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

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`.

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
-------------+----------------------------------   F(1, 885)       =    341.40
       Model |  10960.2737         1  10960.2737   Prob > F        =    0.0000
    Residual |  28411.6181       885  32.1035233   R-squared       =    0.2784
-------------+----------------------------------   Adj R-squared   =    0.2776
       Total |  39371.8918       886  44.4378011   Root MSE        =     5.666

------------------------------------------------------------------------------
       math5 |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       math3 |   .6081306   .0329126    18.48   0.000     .5435347    .6727265
       _cons |   30.34656   .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)                                    (1)
             {sigma2} ~ igamma(.01,.01)
------------------------------------------------------------------------------
(1) Parameters are elements of the linear form xb_math5.
```

```
Bayesian linear regression                       MCMC iterations  =     12,500
Random-walk Metropolis-Hastings sampling         Burn-in          =      2,500
                                                 MCMC sample size =     10,000
                                                 Number of obs    =        887
                                                 Acceptance rate  =      .3312
                                                 Efficiency:  min =      .1099
                                                              avg =      .1529
Log marginal likelihood = -2817.2335                          max =      .2356
```

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

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 $\boldsymbol{\theta}$ 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 $\boldsymbol{\theta}$ formulated as a **prior distribution** $p(\boldsymbol{\theta})$.
- After data y are observed, the information about $\boldsymbol{\theta}$ is updated based on the **likelihood** $f(y|\boldsymbol{\theta})$.
- Information is updated by using the Bayes rule to form a **posterior distribution** $p(\boldsymbol{\theta}|y)$:

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

where $p(y)$ is the **marginal distribution** of the data y.

Inference

- Estimating a posterior distribution $p(\boldsymbol{\theta}|y)$ 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** (CrI)—(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.

- 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**:

$$p(y^{new}|y) = \int f(y^{new}|\boldsymbol{\theta})p(\boldsymbol{\theta}|y)d\boldsymbol{\theta}$$

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

Command	*Description*
Estimation	
`bayes:`	Bayesian regression models using the bayes prefix (new in Stata 15)
`bayesmh`	General Bayesian models using MH
`bayesmh` *evaluators*	User-written Bayesian models using MH
Postestimation	
`bayesgraph`	Graphical convergence diagnostics
`bayesstats ess`	Effective sample sizes and more
`bayesstats summary`	Summary statistics
`bayesstats ic`	Information criteria and Bayes factors
`bayestest model`	Model posterior probabilities
`bayestest interval`	Interval hypothesis testing

- 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!

Bayesian linear regression

- 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                       MCMC iterations  =     12,500
Random-walk Metropolis-Hastings sampling         Burn-in          =      2,500
                                                 MCMC sample size =     10,000
                                                 Number of obs    =        887
                                                 Acceptance rate  =      .3312
                                                 Efficiency:  min =      .1099
                                                              avg =      .1529
Log marginal likelihood = -2817.2335                          max =      .2356
```

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

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)                                      (1)
             {sigma2} ~ igamma(1,2)
------------------------------------------------------------------------------
(1) Parameters are elements of the linear form xb_math5.
```

```
Bayesian linear regression                       MCMC iterations  =     12,500
Random-walk Metropolis-Hastings sampling         Burn-in          =      2,500
                                                 MCMC sample size =     10,000
                                                 Number of obs    =        887
                                                 Acceptance rate  =      .3503
                                                 Efficiency:  min =      .1189
                                                              avg =      .1471
Log marginal likelihood = -2815.3081                          max =      .2005
```

	Mean	Std. Dev.	MCSE	Median	Equal-tailed [95% Cred.	Interval]
math5						
math3	.6076875	.033088	.000948	.6076282	.5405233	.673638
_cons	30.326	.1931568	.005602	30.32804	29.93212	30.70529
sigma2	32.09694	1.530839	.034185	32.03379	29.27687	35.37723

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

Burn-in ...
Simulation ...

Model summary
------------------------------------------------------------------------------
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.
```

```
Bayesian linear regression                        MCMC iterations  =     12,500
Random-walk Metropolis-Hastings sampling          Burn-in          =      2,500
                                                  MCMC sample size =     10,000
                                                  Number of obs    =        887
                                                  Acceptance rate  =      .3401
                                                  Efficiency:  min =      .1034
                                                               avg =      .1405
Log marginal likelihood = -2806.8234                           max =       .211
```

	Mean	Std. Dev.	MCSE	Median	Equal-tailed [95% Cred.	Interval]
math5						
math3	.6064431	.0306863	.000954	.6068399	.5455701	.6676897
_cons	30.34391	.1856718	.005676	30.3475	29.96434	30.71451
sigma2	32.15952	1.55488	.033853	32.10525	29.25887	35.33335

