Fitting Bayesian regression models using the bayes prefix

Yulia Marchenko

Executive Director of Statistics
StataCorp LP

2017 UK Stata Users Group Meeting
In a nutshell

What is Bayesian analysis?

Stata’s Bayesian suite of commands

Bayesian linear regression

Postestimation

Bayesian autoregressive models

Bayesian multilevel models

Bayesian survival models

Concluding remarks

References
Stata 15 provides a convenient and elegant way of fitting Bayesian regression models by prefixing your estimation command with `bayes`.

- You fit linear regression by typing
  ```
  . regress y x
  ```
  You can now fit Bayesian linear regression by typing
  ```
  . bayes: regress y x
  ```
- Default priors are provided for convenience; you should carefully think about the priors and often specify your own:
  ```
  . bayes, prior(...) prior(...) ... : regress y x
  ```
- 45 estimation commands are supported including GLM, survival models, multilevel models, and more.
- All Bayesian postestimation features work after `bayes`: just like they do after `bayesmh`. 
In a nutshell

Classical linear regression

- **Data**: Math scores of pupils in the third and fifth years from 48 different schools in Inner London (Mortimore et al. 1988).
- **Linear regression of five-year math scores** (`math5`) on **three-year math scores** (`math3`).

```
. regress math5 math3

Source | SS       | df  | MS         | Number of obs = 887
-------|----------|-----|------------|------------------
Model  | 10960.27 | 1   | 10960.27   | F(1, 885) = 341.40
Residual | 28411.62 | 885 | 32.1035    | Prob > F = 0.0000
Total   | 39371.89 | 886 | 44.4378    | R-squared = 0.2784
               |         |     |            | Adj R-squared = 0.2776
               |         |     |            | Root MSE = 5.666

math5    | Coef.  | Std. Err. | t    | P>|t|  | [95% Conf. Interval]
---------|--------|-----------|------|------|---------------------
math3    | .60813 | .032912   | 18.48| 0.000| .5435347 - .6727265
_cons    | 30.3465| .1906157  | 159.20| 0.000| 29.97245 - 30.72067
```
Bayesian linear regression

. set seed 15
. bayes: regress math5 math3

Burn-in ...
Simulation ...
Model summary

Likelihood:
  math5 ~ regress(xb_math5,{sigma2})

Priors:
  {math5:math3 _cons} ~ normal(0,10000)  \hspace{1cm} \text{(1)}
  {sigma2} ~ igamma(.01,.01)

(1) Parameters are elements of the linear form xb_math5.
Bayesian linear regression
Random-walk Metropolis-Hastings sampling

<table>
<thead>
<tr>
<th></th>
<th>MCMC iterations = 12,500</th>
<th>Burn-in = 2,500</th>
<th>MCMC sample size = 10,000</th>
<th>Number of obs = 887</th>
<th>Acceptance rate = .3312</th>
<th>Efficiency: min = .1099</th>
<th>avg = .1529</th>
<th>max = .2356</th>
</tr>
</thead>
<tbody>
<tr>
<td>Log marginal likelihood = -2817.2335</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>Mean</th>
<th>Std. Dev.</th>
<th>MCSE</th>
<th>Median</th>
<th>[95% Cred. Interval]</th>
</tr>
</thead>
<tbody>
<tr>
<td>math5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>math3</td>
<td>.6070097</td>
<td>.0323707</td>
<td>.000976</td>
<td>.6060445</td>
<td>.5440594 - .6706959</td>
</tr>
<tr>
<td>_cons</td>
<td>30.3462</td>
<td>.1903067</td>
<td>.005658</td>
<td>30.34904</td>
<td>29.97555 - 30.71209</td>
</tr>
<tr>
<td>sigma2</td>
<td>32.17492</td>
<td>1.538155</td>
<td>.031688</td>
<td>32.0985</td>
<td>29.3045 - 35.38031</td>
</tr>
</tbody>
</table>

Note: Default priors are used for model parameters.

- **bayes**: `regress` is not `regress`!
- Let’s review a few Bayesian concepts before we interpret results.
What is Bayesian analysis?

Bayesian analysis is a statistical paradigm that answers research questions about unknown parameters using probability statements.

- What is the probability that a person accused of a crime is guilty?
- What is the probability that treatment A is more cost effective than treatment B for a specific health care provider?
- What is the probability that the odds ratio is between 0.3 and 0.5?
- What is the probability that three out of five quiz questions will be answered correctly by students?
- And more.
You may be interested in Bayesian analysis if

- you have some prior information available from previous studies that you would like to incorporate in your analysis. For example, in a study of preterm birthweights, it would be sensible to incorporate the prior information that the probability of a mean birthweight above 15 pounds is negligible. Or,

- your research problem may require you to answer a question: What is the probability that my parameter of interest belongs to a specific range? For example, what is the probability that an odds ratio is between 0.2 and 0.5? Or,

- you want to assign a probability to your research hypothesis. For example, what is the probability that a person accused of a crime is guilty?

- And more.
Assumptions

- Observed data sample $y$ is fixed and model parameters $\theta$ are random.
- $y$ is viewed as a result of a one-time experiment.
- A parameter is summarized by an entire distribution of values instead of one fixed value as in classical frequentist analysis.
There is some prior (before seeing the data!) knowledge about $\theta$ formulated as a **prior distribution** $p(\theta)$.

After data $y$ are observed, the information about $\theta$ is updated based on the **likelihood** $f(y|\theta)$.

Information is updated by using the Bayes rule to form a **posterior distribution** $p(\theta|y)$:

$$p(\theta|y) = \frac{f(y|\theta)p(\theta)}{p(y)}$$

where $p(y)$ is the **marginal distribution** of the data $y$. 
Estimating a posterior distribution $p(\theta|y)$ is at the heart of Bayesian analysis.

Various summaries of this distribution are used for inference.

Point estimates: posterior means, modes, medians, percentiles.

