Bayesian hierarchical models in Stata

Nikolay Balov

StataCorp LP

2016 Stata Conference
Why hierarchical models?

- Hierarchical models represent complex, multilevel data structures.
- Examples:
  - Predict the risk of death after surgery for a group of hospitals and then rank the hospitals according to their performance
  - Estimate the rate of weight gain in children from a panel data of different age groups
  - Estimate student abilities based on their performance on a test panel of different questions
- I will apply a Bayesian approach to answer this kind of questions.
Why Bayesian hierarchical models?

- Bayesian models combine prior knowledge about model parameters with evidence from data.
- They are especially well suited for analysis of multilevel models:
  - Flexibility in specifying multilevel structures of parameters using priors
  - Ability to handle small samples and model misspecification (overparametrization of the likelihood can be resolved with well chosen priors).
  - Provide intuitive and easy to interpret answers. (credible interval vs. confidence interval).

- Some challenges of the Bayesian approach:
  - Computational burden of simulating posterior distributions with many parameters
  - Difficulties in specifying prior distributions; potential **subjectivity in selecting priors**.
Main problem of interest

I will focus on

prior specification and efficient simulation of model parameters associated with grouping variables ("random-effects" parameters).

This methodological problem is at the heart of multilevel (hierarchical) modeling.
Outline

- Motivating example: Hospital ranking
- Overview of Bayesian analysis in Stata
- Bayesian multilevel models
  - Sources of hierarchy in data
  - Hierarchical prior structures involving random-effects (RE)
  - Efficient MCMC sampling of RE parameters
- Analysis of the hospital ranking problem
  - Completely uninformative prior
  - Weakly informative prior
  - Hierarchical prior
  - Model comparison
- Other hierarchical model examples
  - Random-slope with unstructured covariance
  - Weight gain in children: Growth curve model
  - Federal interest rates: Gaussian 2-mixture model
  - Educational research example: 3PL IRT model
Motivating example: Hospital ranking

Mortality rate after cardiac surgery in babies from 12 hospitals (WinBUGS).

```
. input hospital n_ops deaths
  1  47  0
  2 148  18
  3 119  8
  4 810  46
  5 211  8
  6 196  13
  7 148  9
  8 215  31
  9 207  14
 10  97  8
 11 256  29
 12 360  24
```

- Estimate the risk of death in each hospital
- Rank hospitals according to their risk probabilities
Hospital ranking: Frequentist approach

The likelihood model is

\[ \text{deaths}_i \sim \text{Binomial}(\theta_i, n_{\text{ops}}_i) \]

where, for \( i = 1, \ldots, 12 \), \( \theta_i \) is probability of death.

```
.fvset base none hospital
.binreg deaths i.hospital, nocons n(n_ops) or
```

| death | Odds Ratio | Std. Err. | z  | P>|z| | [95% Conf. Interval] |
|-------|------------|-----------|----|------|----------------------|
| hospital |            |           |    |      |                      |
| 1     | 3.17e-09   | 4.98e-06  | -0.01 | 0.990 | 0                    |
| 2     | .1384615   | .0348219  | -7.86 | 0.000 | .0845784 .2266725   |
| ...   |            |           |      |      |                      |
| 12    | .0714286   | .015092   | -12.49 | 0.000 | .0472088 .108074    |

Risk probability for the first hospital is estimated to be zero.
Hospital ranking: Mixed-effects approach

A random-intercept model pools information across hospitals and provides more believable predictions for the risk probabilities.

```
.meglm deaths || hospital:, family(binomial n_ops) link(logit)
.predict theta, xb
.predict re, reffects
.replace theta = invlogit(theta+re)
.list hospital n_ops deaths theta
```

<table>
<thead>
<tr>
<th>hospital</th>
<th>n_ops</th>
<th>deaths</th>
<th>theta</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>47</td>
<td>0</td>
</tr>
<tr>
<td>2</td>
<td>2</td>
<td>148</td>
<td>18</td>
</tr>
<tr>
<td>3</td>
<td>3</td>
<td>119</td>
<td>8</td>
</tr>
<tr>
<td>4</td>
<td>4</td>
<td>810</td>
<td>46</td>
</tr>
<tr>
<td>11</td>
<td>11</td>
<td>256</td>
<td>29</td>
</tr>
<tr>
<td>12</td>
<td>12</td>
<td>360</td>
<td>24</td>
</tr>
</tbody>
</table>

We obtain point estimates of the risk probabilities.
Hospital ranking: Limitations of the standard approaches

Although the mixed-effects model predicts hospital risk probabilities that can be used for ranking, it is impossible to quantify the credibility of the predicted hospital ranking. The frequentist approach cannot answer questions such as

- How probable is the risk of death for the first hospital to be lower than the second hospital?
- What is the probability the first hospital to have rank one, that is, to perform best across all twelve hospitals?

