

Hierarchical Modeling

Tuesday, September 25, 2012 10:50 AM

- Previous lecture on simple Hierarchical Linear Modeling

arm (Gelman & Hill)

- Goal today: use Bayesian methodology to construct hierarchical models in the GLM framework

What is a Hierarchical Model?

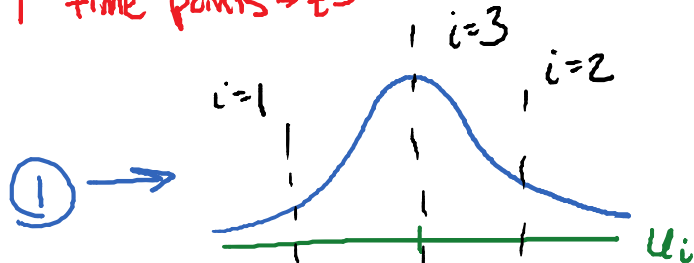
Tuesday, September 25, 2012 11:27 AM

- A hierarchical model is a model with a hierarchy (har har)

- The hierarchy in question is a hierarchy of parameters which are related to one another via unidirectional path-like relations

- Example: simple random effects model

N units $\rightarrow i$
T time points $\rightarrow t$ } TSCS



$$y_{it} = \beta_0 + \beta_1 X_{it} + \Sigma_{it}$$

unit heterogeneity

$$y_{it} = \beta_0 + \beta_1 X_{it} + \gamma_i + u_{it}$$

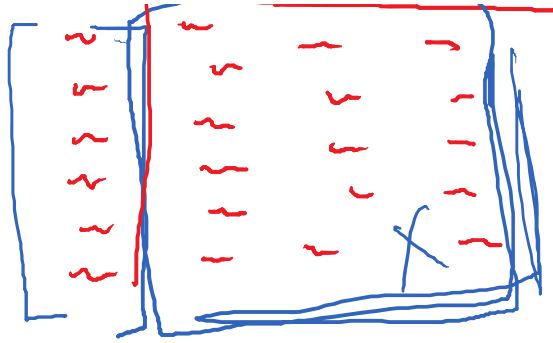
FE: create dummy vars for each unit i

$$RE: \gamma \sim \phi(0, \sigma_\gamma^2)$$

$$\int f(\beta | \gamma) f(\gamma) d\gamma = f(\beta)$$

$\hat{\beta}$	$\hat{\gamma}_1$	$\hat{\gamma}_2$	$\hat{\gamma}_3$
~	~	~	~
~	~	~	~

$$E[\beta]$$

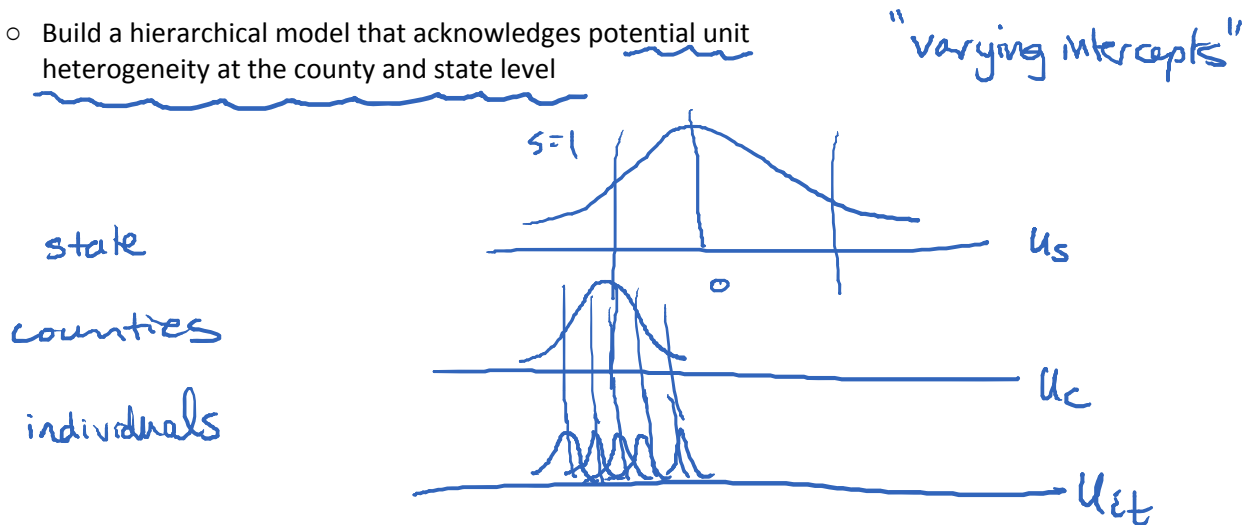


- Use Bayesian methodology to construct this model, then recover posterior densities for parameters of interest using Monte Carlo integration to integrate out incidental parameters

Example: Hierarchical Logit Model of Vote Choice

Tuesday, September 25, 2012 11:31 AM

- Consider the following situation:
 - Polling data for 4 states
 - Between 10 and 60 counties in each state
 - Between 5 and 500 respondents in each county
 - Build a hierarchical model that acknowledges potential unit heterogeneity at the county and state level



- Questions to consider:
 - How likely is each person to vote for the incumbent?

~~~~~

- What proportion of each state will vote for the incumbent?

~~~~~