Interval estimates: **credible intervals** (Crl)—(fixed) ranges to which a parameter is known to belong with a pre-specified probability.

Monte-Carlo standard error (MCSE)—represents precision about posterior mean estimates.
Fitting Bayesian regression models using the bayes prefix

Components of Bayesian analysis

Inference

- Hypothesis testing—assign probability to any hypothesis of interest.
- Model comparison: model posterior probabilities, Bayes factors.
- Predictions and model checking are based on a **posterior predictive distribution**:

\[
p(y_{\text{new}} | y) = \int f(y_{\text{new}} | \theta)p(\theta | y)\,d\theta
\]
Advantages

Bayesian inference:

- is universal—it is based on the Bayes rule which applies equally to all models;
- incorporates prior information;
- provides the entire posterior distribution of model parameters;
- is exact, in the sense that it is based on the actual posterior distribution rather than on asymptotic normality in contrast with many frequentist estimation procedures; and
- provides straightforward and more intuitive interpretation of the results in terms of probabilities.
Disadvantages

- Potential subjectivity in specifying prior information—noninformative priors or sensitivity analysis to various choices of informative priors.
- Computationally demanding—involves intractable integrals that can only be computed using intensive numerical methods such as Markov chain Monte Carlo (MCMC).
Stata’s Bayesian suite of commands

<table>
<thead>
<tr>
<th>Command</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Estimation</strong></td>
<td></td>
</tr>
<tr>
<td>bayes:</td>
<td>Bayesian regression models using the bayes prefix <em>(new in Stata 15)</em></td>
</tr>
<tr>
<td>bayesmh</td>
<td>General Bayesian models using MH</td>
</tr>
<tr>
<td>bayesmh evaluators</td>
<td>User-written Bayesian models using MH</td>
</tr>
<tr>
<td><strong>Postestimation</strong></td>
<td></td>
</tr>
<tr>
<td>bayesgraph</td>
<td>Graphical convergence diagnostics</td>
</tr>
<tr>
<td>bayesstats ess</td>
<td>Effective sample sizes and more</td>
</tr>
<tr>
<td>bayesstats summary</td>
<td>Summary statistics</td>
</tr>
<tr>
<td>bayesstats ic</td>
<td>Information criteria and Bayes factors</td>
</tr>
<tr>
<td>bayestest model</td>
<td>Model posterior probabilities</td>
</tr>
<tr>
<td>bayestest interval</td>
<td>Interval hypothesis testing</td>
</tr>
</tbody>
</table>
Over 50 built-in likelihoods: normal, logit, ologit, Poisson, …
Many built-in priors: normal, gamma, Wishart, Zellner’s \( g \), …
Continuous, binary, ordinal, categorical, count, censored, truncated, zero-inflated, and survival outcomes.
Univariate, multivariate, and multiple-equation models.
Linear, nonlinear, generalized linear and nonlinear, sample-selection, panel-data, and multilevel models.
Continuous univariate, multivariate, and discrete priors.
User-defined models: likelihoods and priors.

MCMC methods:
Adaptive MH.
Adaptive MH with Gibbs updates—hybrid.
Full Gibbs sampling for some models.
Built-in models

Fitting regression models
bayes: stata_command ...

Fitting general models
bayesmh ..., likelihood() prior() ...

User-defined models

Posterior evaluator
bayesmh ..., evaluator() ...

Likelihood evaluator with built-in priors
bayesmh ..., llevaluator() prior() ...

Postestimation features are the same whether you use a built-in model or program your own!
Recall our Bayesian linear regression of `math5` on `math3`.

Let’s describe results in more detail.

```
. set seed 15
. bayes: regress math5 math3

Burn-in ...
Simulation ...
Model summary
```

Likelihood:

```
  math5 ~ regress(xb_math5,{sigma2})
```

Priors:

```
{math5:math3 _cons} ~ normal(0,10000)  (1)
{sigma2} ~ igamma(.01,.01)
```

(1) Parameters are elements of the linear form `xb_math5`.
Bayesian linear regression
Random-walk Metropolis-Hastings sampling

<table>
<thead>
<tr>
<th></th>
<th>Mean</th>
<th>Std. Dev.</th>
<th>MCSE</th>
<th>Median</th>
<th>Equal-tailed [95% Cred. Interval]</th>
</tr>
</thead>
<tbody>
<tr>
<td>math5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>math3</td>
<td>.6070097</td>
<td>.0323707</td>
<td>.000976</td>
<td>.6060445</td>
<td>.5440594 .6706959</td>
</tr>
<tr>
<td>_cons</td>
<td>30.3462</td>
<td>.1903067</td>
<td>.005658</td>
<td>30.34904</td>
<td>29.97555 30.71209</td>
</tr>
<tr>
<td>sigma2</td>
<td>32.17492</td>
<td>1.538155</td>
<td>.031688</td>
<td>32.0985</td>
<td>29.3045 35.38031</td>
</tr>
</tbody>
</table>

Note: Default priors are used for model parameters.

- The output from `bayes:` is the same as the output from `bayesmh`. 
Default priors

- Default priors are provided for convenience. For example, to specify your own priors, you need to know the names of parameters, and `bayes:` provides this information in the output.

- Normal priors with zero mean and variance 10,000 are used for regression coefficients and inverse-gamma priors with shape and scale parameters of 0.01 are used for variances.

- The priors are chosen to be fairly uninformative but may become informative for parameters of large magnitude.

- Default priors may not always be suitable for your particular model.