Can a Bayesian approach help?
A Bayesian model for data \( y \) and model parameters \( \theta \) includes

- Likelihood function \( L(\theta; y) = P(y|\theta) \)
- Prior probability distribution \( \pi(\theta) \)
- Bayes rule for the posterior distribution

\[
P(\theta|y) \propto L(\theta; y)\pi(\theta)
\]

- Posterior distribution \( P(\theta|y) \) provides full description of \( \theta \)
- MCMC methods are usually used for simulating \( P(\theta|y) \)
Bayesian analysis in Stata

<table>
<thead>
<tr>
<th>Command</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Estimation</strong></td>
<td></td>
</tr>
<tr>
<td>bayesmh</td>
<td>Bayesian regression using MH</td>
</tr>
<tr>
<td><strong>Postestimation</strong></td>
<td></td>
</tr>
<tr>
<td>bayesgraph</td>
<td>Graphical diagnostics</td>
</tr>
<tr>
<td>bayesstats ess</td>
<td>Effective sample sizes</td>
</tr>
<tr>
<td>bayesstats ic</td>
<td>Bayesian information criteria</td>
</tr>
<tr>
<td>bayesstats summary</td>
<td>Summary statistics</td>
</tr>
<tr>
<td>bayestest interval</td>
<td>Interval hypothesis testing</td>
</tr>
<tr>
<td>bayestest model</td>
<td>Model posterior probabilities</td>
</tr>
</tbody>
</table>
Bayesian estimation in Stata

- Built-in likelihood models
  
  \texttt{bayesmh \ldots, likelihood() prior() \ldots}

- User-defined models
  
  \texttt{bayesmh \ldots, \{evaluator() \mid llevaluator()\} \ldots}

- You can access the GUI by typing
  
  \texttt{. db bayesmh}

  or from the statistical menu.

- \texttt{bayesmh} performs MCMC estimation using adaptive Metropolis-Hastings (MH) algorithm.
Prior distributions

- **Completely uninformative** priors: the flat prior option
  
  \[
  \text{prior}(\{\text{params}\}, \text{flat})
  \]

- **Weakly informative** priors: \( N(0, 1e6) \)
  
  \[
  \text{prior}(\{\text{params}\}, \text{normal}(0, 1e6))
  \]

- **Informative** priors: \( N(-1, 1), \text{InvGamma}(10, 10), \ldots \)

- **Hierarchical** priors using hyper-parameters: \( N(\mu, \sigma^2) \)
  
  \[
  \text{prior}(\{\text{params}\}, \text{normal}(\{\mu\}, \{\text{sig2}\}))
  \]
  \[
  \text{prior}(\{\mu\}, \text{normal}(0, 100))
  \]
  \[
  \text{prior}(\{\text{sig2}\}, \text{igamma}(0.01, 0.01))
  \]

- Hierarchical priors are essential in Bayesian multilevel modeling
Two sources of hierarchy in Bayesian models

- **Multilevel data structure**, where observations are grouped by one or more categorical variables; it is represented in the **likelihood**. For example, observations of students clustered in schools.
  - Frequentist: fixed-effects and random-effects (RE) parameters.
  - Bayesian: all model parameters are random, and the distinction is in their prior specification.

- **Model parameter hierarchy**, where the prior of lower-level parameters involves higher-level hyper-parameters.

\[
\text{prior}(\{RE\_params\}, \text{normal}(\{RE\_cons\}, \{RE\_var\}))
\]
\[
\text{prior}(\{RE\_cons\}, \text{normal}(0, 100))
\]
\[
\text{prior}(\{RE\_var\}, \text{igamma}(0.01, 0.01))
\]
Consider a simple random-intercept regression (2-level) model

\[ y = X\beta + Zu + \epsilon \]

where \( Z \) is \( n \times q \) design matrix and \( u_j, j \in \{1, \ldots, q\} \), are “random-effects” parameters.

