beta_{hi})}{\sum_j \exp(X_{htj}\beta_{hj})}$$

$$\beta_h \sim N(\bar{\beta}, V_\beta)$$

$$\bar{\beta} \sim N(b_0, S_0) \qquad V_\beta \sim IW(\nu_0, V_0)$$

The posterior:

$$p(\{\beta_h\}, \bar{\beta}, V_\beta) \propto \prod_{h=1}^{N} \left( \prod_{h=1}^{T_h} \underbrace{[y_{ht} | X_{ht}, \beta_h]}_{\text{logit model}} \right) \underbrace{[\beta_h | \bar{\beta}, V_\beta]}_{\text{normal heterogeneity}} \underbrace{[\bar{\beta}\,][V_\beta]}_{\text{priors}}$$

4

## Heterogeneous logit model

Priors:

$$\beta_h \sim N(\bar{\beta}, V_\beta)$$

$$\bar{\beta} \sim N(\bar{\bar{\beta}}, A)$$

$$V_\beta \sim IW(\upsilon, \upsilon I)$$

$$\begin{tikzcd} \bar{\beta} \arrow[dr] & & \\ & \{\beta_h\} \arrow[r] & y \\ V_\beta \arrow[ur] & & \end{tikzcd}$$

GS:

$$\beta_h \,\middle|\, \bar{\beta}, V_\beta$$

$$\bar{\beta}, V_\beta \,\middle|\, \{\beta_h\}$$



## Drawing $\beta_h$

Use RW Metropolis:

$V_\beta$

$$\beta_h^{cand} = \beta_h^{old} + \varepsilon, \varepsilon \sim N(0, ?)$$

$$\alpha(\beta_h^{cand}, \beta_h^{old}) = \min\left\{1, \frac{\pi(\beta_h^{cand})}{\pi(\beta_h^{old})}\right\}$$

Increment Cov matrix: One simple idea is just to use the prior – assumes unit likelihoods are relatively uninformative

$$\pi(\beta^{cand}) = \left( \prod_{t=1}^{T_h} \frac{\exp[x_{iht}'\beta^{cand}]}{\sum_j \exp[x_{jht}'\beta^{cand}]} \right) \times \exp\left( \frac{-1}{2} (\beta^{cand} - \bar{\beta})' V_\beta^{-1} (\beta^{cand} - \bar{\beta}) \right)$$

## Random effects with regressors

$\Delta$ is a matrix of regression coefficients related covariates (Z) to mean of random-effects distribution.

$z_h$ are covariates for respondent h

$$b_h = D' z_h + u_h \text{ or } B = ZD + U$$
$$U \sim N(0, V_b)$$

7

## Hierarchical Linear Model

Consider a regression:

$$y_{it} = X_{it}\beta_i + \epsilon_{it} \qquad \epsilon_{it} \sim N(0, \sigma^2)$$

$$\beta_i = \Delta' z_i + v_i \quad v_i \sim \text{iid}N(0, V_\beta)$$ Tie together via Prior

or

$$B = ZD + V \quad B = \begin{bmatrix} b_1' \\ \vdots \\ b_m' \end{bmatrix} \quad Z = \begin{bmatrix} z_1' \\ \vdots \\ z_m' \end{bmatrix} \quad D = \begin{bmatrix} d_1 & \cdots & d_k \end{bmatrix} \quad v_i' \sim N(0, V_b)$$

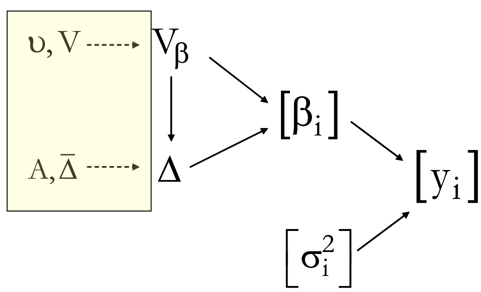
8

# Priors

$$V_\beta \sim IW(\nu, V)$$

$$vec(\Delta)\,|\,V_\beta \sim N\left(vec(\bar{\Delta}), V_\beta \otimes A^{-1}\right)$$

$$\sigma_i^2 \sim \nu_i s_{0,i}^2 / \chi^2_{\nu_i}$$




9

# Posterior

Posterior is the joint distribution of all model parameters:
$\{\beta_i\ \forall i\}, \sigma^2, \Delta, V_\beta$