- You should always carefully evaluate the choice of priors and specify the priors that are appropriate for your model and research question.
Custom priors

- Modify parameters of the default normal and inverse-gamma priors:

  . set seed 15
  . bayes, normalprior(10) igammaprior(1 2): regress math5 math3

  Burn-in ...
  Simulation ...
  Model summary

  Likelihood:
  math5 ~ regress(xb_math5, {sigma2})

  Priors:
  {math5:math3 _cons} ~ normal(0,100)
  {sigma2} ~ igamma(1,2)  \hspace{1cm} (1)

  (1) Parameters are elements of the linear form xb_math5.
Bayesian linear regression
Random-walk Metropolis-Hastings sampling

MCMC iterations = 12,500
Burn-in = 2,500
MCMC sample size = 10,000
Number of obs = 887
Acceptance rate = .3503
Efficiency: min = .1189
avg = .1471
max = .2005

Log marginal likelihood = -2815.3081

<table>
<thead>
<tr>
<th></th>
<th>Mean</th>
<th>Std. Dev.</th>
<th>MCSE</th>
<th>Median</th>
<th>95% Cred. Interval</th>
</tr>
</thead>
<tbody>
<tr>
<td>math5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>math3</td>
<td>.6076875</td>
<td>.033088</td>
<td>.000948</td>
<td>.6076282</td>
<td>.5405233 - .673638</td>
</tr>
<tr>
<td>_cons</td>
<td>30.326</td>
<td>.1931568</td>
<td>.005602</td>
<td>30.32804</td>
<td>29.93212 - 30.70529</td>
</tr>
<tr>
<td>sigma2</td>
<td>32.09694</td>
<td>1.530839</td>
<td>.034185</td>
<td>32.03379</td>
<td>29.27687 - 35.37723</td>
</tr>
</tbody>
</table>

Note: Default priors are used for model parameters.
Specify your own priors:

```
. set seed 15
. bayes, prior({math5:math3}, uniform(-1,1)) ///
>   prior({math5:_cons}, uniform(-50,50)) ///
>   prior({sigma2}, jeffreys): regress math5 math3
```

Likelihood:
- `math5` ~ `regress(xb_math5,{sigma2})`

Priors:
- `{math5:math3}` ~ `uniform(-1,1)` (1)
- `{math5:_cons}` ~ `uniform(-50,50)` (1)
- `{sigma2}` ~ `jeffreys`

(1) Parameters are elements of the linear form `xb_math5`. 
Fitting Bayesian regression models using the `bayes` prefix

Bayesian linear regression
Random-walk Metropolis-Hastings sampling

<table>
<thead>
<tr>
<th></th>
<th>MCMC iterations</th>
<th>Burn-in</th>
<th>MCMC sample size</th>
<th>Number of obs</th>
<th>Acceptance rate</th>
<th>Efficiency: min</th>
<th>Average</th>
<th>max</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>12,500</td>
<td>2,500</td>
<td>10,000</td>
<td>887</td>
<td>.3401</td>
<td>.1034</td>
<td>.1405</td>
<td>.211</td>
</tr>
</tbody>
</table>

Log marginal likelihood = -2806.8234

<table>
<thead>
<tr>
<th></th>
<th>Mean</th>
<th>Std. Dev.</th>
<th>MCSE</th>
<th>Median</th>
<th>[95% Cred. Interval]</th>
</tr>
</thead>
<tbody>
<tr>
<td>math5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>math3</td>
<td>.6064431</td>
<td>.0306863</td>
<td>.000954</td>
<td>.6068399</td>
<td>.5455701</td>
</tr>
<tr>
<td>_cons</td>
<td>30.34391</td>
<td>.1856718</td>
<td>.005676</td>
<td>30.3475</td>
<td>29.96434</td>
</tr>
<tr>
<td>sigma2</td>
<td>32.15952</td>
<td>1.55488</td>
<td>.033853</td>
<td>32.10525</td>
<td>29.25887</td>
</tr>
</tbody>
</table>
Use more efficient Gibbs sampling:

```plaintext
. set seed 15
. bayes, gibbs: regress math5 math3
```

Bayesian linear regression
Gibbs sampling

<table>
<thead>
<tr>
<th>MCMC iterations = 12,500</th>
</tr>
</thead>
<tbody>
<tr>
<td>Burn-in = 2,500</td>
</tr>
<tr>
<td>MCMC sample size = 10,000</td>
</tr>
<tr>
<td>Number of obs = 887</td>
</tr>
<tr>
<td>Acceptance rate = 1</td>
</tr>
<tr>
<td>Efficiency: min = 1</td>
</tr>
<tr>
<td>avg = 1</td>
</tr>
<tr>
<td>max = 1</td>
</tr>
</tbody>
</table>

Log marginal likelihood = -2817.184

<table>
<thead>
<tr>
<th>Mean</th>
<th>Std. Dev.</th>
<th>MCSE</th>
<th>Median</th>
<th>Equal-tailed [95% Cred. Interval]</th>
</tr>
</thead>
<tbody>
<tr>
<td>math5</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>math3</td>
<td>.6085104</td>
<td>.0333499</td>
<td>.000333</td>
<td>.6087819</td>
</tr>
<tr>
<td>_cons</td>
<td>30.34419</td>
<td>.1916673</td>
<td>.001917</td>
<td>30.34441</td>
</tr>
<tr>
<td>sigma2</td>
<td>32.16765</td>
<td>1.551119</td>
<td>.015511</td>
<td>32.10778</td>
</tr>
</tbody>
</table>

Note: Default priors are used for model parameters.
All Bayesian postestimation features work after `bayes`: just like they do after `bayesmh`.

```stata
.bayesgraph diagnostics {sigma2}
```
Although not as conveniently, we could already fit Bayesian linear regression using `bayesmh`.

What we couldn’t do, and still can’t, is to use time-series operators with `bayesmh`.

We can with `bayes: regress`!

Let’s use time-series operators to fit an autoregressive model.
• Data: Quarterly coal consumption (in millions of tons) in a given year in the United Kingdom from 1960 to 1986 (e.g., Harvey [1989]). (Variable `lcoal` is transformed using `log(coal/1000)`.