\( u_j \)'s are assigned a hierarchical prior, typically

\[ u_j | \mu, \sigma_u^2 \sim i.i.d. \ N(\mu, \sigma_u^2) \]

where \( \mu \) and \( \sigma_u^2 \) are hyper-parameters.
Block sampling of random-effects parameters

- RE parameters $u_j$’s are, typically, **highly dependent** in the prior and posterior, which complicates MCMC simulation.

$$\pi(u_1, \ldots u_q) \neq \prod_{j=1}^{q} \pi(u_j)$$

- `bayesmh` employs an adaptive random-walk Metropolis sampling algorithm in which model parameters are grouped in blocks.
- If $u_j$’s are grouped in one block, the sampling becomes extremely inefficient as $q$ increases - **the curse of dimensionality**.
- When $u_j$’s are sampled individually, the computational complexity of one MCMC iteration is $O(nq)$, where $n$ is the sample size.
- The solution: use the `reffects()` option in `bayesmh`. 
Efficient sampling of RE parameters in \texttt{bayesmh}

\textbullet \ \texttt{bayesmh} employs the conditional independence of random-effects parameters in both prior and posterior

\[
\pi(u_1, \ldots, u_q | \mu, \sigma_u^2) = \prod_{j=1}^{q} \pi(u_j | \mu, \sigma_u^2)
\]

\[
P(u_1, \ldots, u_q | \mu, \sigma_u^2, y) = \prod_{j=1}^{q} P(u_j | \mu, \sigma_u^2, y_j)
\]

where \(y_j\) is a subsample of \(y\) having effect \(u_j\).

\textbullet \ In such cases the computational complexity of one MCMC iteration is now only \(O(n)\), a huge improvement from \(O(nq)\).
Specifying RE parameters in `bayesmh`

- Suboption `reffects` of option `block()`
  ```stata
  . fvset base none u
  . bayesmh y ... i.u, likelihood(...) ...
    block({y:i.u}, reffects) ...
  ```

- Global `reffects()` option
  ```stata
  . bayesmh y ..., reffects(u) ...
  ```

- Option `redefine()`: specify RE linear forms to be used as latent variables in expressions
  ```stata
  . fvset base none u
  . bayesmh y = ({re:}), redefine(re:i.u) ...
  ```
Back to the hospital ranking example

Recall our earlier example of mortality rate after cardiac surgery.

```
. input hospital n_ops deaths
  1   47   0
  2  148   18
  3  119    8
  4  810   46
  5  211    8
  6  196   13
  7  148    9
  8  215   31
  9  207   14
 10   97    8
 11  256   29
 12  360   24
. end
```

The standard frequentist approach is unable to answer satisfactory our research questions.
Hospital ranking models

I will fit three Bayesian models with increasing complexity according to their prior specification

- Model 1: Completely uninformative, flat, prior
- Model 2: Slightly informative prior
- Model 3: Hierarchical prior

I will discard the first model as improper. Then, I will compare the second and the third models and show that the latter, the hierarchical model, is the best fit for the data.
Model 1: Uninformative priors

We assume that death incidents are independent across hospitals and apply uninformative, flat, prior for the risk effects.

.fvset none hospital
.bayesmh deaths i.hospital, likelihood(binomial(n_ops)) ///
prior({deaths:i.hospital}, flat) noconstant

The above specification has poor sampling efficiency. To improve the MCMC sampling efficiency we apply the global reffects() option

.set seed 12345
.bayesmh deaths, reffects(hospital) likelihood(binomial(n_ops)) ///
prior({deaths:i.hospital}, flat) noconstant ///
showreffects({deaths:i.hospital})
Bayesian binomial regression
Random-walk Metropolis-Hastings sampling

MCMC iterations = 12,500
Burn-in = 2,500
MCMC sample size = 10,000
Number of obs = 12
Acceptance rate = .3138
Efficiency: min = .001144
avg = .1483
max = .2025

Log marginal likelihood = -25.093932

| Equal-tailed |
|-------------|-----------------|-----------------|--------------|-----------------|-----------------|-----------------|-----------------|
| deaths      | Mean            | Std. Dev.       | MCSE          | Median         | [95% Cred. Interval] |
|-------------|-----------------|-----------------|--------------|-----------------|-----------------|-----------------|
| hospital    |                 |                 |               |                 |                 |                 |
| 1           | -165.8625       | 56.62666        | 16.7452      | -177.5466       | -237.5561       | -29.43683       |
| 2           | -1.998605       | .256157         | .0063        | -1.985625       | -2.51977        | -1.521995       |
| 3           | -2.691607       | .3765987        | .008468      | -2.663127       | -3.487504       | -2.024282       |
| ...         |                 |                 |               |                 |                 |                 |
| 11          | -2.072715       | .1923903        | .005107      | -2.068274       | -2.461135       | -1.719813       |
| 12          | -2.654584       | .2146438        | .005511      | -2.651604       | -3.079447       | -2.254491       |