Joint likelihood of all these model parameters is

$$(y|\beta, \sigma^2) \times (\beta|\Delta, V_\beta)$$

$(y|\beta)$ is a normal regression

$$\frac{1}{2\pi\sigma} \exp\left(-\frac{1}{2}\frac{(y - x\beta)^2}{\sigma^2}\right)$$

$(\beta|\Delta, V_\beta)$ is also a normal regression

$$\frac{1}{\sqrt{(2\pi)^k |V_\beta|}^{\,M}} \exp\left(-\frac{1}{2}\sum_{i=1}^{M}(\beta_i - Z_i\Delta)V_\beta^{-1}(\beta_i - Z_i\Delta)'\right)$$

10

## GS for the Hierarchical Linear Model

$\beta_i \mid \sigma_i^2, \Delta, V_\beta, y_i, X_i$

Univariate Regression with an *informative* prior!

$\sigma_i^2 \mid \beta_i, y_i, X_i$

note independence from hierarchical parms!

$\Delta, V_\beta \mid \{\beta_i\}, Z$

rmultireg with $\{\beta_i\}$ as data and Z as "X"

implemented in rhierLinearModel



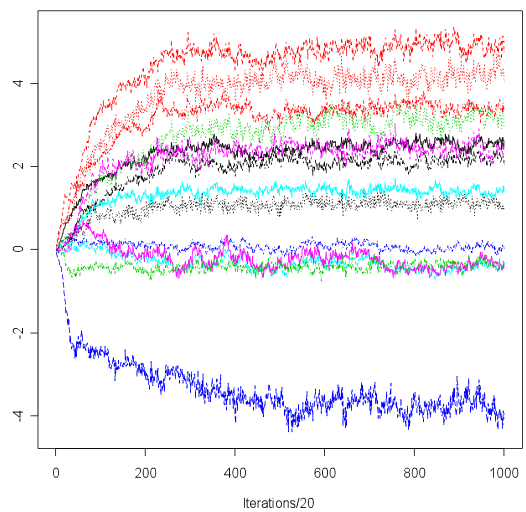
## Adaptive Shrinkage

With fixed values of $\Delta, V_\beta$, we have m independent Bayes regressions with informative priors.

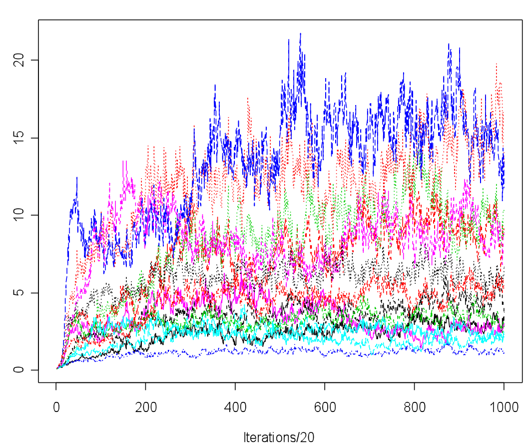
In the hierarchical setting, we "learn" about the location and spread of the $\{\beta_i\}$.

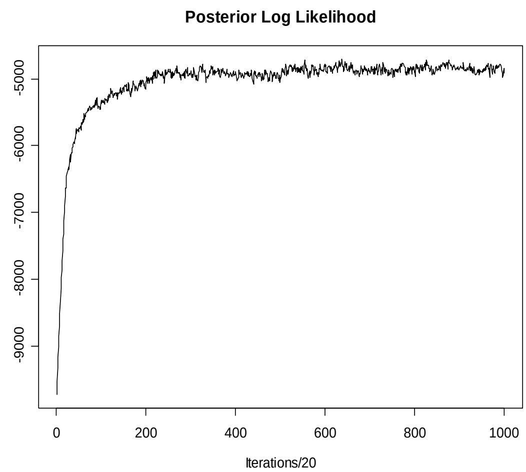
The extent of shrinkage, for any one unit, depends on dispersion of betas across units and the amount of information available for that unit.

## Average Respondent Part-Worths




## V-beta Draws




Posterior Log Likelihood
-5000
-6000
-7000
-8000
-9000
0
200
400
600
800
1000
Iterations/20