• Bayesian AR(1) model:

```
. bayes: regress lcoal L.lcoal
Burn-in ...
Simulation ...
Model summary

Likelihood:
  lcoal ~ regress(xb_lcoal,{sigma2})

Priors:
  {lcoal:L.lcoal _cons} ~ normal(0,10000) (1)
  {sigma2} ~ igamma(.01,.01)
```

(1) Parameters are elements of the linear form `xb_lcoal`. 
Fitting Bayesian regression models using the bayes prefix

Bayesian autoregressive models

AR(1) model

Bayesian linear regression
Random-walk Metropolis-Hastings sampling

MCMC iterations = 12,500
Burn-in = 2,500
MCMC sample size = 10,000
Number of obs = 107
Acceptance rate = .3285
Efficiency: min = .1199
               avg = .1448
               max = .1905

Log marginal likelihood = -75.889709

<table>
<thead>
<tr>
<th></th>
<th>Mean</th>
<th>Std. Dev.</th>
<th>MCSE</th>
<th>Median</th>
<th>[95% Cred. Interval]</th>
</tr>
</thead>
<tbody>
<tr>
<td>_coal</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>L1.</td>
<td>.7143121</td>
<td>.0649968</td>
<td>.001877</td>
<td>.7123857</td>
<td>.5884089 - .8436602</td>
</tr>
<tr>
<td>_cons</td>
<td>-.6896604</td>
<td>.1561023</td>
<td>.004433</td>
<td>-.6935272</td>
<td>-.9970502 - -.3879924</td>
</tr>
<tr>
<td>sigma2</td>
<td>.1702592</td>
<td>.0243144</td>
<td>.000557</td>
<td>.1672834</td>
<td>.1299619 - .2248287</td>
</tr>
</tbody>
</table>

Note: Default priors are used for model parameters.

Store results for later comparison.

. bayes, saving(lag1_mcmc)
note: file lag1_mcmc.dta saved
. estimates store lag1
Bayesian AR(2) model:

```
.bayes, saving(lag2_mcmc): regress lcoal L.lcoal L2.lcoal
```

Burn-in ...
Simulation ...
file lag2_mcmc.dta saved
Model summary

Likelihood:
  lcoal ~ regress(xb_lcoal,{sigma2})

Priors:
  {lcoal:L.lcoal L2.lcoal _cons} ~ normal(0,10000) (1)
  {sigma2} ~ igamma(.01,.01)

(1) Parameters are elements of the linear form xb_lcoal.
Fitting Bayesian regression models using the `bayes` prefix

Bayesian autoregressive models

**AR(2) model**

Bayesian linear regression

Random-walk Metropolis-Hastings sampling

MCMC iterations = 12,500

Burn-in = 2,500

MCMC sample size = 10,000

Number of obs = 106

Acceptance rate = .3614

Efficiency:

- min = .07552
- avg = .1172
- max = .1966

Log marginal likelihood = -82.507817

<table>
<thead>
<tr>
<th></th>
<th>Mean</th>
<th>Std. Dev.</th>
<th>MCSE</th>
<th>Median</th>
<th>Equal-tailed [95% Cred. Interval]</th>
</tr>
</thead>
<tbody>
<tr>
<td>lcoal</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>lcoal</td>
<td>.6954794</td>
<td>.0967804</td>
<td>.003522</td>
<td>.6958134</td>
<td>.5008727 .8832852</td>
</tr>
<tr>
<td>L1.</td>
<td>.0372711</td>
<td>.0970813</td>
<td>.003091</td>
<td>.0350822</td>
<td>-.1491183 .2311099</td>
</tr>
<tr>
<td>L2.</td>
<td>-.6414813</td>
<td>.1760301</td>
<td>.005622</td>
<td>-.6465191</td>
<td>-.9783713 -.2926136</td>
</tr>
<tr>
<td>_cons</td>
<td>-.6414813</td>
<td>.1760301</td>
<td>.005622</td>
<td>-.6465191</td>
<td>-.9783713 -.2926136</td>
</tr>
<tr>
<td>sigma2</td>
<td>.1727567</td>
<td>.0248036</td>
<td>.000559</td>
<td>.1701944</td>
<td>.1296203 .2264395</td>
</tr>
</tbody>
</table>

Note: Default priors are used for model parameters.

. estimates store lag2
Bayesian AR(3) model:

. bayes, saving(lag3_mcmc): regress lcoal L(1/3).lcoal
. estimates store lag3

Bayesian AR(4) model:

. bayes, saving(lag4_mcmc): regress lcoal L(1/4).lcoal
. estimates store lag4

Bayesian AR(5) model:

. bayes, saving(lag5_mcmc): regress lcoal L(1/5).lcoal
. estimates store lag5
Compute model posterior probabilities:

```bash
.bayestest model lag1 lag2 lag3 lag4 lag5
```

Bayesian model tests

|     | log(ML)  | P(M)  | P(M|y) |
|-----|----------|-------|--------|
| lag1| -75.8897 | 0.2000| 0.0000 |
| lag2| -82.5078 | 0.2000| 0.0000 |
| lag3| -59.6688 | 0.2000| 0.0000 |
| lag4| -13.8944 | 0.2000| 0.9990 |
| lag5| -20.8194 | 0.2000| 0.0010 |

Note: Marginal likelihood (ML) is computed using Laplace-Metropolis approximation.
We can incorporate the estimation of a lag directly in our Bayesian model through prior distributions.

\[
\text{. bayes, prior({lcoal:L1.lcoal}, normal(0, cond({lag}>=1,100,0.01))) ///}
> \text{ prior({lcoal:L2.lcoal}, normal(0, cond({lag}>=2,100,0.01))) ///}
> \text{ prior({lcoal:L3.lcoal}, normal(0, cond({lag}>=3,100,0.01))) ///}
> \text{ prior({lcoal:L4.lcoal}, normal(0, cond({lag}>=4,100,0.01))) ///}
> \text{ prior({lcoal:L5.lcoal}, normal(0, cond({lag}>=5,100,0.01))) ///}
> \text{ prior({lag}, index(0.2,0.2,0.2,0.2,0.2)):