Note: There is a high autocorrelation after 500 lags.
## Model 1: Sampling efficiency

```bash
.bayesstats ess
```

### Efficiency summaries  
MCMC sample size = 10,000

<table>
<thead>
<tr>
<th>deaths</th>
<th>ESS</th>
<th>Corr. time</th>
<th>Efficiency</th>
</tr>
</thead>
<tbody>
<tr>
<td>hospital</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>11.44</td>
<td>874.46</td>
<td>0.0011</td>
</tr>
<tr>
<td>2</td>
<td>1653.45</td>
<td>6.05</td>
<td>0.1653</td>
</tr>
<tr>
<td>3</td>
<td>1978.00</td>
<td>5.06</td>
<td>0.1978</td>
</tr>
<tr>
<td>...</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>1419.06</td>
<td>7.05</td>
<td>0.1419</td>
</tr>
<tr>
<td>12</td>
<td>1516.84</td>
<td>6.59</td>
<td>0.1517</td>
</tr>
</tbody>
</table>

The very small ESS for the first hospital suggests **nonconvergence**.
Model 1: Diagnostic plot confirms nonconvergence

\`. bayesgraph diagnostic {deaths:1bn.hospital}\`
Model 2: Weakly informative priors

We again assume that death incidents are independent across hospitals but this time we apply slightly informative, \texttt{normal(0, 100)}, prior for the probabilities of death.

\begin{verbatim}
. set seed 12345
. bayesmh deaths, reffects(hospital) likelihood(binomial(n_ops)) ///
   prior({deaths:i.hospital}, normal(0, 100)) noconstant ///
   showreffects({deaths:i.hospital}) saving(model2)
. estimates store model2
\end{verbatim}

We also save the simulation results in \texttt{model2.dta} and store estimation results as \texttt{model2}. 
Model 2: Sampling efficiency

. bayesstats ess

Efficiency summaries MCMC sample size = 10,000

<table>
<thead>
<tr>
<th>deaths</th>
<th>ESS</th>
<th>Corr. time</th>
<th>Efficiency</th>
</tr>
</thead>
<tbody>
<tr>
<td>hospital</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>129.62</td>
<td>77.15</td>
<td>0.0130</td>
</tr>
<tr>
<td>2</td>
<td>1587.85</td>
<td>6.30</td>
<td>0.1588</td>
</tr>
<tr>
<td>3</td>
<td>1936.80</td>
<td>5.16</td>
<td>0.1937</td>
</tr>
<tr>
<td>...</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>1483.44</td>
<td>6.74</td>
<td>0.1483</td>
</tr>
<tr>
<td>12</td>
<td>1541.34</td>
<td>6.49</td>
<td>0.1541</td>
</tr>
</tbody>
</table>

The ESS for the first hospital is greatly improved.
Model 2: Diagnostic plot for the first hospital

```
.bayesgraph diagnostic {deaths:1bn.hospital}
```

![Diagnostic plot for Model 2](image)
Model 2: Summaries

Note that the parameters \{deaths:i.hospital\} are regression coefficients in a generalized linear model with logit link. We apply \texttt{invlogit()} transformation to obtain risk probabilities.

\begin{verbatim}
. bayesstats summary (hosp1_risk:invlogit({deaths:1bn.hospital}))) ///
   (hosp2_risk:invlogit({deaths:2.hospital}))) ///
   (hosp3_risk:invlogit({deaths:3.hospital}))), nolegend
\end{verbatim}

Posterior summary statistics

\begin{verbatim}
Prior summary statistics
MCMC sample size = 10,000
\end{verbatim}

\begin{verbatim}
<p>| Equal-tailed |</p>
<table>
<thead>
<tr>
<th>Mean</th>
<th>Std. Dev.</th>
<th>MCSE</th>
<th>Median</th>
<th>[95% Cred. Interval]</th>
</tr>
</thead>
<tbody>
<tr>
<td>hosp1_risk</td>
<td>.0021345</td>
<td>.0073743</td>
<td>.000265</td>
<td>.0000308</td>
</tr>
<tr>
<td>hosp2_risk</td>
<td>.1214157</td>
<td>.0266825</td>
<td>.000669</td>
<td>.1192722</td>
</tr>
<tr>
<td>hosp3_risk</td>
<td>.066891</td>
<td>.0228277</td>
<td>.000514</td>
<td>.0650115</td>
</tr>
</tbody>
</table>
\end{verbatim}
Model 3: Hierarchical approach