|| ≠

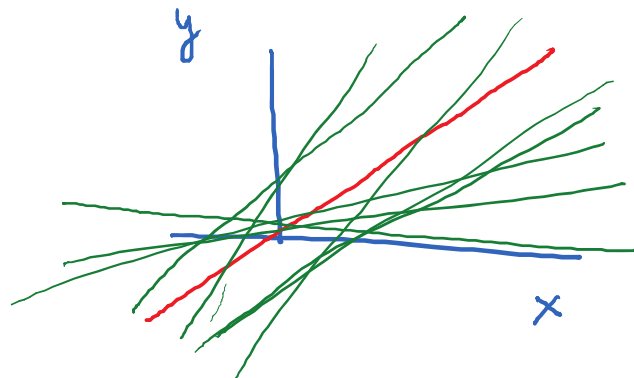
- What proportion of each county will vote for the incumbent?

|| 29

Example: Hierarchical Model of Corruption

Tuesday, September 25, 2012 4:36 PM

- Consider the following scenario, described in my working paper with Gina Chirillo:
 - TSCS data set: multiple countries observed at multiple times
 - Primary DV: WBGI control of corruption measure
 - Primary IVs:
 - % Women in Government
 - POLITY IV
 - Idea: see whether relationship between Corruption and % Women is different in different institutional contexts (viz. ,POLITY scores)
 - POLITY is a proxy for democratic institutions, which we believe disincentivize corruption
- Construct a hierarchical model with...
 - Varying intercepts for region
 - Varying slopes on % Women, as a function of POLITY



Missing Data Imputation with a Bayesian Model

Tuesday, September 25, 2012 5:58 PM

- If some data is missing in a model, we can systematically impute that data in our Bayesian model and use it in the likelihood, rather than simply discarding these cases
- Assume data is Missing (Conditionally) at Random: $\Pr(\text{miss} = 1|Y_{\text{miss}}, Y_{\text{obs}}) = \Pr(\text{miss} = 1|Y_{\text{obs}})$
- $f(\beta|Y_{\text{obs}}) = \int f(\beta|Y_{\text{obs}}, Y_{\text{miss}})f(Y_{\text{miss}}|Y_{\text{obs}})dY_{\text{miss}}$

- If data is not MAR, integrating out missing values will require integrating out the pattern of missingness using some kind of model:

$$f(Y_{\text{miss}}|Y_{\text{obs}}) = \int f(Y_{\text{miss}}|Y_{\text{obs}}, R) \Pr(R|Y_{\text{miss}})$$

where R is a missingness indicator. In short, will need to add a model of missingness to the overall Bayesian model.

- Upshot: can create an "imputation model" inside of your Bayesian model to recover the data