\text{ ///
> \text{ regress lcoal L(1/5).lcoal

\text{note: operator L1. is replaced with L. in parameter name L1.lcoal}

\text{Burn-in ...}
\text{Simulation ...}
\text{Model summary}
\]

Likelihood:
\[
\text{lcoal} \sim \text{regress(xb_lcoal,}\{\text{sigma2}\})
\]

Priors:
\[
\text{lcoal:L1.lcoal} \sim \text{normal(0,}\text{cond}\{\text{lag}\geq 1,100,0.01}\}\text{\}) (1)
\text{lcoal:L2.lcoal} \sim \text{normal(0,}\text{cond}\{\text{lag}\geq 2,100,0.01}\}\text{\}) (1)
\text{lcoal:L3.lcoal} \sim \text{normal(0,}\text{cond}\{\text{lag}\geq 3,100,0.01}\}\text{\}) (1)
\text{lcoal:L4.lcoal} \sim \text{normal(0,}\text{cond}\{\text{lag}\geq 4,100,0.01}\}\text{\}) (1)
\text{lcoal:L5.lcoal} \sim \text{normal(0,}\text{cond}\{\text{lag}\geq 5,100,0.01}\}\text{\}) (1)
\text{lcoal:_cons} \sim \text{normal(0,10000)} (1)
\text{\{sigma2\}} \sim \text{igamma(0.01,0.01)} (1)

Hyperprior:
\[
\text{\{lag\}} \sim \text{index(0.2,0.2,0.2,0.2,0.2)}
\]

(1) Parameters are elements of the linear form \text{xb_lcoal}. 

Yulia Marchenko (StataCorp)
### Bayesian Linear Regression

Fitting Bayesian regression models using the `bayes` prefix

**Bayesian autoregressive models**

**Lag selection**

<table>
<thead>
<tr>
<th></th>
<th>Mean</th>
<th>Std. Dev.</th>
<th>MCSE</th>
<th>Median</th>
<th>Equal-tailed [95% Cred. Interval]</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>lcoal</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>_lcoal</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>L1. lcoal</td>
<td>.2062446</td>
<td>.0784492</td>
<td>.011311</td>
<td>.2050062</td>
<td>.0487352 .3605725</td>
</tr>
<tr>
<td>L2. lcoal</td>
<td>-.0738366</td>
<td>.0588681</td>
<td>.002764</td>
<td>-.0739381</td>
<td>-.1877364 .0391768</td>
</tr>
<tr>
<td>L3. lcoal</td>
<td>.100462</td>
<td>.0597828</td>
<td>.004398</td>
<td>.1003963</td>
<td>-.0142032 .2216838</td>
</tr>
<tr>
<td>L4. lcoal</td>
<td>.7994076</td>
<td>.0606384</td>
<td>.006607</td>
<td>.8031808</td>
<td>.6651497 .910174</td>
</tr>
<tr>
<td>L5. lcoal</td>
<td>-.0729926</td>
<td>.0698683</td>
<td>.009211</td>
<td>-.0708155</td>
<td>-.2074388 .060126</td>
</tr>
<tr>
<td>_cons</td>
<td>-.1401982</td>
<td>.0812334</td>
<td>.015212</td>
<td>-.1438271</td>
<td>-.2877263 .0403175</td>
</tr>
<tr>
<td><strong>sigma2</strong></td>
<td>.0343128</td>
<td>.0051157</td>
<td>.000123</td>
<td>.0338508</td>
<td>.0256253 .0456132</td>
</tr>
<tr>
<td><strong>lag</strong></td>
<td>4.0194</td>
<td>.1379331</td>
<td>.004424</td>
<td>4</td>
<td>4</td>
</tr>
</tbody>
</table>

**Note:** Default priors are used for some model parameters.

**Note:** There is a high autocorrelation after 500 lags.
Recall our earlier example of math scores. There are multiple observations for each school.

Classical random-intercept model:

```
. mixed math5 math3 || school:
```

Mixed-effects ML regression

|                | Estimate  | Std. Err. | z     | P>|z|   | [95% Conf. Interval] |
|----------------|-----------|-----------|-------|-------|---------------------|
| math5          |           |           |       |       |                     |
| math3          | 0.6088066 | 0.0326392 | 18.65 | 0.000 | 0.5448349 - 0.6727783 |
| _cons          | 30.36495  | 0.3491544 | 86.97 | 0.000 | 29.68062 - 31.04928  |

Random-effects Parameters

<table>
<thead>
<tr>
<th></th>
<th>Estimate</th>
<th>Std. Err.</th>
<th>[95% Conf. Interval]</th>
</tr>
</thead>
<tbody>
<tr>
<td>school: Identity</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>var(_cons)</td>
<td>4.026853</td>
<td>1.189895</td>
<td>2.256545 - 7.186004</td>
</tr>
<tr>
<td>var(Residual)</td>
<td>28.12721</td>
<td>1.37289</td>
<td>25.5611 - 30.95094</td>
</tr>
</tbody>
</table>

LR test vs. linear model: chibar2(01) = 56.38 Prob >= chibar2 = 0.0000
Bayesian random-intercept model:

```
.bayes, melabel: mixed math5 math3 || school:
```

Note: Gibbs sampling is used for regression coefficients and variance components

Bayesian multilevel regression

<table>
<thead>
<tr>
<th>MCMC iterations</th>
<th>12,500</th>
</tr>
</thead>
<tbody>
<tr>
<td>Burn-in</td>
<td>2,500</td>
</tr>
<tr>
<td>MCMC sample size</td>
<td>10,000</td>
</tr>
</tbody>
</table>

Group variable: school

<table>
<thead>
<tr>
<th>Number of groups</th>
<th>48</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of obs</td>
<td>887</td>
</tr>
<tr>
<td>Acceptance rate</td>
<td>.8091</td>
</tr>
<tr>
<td>Efficiency: min</td>
<td>.03366</td>
</tr>
<tr>
<td>avg</td>
<td>.3331</td>
</tr>
<tr>
<td>max</td>
<td>.6298</td>
</tr>
</tbody>
</table>