It is more realistic to assume that the risks of death across hospitals are related. After all, the surgical procedures followed in different hospital are probably similar. This observation motivates the following random-effects model

$$\text{deaths}_i \sim \text{Binomial}(\text{invlogit}(u_i), n_{-ops}_i)$$

$$u_i \sim \text{Normal}(\mu, \sigma^2)$$

This is a two-level model with RE parameters $u_i$’s and hyper-parameters $\mu$ and $\sigma^2$. Moreover, we assume exchangiability of $u_i$’s

$$u_i | \mu, \sigma^2 \sim \text{i.i.d. Normal}(\mu, \sigma^2)$$
Model 3: Specification

. set seed 12345
. bayesmh deaths, reffects(hospital) likelihood(binomial(n_ops)) ///
   prior({deaths:i.hospital}, normal({mu}, {sig2})) noconstant ///
   prior({mu}, normal(0, 1e6)) ///
   prior({sig2}, igamma(0.001, 0.001)) ///
   block({mu}) block({sig2}) ///
   saving(model3, replace)
. estimates store model3

- The RE parameters $u_i$’s are represented by \{deaths:i.hospital\}.
- We apply uninformative hyperpriors for \{mu\} and \{sig2\}.
Model 3: Estimation results

Bayesian binomial regression
Random-walk Metropolis-Hastings sampling

MCMC iterations = 12,500
Burn-in = 2,500
MCMC sample size = 10,000
Number of obs = 12
Acceptance rate = .3743
Efficiency: min = .02602
avg = .05918
max = .09235

Log marginal likelihood = -48.442035

<table>
<thead>
<tr>
<th>Equal-tailed</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
</tr>
<tr>
<td>-------------</td>
</tr>
<tr>
<td>mu</td>
</tr>
<tr>
<td>sig2</td>
</tr>
</tbody>
</table>
Model 3: Diagnostic plot for the first hospital

```
  . bayesgraph diagnostic {deaths:1bn.hospital}
```

![Diagnostic plot for the first hospital](image)
Bayesian information criteria

We compare model2 and model3

. bayesstats ic model2 model3

Bayesian information criteria

<table>
<thead>
<tr>
<th></th>
<th>DIC</th>
<th>log(ML)</th>
<th>log(BF)</th>
</tr>
</thead>
<tbody>
<tr>
<td>model2</td>
<td>74.76517</td>
<td>-66.21896</td>
<td>.</td>
</tr>
<tr>
<td>model3</td>
<td>74.26092</td>
<td>-48.44204</td>
<td>17.77692</td>
</tr>
</tbody>
</table>

Note: Marginal likelihood (ML) is computed using Laplace-Metropolis approximation.

model3 is a better fit than model2 with respect to both DIC and marginal likelihood ML.
Bayesian model comparison

We compare model2 and model3

. bayestest model model2 model3

Bayesian model tests

|                 | log(ML)  | P(M)  | P(M|y)  |
|-----------------|----------|-------|---------|
| model2          | -66.2190 | 0.5000| 0.0000  |
| model3          | -48.4420 | 0.5000| 1.0000  |

Note: Marginal likelihood (ML) is computed using Laplace-Metropolis approximation.

Conclusion: model3 is overwhelmingly better than model2 based on the Bayes factors and model probabilities.
Model 3: Summaries

. bayesstats summary (hosp1_risk:invlogit({deaths:1bn.hospital})) ///
(hosp2_risk:invlogit({deaths:2.hospital})) ///
(hosp3_risk:invlogit({deaths:3.hospital})), nolegend

Posterior summary statistics

<table>
<thead>
<tr>
<th>Equal-tailed</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean Std. Dev. MCSE Median [95% Cred. Interval]</td>
</tr>
<tr>
<td>---------------------------</td>
</tr>
<tr>
<td>hosp1_risk</td>
</tr>
</tbody>
</table>
hosp2_risk | .1037734 .0227254 .000705 .1009743 .0667345 .1555239 |
hosp3_risk | .0704388 .0174802 .000423 .0695322 .0403892 .1094492 |

The posterior mean risk for the first hospital is estimated to be about 5%.

These posterior means are very close to the predicted with `meglm`.

Nikolay Balov (Stata)
Bayesian hierarchical models in Stata
2016 Stata Conference
Model 3: Histogram plots of the risk effects

.bayesgraph histogram {deaths:i.hospital}, ///
   by parm(legend(off) noxrescale noyrescale ///
   title(Posterior distributions of risk effects)) ///
   normal

Posterior distributions of risk effects

Graphs by parameter
Model 3: Hospital comparison

We can test whether the risk probability for the first hospital is lower than that for the second hospital.

```
. bayestest interval (prob12:{deaths:1bn.hospital}-{deaths:2.hospital}), ///
   upper(0) nolegend

Interval tests       MCMC sample size = 10,000
```

<table>
<thead>
<tr>
<th></th>
<th>Mean</th>
<th>Std. Dev.</th>
<th>MCSE</th>
</tr>
</thead>
<tbody>
<tr>
<td>prob12</td>
<td>.961</td>
<td>0.19360</td>
<td>.0054645</td>
</tr>
</tbody>
</table>

We estimate the posterior probability \( P(u_1 < u_2) \) to be 96%.
What is the probability of the first hospital to have rank 1?

```
. bayestest interval (\{deaths:1bn.hospital\} - min( \\
   \{deaths:2.hospital\},\{deaths:3.hospital\}, \\
   \{deaths:4.hospital\},\{deaths:5.hospital\}, \\
   \{deaths:6.hospital\},\{deaths:7.hospital\}, \\
   \{deaths:8.hospital\},\{deaths:9.hospital\}, \\
   \{deaths:10.hospital\},\{deaths:11.hospital\}, \\
   \{deaths:12.hospital\})), upper(0)
```

<table>
<thead>
<tr>
<th></th>
<th>Mean</th>
<th>Std. Dev.</th>
<th>MCSE</th>
</tr>
</thead>
<tbody>
<tr>
<td>prob1</td>
<td>.3588</td>
<td>0.47967</td>
<td>.0120528</td>
</tr>
</tbody>
</table>

We estimate the posterior probability $P(u_1 \leq \min(u))$ to be 36%. The Bayesian approach gives us more informative quantitative answers than any of the standard frequentist approaches.
The advantage of hierarchical priors

- Flat or uninformative priors may result in improper posterior.
- Strong informative priors may be subjective and introduce bias.
- Hierarchical priors provide a compromise between these two ends by using informative prior family of distributions and uninformative hyper-priors for the hyper-parameters

\[
\text{prior(} \{\text{RE}_\text{params}\}, \text{normal(} \{\text{RE}_\text{cons}\}, \{\text{RE}_\text{var}\})\text{)}
\]
\[
\text{prior(} \{\text{RE}_\text{cons}\}, \text{normal(} 0, 100)\text{)}
\]
\[
\text{prior(} \{\text{RE}_\text{var}\}, \text{igamma(} 0.01, 0.01)\text{)}
\]

- The hierarchical prior specification provides pooling of information across REs to enhance model estimation.
Other hierarchical models using bayesmh
Random-intercept model

- Modeling weight growth based on panel data
- Data: weight measurements of 48 pigs identified by id on 9 successive weeks (e.g. Diggle et al. [2002]).
- Consider a random intercept model with group variable id

\[ \text{weight}_{ij} = b_{1\text{week}} + u_j + \epsilon_{ij} \]

\[ u_j \sim N(b_0, \sigma^2_{\text{cons}}), \quad \epsilon_{ij} \sim N(0, \sigma^2) \]

where \( j = 1, \ldots, 48 \) and \( i = 1, \ldots, n_j = 9 \).