Log marginal likelihood

<table>
<thead>
<tr>
<th></th>
<th>Mean</th>
<th>Std. Dev.</th>
<th>MCSE</th>
<th>Median</th>
<th>Equal-tailed [95% Cred. Interval]</th>
</tr>
</thead>
<tbody>
<tr>
<td>math5</td>
<td>.6087689</td>
<td>.0326552</td>
<td>.000436</td>
<td>.6087444</td>
<td>.5450837 - .6729982</td>
</tr>
<tr>
<td>math3</td>
<td>30.39202</td>
<td>.3597873</td>
<td>.01961</td>
<td>30.38687</td>
<td>29.67802 - 31.10252</td>
</tr>
<tr>
<td>_cons</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>school</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>var(_cons)</td>
<td>4.272626</td>
<td>1.299061</td>
<td>.039697</td>
<td>4.122282</td>
<td>2.247659 - 7.220809</td>
</tr>
<tr>
<td>var(Residual)</td>
<td>28.23014</td>
<td>1.37812</td>
<td>.017365</td>
<td>28.18347</td>
<td>25.63394 - 31.04375</td>
</tr>
</tbody>
</table>

Note: Default priors are used for model parameters.
Default output (without option `melabel`):

```
.bayes
Multilevel structure

school
   {U0}: random intercepts

Model summary

Likelihood:
   math5 ~ normal(xb_math5,{e.math5: sigma2})

Priors:
   {math5:math3 _cons} ~ normal(0,10000) (1)
   {U0} ~ normal(0,{U0: sigma2}) (1)
   {e.math5: sigma2} ~ igamma(.01,.01)

Hyperprior:
   {U0: sigma2} ~ igamma(.01,.01)
```

(1) Parameters are elements of the linear form `xb_math5`. 

Yulia Marchenko (StataCorp)
Bayesian multilevel regression
Metropolis-Hastings and Gibbs sampling

MCMC iterations = 12,500
Burn-in = 2,500
MCMC sample size = 10,000

Group variable: school
Number of groups = 48

Obs per group:
min = 5
avg = 18.5
max = 62

Number of obs = 887
Acceptance rate = .8091
Efficiency: min = .03366
avg = .3331
max = .6298

Log marginal likelihood

<table>
<thead>
<tr>
<th></th>
<th>Mean</th>
<th>Std. Dev.</th>
<th>MCSE</th>
<th>Median</th>
<th>Equal-tailed [95% Cred. Interval]</th>
</tr>
</thead>
<tbody>
<tr>
<td>math5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>math3</td>
<td>.6087689</td>
<td>.0326552</td>
<td>.000436</td>
<td>.6087444</td>
<td>.5450837  .6729982</td>
</tr>
<tr>
<td>_cons</td>
<td>30.39202</td>
<td>.3597873</td>
<td>.01961</td>
<td>30.38687</td>
<td>29.67802  31.10252</td>
</tr>
<tr>
<td>school</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>U0:sigma2</td>
<td>4.272626</td>
<td>1.299061</td>
<td>.039697</td>
<td>4.122282</td>
<td>2.247659  7.220809</td>
</tr>
<tr>
<td>e.math5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>sigma2</td>
<td>28.23014</td>
<td>1.37812</td>
<td>.017365</td>
<td>28.18347</td>
<td>25.63394  31.04375</td>
</tr>
</tbody>
</table>

Note: Default priors are used for model parameters.
Display estimates of the first 12 “random effects”:

```
. bayes, showreffects({U0[1/12]})
```

<table>
<thead>
<tr>
<th></th>
<th>Mean</th>
<th>Std. Dev.</th>
<th>MCSE</th>
<th>Median</th>
<th>Equal-tailed Mean 95% Cred. Interval</th>
</tr>
</thead>
<tbody>
<tr>
<td>math5</td>
<td>0.608769</td>
<td>0.0326552</td>
<td>0.000436</td>
<td>0.608744</td>
<td>0.5450837 to 0.6729982</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>math3</td>
<td>0.608769</td>
<td>0.0326552</td>
<td>0.000436</td>
<td>0.608744</td>
<td>0.5450837 to 0.6729982</td>
</tr>
<tr>
<td>_cons</td>
<td>30.39202</td>
<td>0.3597873</td>
<td>0.01961</td>
<td>30.38687</td>
<td>29.67802 to 31.10252</td>
</tr>
<tr>
<td>U0[school]</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>-2.685824</td>
<td>0.9776969</td>
<td>0.031227</td>
<td>-2.672364</td>
<td>-4.633162 to 0.7837494</td>
</tr>
<tr>
<td>2</td>
<td>0.015465</td>
<td>1.290535</td>
<td>0.03201</td>
<td>0.0041493</td>
<td>-2.560203 to 2.556316</td>
</tr>
<tr>
<td>3</td>
<td>1.049006</td>
<td>1.401383</td>
<td>0.033731</td>
<td>1.021202</td>
<td>-1.534088 to 3.84523</td>
</tr>
<tr>
<td>4</td>
<td>-2.123055</td>
<td>0.9921679</td>
<td>0.028859</td>
<td>-2.144939</td>
<td>-4.069283 to -0.1507593</td>
</tr>
<tr>
<td>5</td>
<td>-0.1504003</td>
<td>0.9650027</td>
<td>0.033881</td>
<td>-0.1468966</td>
<td>-2.093015 to 1.721503</td>
</tr>
<tr>
<td>6</td>
<td>0.5833945</td>
<td>1.192379</td>
<td>0.032408</td>
<td>0.5918357</td>
<td>-1.660335 to 3.049718</td>
</tr>
<tr>
<td>7</td>
<td>1.490231</td>
<td>1.332917</td>
<td>0.033846</td>
<td>1.481793</td>
<td>-1.095757 to 4.272903</td>
</tr>
<tr>
<td>8</td>
<td>0.4198105</td>
<td>0.9783772</td>
<td>0.031891</td>
<td>0.4579817</td>
<td>-1.496317 to 2.403908</td>
</tr>
<tr>
<td>9</td>
<td>-1.996105</td>
<td>1.02632</td>
<td>0.035372</td>
<td>-2.001467</td>
<td>-4.037044 to -0.029627</td>
</tr>
<tr>
<td>10</td>
<td>0.6736806</td>
<td>1.249238</td>
<td>0.031114</td>
<td>0.660939</td>
<td>1.70319 to 3.179273</td>
</tr>
<tr>
<td>11</td>
<td>-0.5650109</td>
<td>0.9926453</td>
<td>0.031783</td>
<td>-0.5839293</td>
<td>-2.646413 to 1.300388</td>
</tr>
<tr>
<td>12</td>
<td>-0.3620733</td>
<td>1.090265</td>
<td>0.03474</td>
<td>-0.3203626</td>
<td>-2.550097 to 1.717532</td>
</tr>
<tr>
<td>school</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>U0:sigma2</td>
<td>4.272626</td>
<td>1.299061</td>
<td>0.039697</td>
<td>4.122282</td>
<td>2.247659 to 7.220809</td>
</tr>
<tr>
<td>e.math5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>sigma2</td>
<td>28.23014</td>
<td>1.37812</td>
<td>0.017365</td>
<td>28.18347</td>
<td>25.63394 to 31.04375</td>
</tr>
</tbody>
</table>