- Noninformative hyperpriors

\[ b_0, b_1 \sim \text{Normal}(0, 100) \]

\[ \sigma^2, \sigma^2_{\text{cons}} \sim \text{InvGamma}(0.01, 0.01) \]
Bayesian random-intercept model

We use the global `reffects(id)` option to introduce the random intercept parameters.

```
.bayesmh weight week, reffects(id) likelihood(normal({var})) noconstant ///
  ///
  prior({weight:i.id}, normal({weight:_cons}, {var_cons})) ///
  ///
  prior({var}, igamma(0.01, 0.01)) block({var}, gibbs) ///
  prior({var_cons}, igamma(0.01, 0.01)) block({var_cons}, gibbs) ///
  ///
  prior({weight:week}, normal(0,1e2)) block({weight:week}, gibbs) ///
  prior({weight:_cons}, normal(0,1e2)) block({weight:_cons},gibbs)
```

We request the `noconstant` option and include the parameter `{weight:_cons}` as the mean of the random intercepts.
Two-level, random-slope model with unstructured covariance

- Mixed-effects specification

\[ \text{weight}_{ij} = b_0 + b_1 \text{week} + u_j + v_j \text{week} + \epsilon_{ij} \]

\[(u_j, v_j) \sim \text{MVN}(0, 0, \Sigma_{2x2}), \epsilon_{ij} \sim \text{N}(0, \sigma^2)\]

- We can fit this model by typing

  . mixed weight week || id: week, cov(unstructured)

- Alternative formulation

\[ \text{weight}_{ij} = u_j + v_j \text{week} + \epsilon_{ij} \]

\[(u_j, v_j) \sim \text{MVN}(b_0, b_1, \Sigma_{2x2}), \epsilon_{ij} \sim \text{N}(0, \sigma^2)\]
Bayesian two-level model with unstructured covariance

```stata
.fvset base none id
.bayesmh weight i.id i.id#c.week, likelihood(normal({var_0})) noconstant ///
    prior ({weight:i.id i.id#c.week}, ///
        mvnormal(2, {weight:_cons}, {weight:week}, {covar,m})) ///
    block ({weight: i.id}, reffects) ///
    block ({weight: i.id#c.week}, reffects)
    prior({var_0}, igamma(0.01, 0.01)) block({var_0}, gibbs)
    prior({covar,m}, iwishart(2, 3, I(2)) block({covar,m}, gibbs)
    prior({weight:week _cons}, normal(0, 1e2))
    block({weight:_cons}) block({weight:week})
```

Because we use factor notation to introduce random slopes and intercepts, we need to suppress the base level of `id`.
Weight gain in children: Quadratic growth curve model

Data: weight gain in Asian children in UK (e.g. S. Rabe-Hesketh et al. [2008]).

. use http://www.stata-press.com/data/mlmus2/asian, clear
. gen age2 = age^2

A random-slope model with unstructured covariance

. bayesmh weight age2 i.id i.id#c.age, likelihood(normal({var_0})) noconstant ///
   prior ({weight:i.id i.id#c.age},
   mvnnormal(2, {weight:_cons}, {weight:age}, {covar,m})) ///
   block ({weight: i.id}, reffects) ///
   block ({weight: i.id#c.age}, reffects)

   prior({var_0}, igamma(0.01, 0.01)) block({var_0}, gibbs) ///
   prior({covar,m}, iwishart(2, 3, I(2))) block({covar,m}, gibbs)

   prior({weight:age age2 _cons}, normal(0, 1e4))
   block({weight:_cons}) block({weight:age})
   exclude({weight:i.id i.id#c.age})
Weight gain in children: Estimation results from bayesmh

<table>
<thead>
<tr>
<th>Equal-tailed</th>
<th>Mean</th>
<th>Std. Dev.</th>
<th>MCSE</th>
<th>Median</th>
<th>[95% Cred. Interval]</th>
</tr>
</thead>
<tbody>
<tr>
<td>weight</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>age2</td>
<td>-1.682645</td>
<td>0.0902288</td>
<td>0.02214</td>
<td>-1.68861</td>
<td>-1.840406 -1.460976</td>
</tr>
<tr>
<td>var_0</td>
<td>0.345705</td>
<td>0.0550158</td>
<td>0.003565</td>
<td>0.3409993</td>
<td>0.2534185 0.4691682</td>
</tr>
</tbody>
</table>

The results are similar to those from

. mixed weight age age2 || id: age, mle
Gaussian 2-mixture model

We observe outcome $y$ coming from a mixture of two Gaussian distributions with common variances but different means. The latent mixing variable $z$ is not observed.

$$y|z \sim N(\mu_z, \sigma^2), \quad z \in \{1, 2\},$$

$$z \sim \text{Multinomial}(\pi_1, \pi_2)$$

We want to estimate $\pi_j, \mu_j, j = 1, 2$, and $\sigma^2$. 
Federal interest rates: A two-staged model

Records from the database of the Federal Reserve Bank of Saint Louis from 1954 to 2010 reveal a period in 1970s and 1980s with unusually high rates. **We want to estimate the levels of moderate and high rates.**

```
. webuse usmacro
```

A Markov-switching model with switching intercept: see Example 1 in mswitch manual.

```
. mswitch dr fedfunds
```
Federal interest rates: Gaussian 2-mixture model

. generate id = _n
. fvset base none id

A Gaussian 2-mixture model is applied to the outcome fedfunds

. set seed 12345
. bayesmh fedfunds = (({state:}==1)*{mu1}+({state:}==2)*{mu2}), ///
    likelihood(normal({sig2})) redefine(state:i.id) ///
    prior({state:], index({p1}, (1-{p1}))) ///
    prior({p1], uniform(0, 1)) ///
    prior({mu1} {mu2], normal(0, 100)) ///
    prior({sig2], igamma(0.1, 0.1)) ///
    init({p1} 0.5 {mu1} 1 {mu2} 1 {sig2} 1 {state:] 1) ///
    block({sig2], gibbs) block({p1]) block({mu1}{mu2}) ///
    exclude({state:]) dots
Federal interest rates: Estimation results

Bayesian normal regression

Metropolis-Hastings and Gibbs sampling

MCMC iterations = 12,500
Burn-in = 2,500
MCMC sample size = 10,000
Number of obs = 226
Acceptance rate = .5397
Efficiency: min = .02064
avg = .04739
max = .1073

Log marginal likelihood = .

<table>
<thead>
<tr>
<th>Equal-tailed</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
</tr>
<tr>
<td>------</td>
</tr>
<tr>
<td>mu1</td>
</tr>
<tr>
<td>mu2</td>
</tr>
<tr>
<td>sig2</td>
</tr>
<tr>
<td>p1</td>
</tr>
</tbody>
</table>

Nikolay Balov (Stata)
Bayesian hierarchical models in Stata
2016 Stata Conference
Federal interest rates: Histogram plots

```
. bayesgraph histogram {mu1 mu2},
   byparm(legend(off) noxrescale noyrescale
   title(Posterior distributions of fund rates)
   normal
```

![Posterior distributions of fund rates](image.png)
Educational research example: 3PL IRT model

- Predict the effect of subject ability and question difficulty and discrimination on test performance.
- We observe binary responses $y_{ij}$ of subjects $j = 1, \ldots, K$ with abilities $\theta_j$ on items $i = 1, \ldots, I$ with discrimination parameters $a_i$, difficulty parameters $b_i$, and guessing parameters $c_i$.

$$P(y_{ij} = 1) = c_i + (1 - c_i) \text{InvLogit}\{a_i(\theta_j - b_i)\},$$

$$\theta_j \sim N(0, 1) \quad a_i > 0, \quad c_i \in [0, 1]$$

- Hierarchical priors

$$\log(a_i) \sim N(\mu_a, \sigma_a^2)$$

$$b_i \sim N(\mu_b, \sigma_b^2)$$

$$\log(c_i) \sim N(\mu_c, \sigma_c^2)$$
Bayesian 3PL IRT

\[ P(y_{ij} = 1) = c_i + (1 - c_i) \text{InvLogit}\{a_i(\theta_j - b_i)\} \]

```
. bayesmh y, likelihood(dbernoulli(
   {c:}+(1-{c:})*invlogit({a:}*({theta:}-{b:}))))
   ///
   redefine(a:i.item) redefine(b:i.item)
   redefine(c:i.item) redefine(theta:i.id)
   ///
   prior({theta:i.id}, normal(0, 1))
   prior({a:i.item}, lognormal({mua}, {vara}))
   prior({b:i.item}, normal({mub}, {varb}))
   prior({c:i.item}, lognormal({muc}, {varc}))
   ///
   prior({mua}{mub}{muc}, normal(0, 0.1))
   prior({vara}{varb}{varc}, igamma(10, 1))
```

You can find more details in our Stata blog entry: *Bayesian binary item response theory models using bayesmh.*
Conclusion

The Bayesian hierarchical modeling approach is a powerful tool that facilitates

- the representation of complex multilevel data structures
- the specification of objective priors
- the modeling by exploiting intra-group correlation across panels (pooling information across panels)
- the inference by providing intuitive and comprehensive answers to research questions

The current suite of commands for Bayesian analysis in Stata makes hierarchical modeling accessible for a wide variety of problems.