Note: Default priors are used for model parameters.
Posterior distributions of the first 12 “random effects”:

.bayesgraph histogram {U0[1/12]}, by parm
Bayesian random-coefficient model:

.bayes: mixed math5 math3 || school: math3, covariance(unstructured)
note: Gibbs sampling is used for regression coefficients and variance components

Burn-in 2500 aaaaaaaaa1000aaaaaa2000aaaaa done
Simulation 10000 ........1000...........2000...........3000...........4000...........5
> 000...........6000...........7000...........8000...........9000...........10000 done

Multilevel structure

school
    {U0}: random intercepts
    {U1}: random coefficients for math3

Model summary

Likelihood:
    math5 ~ normal(xb_math5,{e.math5:sigma2})

Priors:
    {math5:math3 _cons} ~ normal(0,10000)
    (1)
    {U0}{U1} ~ mvnormal(2,{U:Sigma,m})
    (1)
    {e.math5:sigma2} ~ igamma(.01,.01)

Hyperprior:
    {U:Sigma,m} ~ iwishart(2,3,I(2))

(1) Parameters are elements of the linear form xb math5.
### Bayesian Multilevel Regression

**Model Specifications:**
- **MCMC iterations:** 12,500
- **Burn-in:** 2,500
- **MCMC sample size:** 10,000
- **Group variable:** school
  - Number of groups: 48
  - Number of obs: 887
  - Acceptance rate: 0.6985
  - Efficiency: min = 0.02935, avg = 0.1559, max = 0.5316

### Log Marginal Likelihood

**Equal-tailed**

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Mean</th>
<th>Std. Dev.</th>
<th>MCSE</th>
<th>Median</th>
<th>Equal-tailed [95% Cred. Interval]</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>School</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>U:Sigma_1_1</td>
<td>4.527905</td>
<td>1.363492</td>
<td>0.046275</td>
<td>4.345457</td>
<td>2.391319 7.765521</td>
</tr>
<tr>
<td>U:Sigma_2_1</td>
<td>-0.322247</td>
<td>0.1510543</td>
<td>0.004913</td>
<td>-0.3055407</td>
<td>-0.6683891 -0.0679181</td>
</tr>
<tr>
<td>U:Sigma_2_2</td>
<td>0.0983104</td>
<td>0.0280508</td>
<td>0.000728</td>
<td>0.0941222</td>
<td>0.0556011 0.1649121</td>
</tr>
<tr>
<td><strong>e.math5</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>sigma2</td>
<td>26.8091</td>
<td>1.34032</td>
<td>0.018382</td>
<td>26.76549</td>
<td>24.27881 29.53601</td>
</tr>
</tbody>
</table>

### Notes
- Default priors are used for model parameters.
Data: Time to hip fracture adjusted for age and for wearing a hip-protective device.

Bayesian exponential survival model:

```
. set seed 15
. bayes: streg protect age, distribution(exponential)
  failure _d: fracture
  analysis time _t: time1
  id: id
```

Burn-in ...
Simulation ...
Model summary

Likelihood:
  `_t ~ streg_exponential(xb__t)
Prior:
  `{_t:protect age _cons} ~ normal(0,10000)

(1) Parameters are elements of the linear form xb__t.
Bayesian exponential PH regression

Random-walk Metropolis-Hastings sampling

MCMC iterations = 12,500
Burn-in = 2,500
MCMC sample size = 10,000
Number of obs = 206

No. of subjects = 148
No. of failures = 37
No. at risk = 1703

Acceptance rate = .1927
Efficiency: min = .05694
avg = .07511
max = .086

Log marginal likelihood = -106.19703

<table>
<thead>
<tr>
<th>_t</th>
<th>Haz. Ratio</th>
<th>Std. Dev.</th>
<th>MCSE</th>
<th>Median</th>
<th>[95% Cred. Interval]</th>
</tr>
</thead>
<tbody>
<tr>
<td>protect</td>
<td>.1279039</td>
<td>.0447223</td>
<td>.001525</td>
<td>.1189394</td>
<td>.0616285 .2328919</td>
</tr>
<tr>
<td>age</td>
<td>1.086308</td>
<td>.0372036</td>
<td>.001559</td>
<td>1.085883</td>
<td>1.018374 1.159326</td>
</tr>
<tr>
<td>_cons</td>
<td>.0043577</td>
<td>.0352772</td>
<td>.001229</td>
<td>.0002529</td>
<td>2.05e-06 .0224516</td>
</tr>
</tbody>
</table>

Note: _cons estimates baseline hazard.
Note: Default priors are used for model parameters.

Store results for later comparison:

. bayes, saving(exp_mcmc)
note: file exp_mcmc.dta saved
. estimates store exp
Bayesian Weibull model:

. set seed 15
. bayes, saving(weib_mcmc): streg protect age, distribution(weibull)
    failure _d: fracture
    analysis time _t: time1
    id: id

Burn-in ...  
Simulation ...  
file weib_mcmc.dta saved

Model summary

Likelihood:
    _t ~ streg_weibull(xb__t,{ln_p})

Priors:
    {_t:protect age _cons} ~ normal(0,10000) (1)
    {ln_p} ~ normal(0,100000)

(1) Parameters are elements of the linear form xb__t.
Bayesian Weibull PH regression
Random-walk Metropolis-Hastings sampling

MCMC iterations = 12,500
Burn-in = 2,500
MCMC sample size = 10,000
Number of obs = 206

No. of subjects = 148
No. of failures = 37
No. at risk = 1703
Acceptance rate = .368
Efficiency: min = .05571
avg = .09994
max = .1767
Log marginal likelihood = -107.88854

<table>
<thead>
<tr>
<th></th>
<th>Haz. Ratio</th>
<th>Std. Dev.</th>
<th>MCSE</th>
<th>Median</th>
<th>Equal-tailed [95% Cred. Interval]</th>
</tr>
</thead>
<tbody>
<tr>
<td>t</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>_t</td>
<td>protect</td>
<td>.0956023</td>
<td>.0338626</td>
<td>.001435</td>
<td>.0899154</td>
</tr>
<tr>
<td></td>
<td>age</td>
<td>1.103866</td>
<td>.0379671</td>
<td>.001313</td>
<td>1.102685</td>
</tr>
<tr>
<td></td>
<td>_cons</td>
<td>.0075815</td>
<td>.0411427</td>
<td>.000979</td>
<td>.000567</td>
</tr>
<tr>
<td>ln_p</td>
<td></td>
<td>.4473869</td>
<td>.1285796</td>
<td>.004443</td>
<td>.4493192</td>
</tr>
</tbody>
</table>

Note: Estimates are transformed only in the first equation.
Note: _cons estimates baseline hazard.
Note: Default priors are used for model parameters.

. estimates store weib
Fitting Bayesian regression models using the `bayes` prefix

Bayesian survival models

Weibull model, group-specific shape parameters

- **Bayesian Weibull model with group-specific shape parameters:**

  ```stata
  . set seed 15
  . bayes, saving(weib_anc_mcmc): streg protect age, distrib(weibull) ancillary(male)
     (failure _d: fracture
      analysis time _t: time1
      id: id)
  
  Burn-in ...
  Simulation ...
  
  file weib_anc_mcmc.dta saved
  
  Model summary
  
  Likelihood:
  _t ~ streg_weibull(xb__t,xb_ln_p)
  
  Priors:
  {_t:protect age _cons} ~ normal(0,10000) (1)
  {ln_p:male _cons} ~ normal(0,10000) (2)
  
  (1) Parameters are elements of the linear form xb__t.
  (2) Parameters are elements of the linear form xb_ln_p.
  ```
Bayesian Weibull PH regression
Random-walk Metropolis-Hastings sampling

MCMC iterations = 12,500
Burn-in = 2,500
MCMC sample size = 10,000
Number of obs = 206

No. of subjects = 148
No. of failures = 37
No. at risk = 1703

Acceptance rate = 0.136
Efficiency: min = 0.006093
avg = 0.02061
max = 0.03044

Log marginal likelihood = -102.48

<table>
<thead>
<tr>
<th></th>
<th>Mean</th>
<th>Std. Dev.</th>
<th>MCSE</th>
<th>Median</th>
<th>[95% Cred. Interval]</th>
</tr>
</thead>
<tbody>
<tr>
<td>_t</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>protect</td>
<td>-2.108707</td>
<td>.3616945</td>
<td>.024969</td>
<td>-2.078421</td>
<td>-2.870089 -1.437823</td>
</tr>
<tr>
<td>age</td>
<td>.0920509</td>
<td>.0330708</td>
<td>.001896</td>
<td>.0944527</td>
<td>.0324366 .1559498</td>
</tr>
</tbody>
</table>

| ln_p    |        |           |       |        |                      |
| male    | -.5933872 | .2344015  | .016873 | -.5561411 | -1.171869 -.247341  |
| _cons   | .4002401  | .1083398  | .013879 | .4053514  | .1776803 .6014997   |

Note: Default priors are used for model parameters.
Note: Adaptation tolerance is not met in at least one of the blocks.
. estimates store weib_anc
Model comparison using Bayes factors:

```
.bayesstats ic weib Anc exp weib
Bayesian information criteria

<table>
<thead>
<tr>
<th></th>
<th>DIC</th>
<th>log(ML)</th>
<th>log(BF)</th>
</tr>
</thead>
<tbody>
<tr>
<td>weib Anc</td>
<td>147.9772</td>
<td>-102.48</td>
<td>.</td>
</tr>
<tr>
<td>exp</td>
<td>171.2604</td>
<td>-106.197</td>
<td>-3.717029</td>
</tr>
<tr>
<td>weib</td>
<td>162.7683</td>
<td>-107.8885</td>
<td>-5.408532</td>
</tr>
</tbody>
</table>
```

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

Weibull model with group-specific shape parameters is strongly preferable to the other models because $\log(\text{BF})$s are negative and $|2 \times \log(\text{BF})| > 6$. 
As of Stata 15, you can use `bayes:` to fit Bayesian regression models more conveniently.

You can continue using `bayesmh` for fitting more general Bayesian models or for programming your own.

Unlike `bayesmh`, `bayes:` provides default priors. You should always evaluate the choice of priors and use the ones appropriate for your model and research question.

All Bayesian postestimation features are available after `bayes:`.

For a full list of commands supported by `bayes:`, see www.stata.com/features/overview/bayesian-estimation/