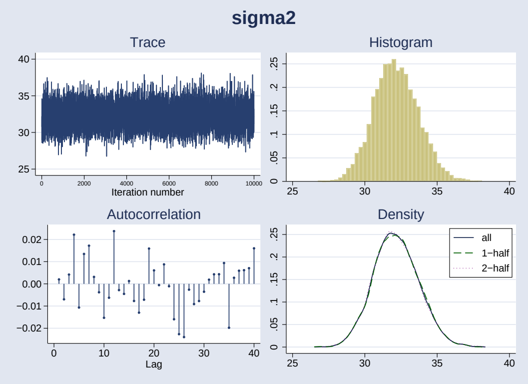
- Use more efficient Gibbs sampling:

```
. set seed 15

. bayes, gibbs: regress math5 math3

Bayesian linear regression                    MCMC iterations  =     12,500
Gibbs sampling                                Burn-in          =      2,500
                                              MCMC sample size =     10,000
                                              Number of obs    =        887
                                              Acceptance rate  =          1
                                              Efficiency:  min =          1
                                                           avg =          1
Log marginal likelihood =  -2817.184                       max =          1

                                                           Equal-tailed
             |      Mean   Std. Dev.     MCSE     Median  [95% Cred. Interval]
-------------+-----------------------------------------------------------------
math5        |
       math3 |  .6085104   .0333499   .000333   .6087819   .5426468   .6731657
       _cons |  30.34419   .1916673   .001917   30.34441   29.97587   30.72617
-------------+-----------------------------------------------------------------
      sigma2 |  32.16765   1.551119   .015511   32.10778     29.238   35.29901
-------------------------------------------------------------------------------
Note: Default priors are used for model parameters.
```

Postestimation

- All Bayesian postestimation features work after `bayes`: just like they do after `bayesmh`.

```
. bayesgraph diagnostics {sigma2}
```



Bayesian autoregressive models

- 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.
```

```
Bayesian linear regression                          MCMC iterations  =     12,500
Random-walk Metropolis-Hastings sampling            Burn-in          =      2,500
                                                    MCMC sample size =     10,000
                                                    Number of obs    =        107
                                                    Acceptance rate  =      .3285
                                                    Efficiency:  min =      .1199
                                                                 avg =      .1448
Log marginal likelihood = -75.889709                             max =      .1905
```

	Mean	Std. Dev.	MCSE	Median	Equal-tailed [95% Cred.	Interval]
lcoal						
lcoal L1.	.7143121	.0649968	.001877	.7123857	.5884089	.8436602
_cons	-.6896604	.1561023	.004433	-.6935272	-.9970502	-.3879924
sigma2	.1702592	.0243144	.000557	.1672834	.1299619	.2248287

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.
```

```
Bayesian linear regression                        MCMC iterations  =     12,500
Random-walk Metropolis-Hastings sampling          Burn-in          =      2,500
                                                  MCMC sample size =     10,000
                                                  Number of obs    =        106
                                                  Acceptance rate  =      .3614
                                                  Efficiency:  min =     .07552
                                                               avg =      .1172
Log marginal likelihood = -82.507817                           max =      .1966
```

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

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:

```
. 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.

```
. bayes,  prior({lcoal:L1.lcoal}, normal(0, cond({lag}>=1,100,0.01))) ///
>         prior({lcoal:L2.lcoal}, normal(0, cond({lag}>=2,100,0.01))) ///
>         prior({lcoal:L3.lcoal}, normal(0, cond({lag}>=3,100,0.01))) ///
>         prior({lcoal:L4.lcoal}, normal(0, cond({lag}>=4,100,0.01))) ///
>         prior({lcoal:L5.lcoal}, normal(0, cond({lag}>=5,100,0.01))) ///
>         prior({lag}, index(0.2,0.2,0.2,0.2,0.2)):                   ///
>         regress lcoal L(1/5).lcoal
note: operator L1. is replaced with L. in parameter name L1.lcoal

Burn-in ...
Simulation ...
Model summary
------------------------------------------------------------------------------
Likelihood:
  lcoal ~ regress(xb_lcoal,{sigma2})
Priors:
   {lcoal:L.lcoal} ~ normal(0,cond({lag}>=1,100,0.01))                     (1)
  {lcoal:L2.lcoal} ~ normal(0,cond({lag}>=2,100,0.01))                     (1)
  {lcoal:L3.lcoal} ~ normal(0,cond({lag}>=3,100,0.01))                     (1)
  {lcoal:L4.lcoal} ~ normal(0,cond({lag}>=4,100,0.01))                     (1)
  {lcoal:L5.lcoal} ~ normal(0,cond({lag}>=5,100,0.01))                     (1)
     {lcoal:_cons} ~ normal(0,10000)                                       (1)
          {sigma2} ~ igamma(.01,.01)
Hyperprior:
  {lag} ~ index(0.2,0.2,0.2,0.2,0.2)
------------------------------------------------------------------------------
(1) Parameters are elements of the linear form xb_lcoal.
```

```
Bayesian linear regression                       MCMC iterations  =     12,500
Random-walk Metropolis-Hastings sampling         Burn-in          =      2,500
                                                 MCMC sample size =     10,000
                                                 Number of obs    =        103
                                                 Acceptance rate  =        .34
                                                 Efficiency:  min =    .002852
                                                              avg =     .04431
Log marginal likelihood = -8.2084752                          max =      .1716
```

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

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                     Number of obs     =        887
Group variable: school                          Number of groups  =         48

                                                Wald chi2(1)      =     347.92
Log likelihood = -2767.8923                     Prob > chi2       =     0.0000

------------------------------------------------------------------------------
       math5 |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       math3 |   .6088066   .0326392    18.65   0.000     .5448349    .6727783
       _cons |   30.36495   .3491544    86.97   0.000     29.68062    31.04928
------------------------------------------------------------------------------

------------------------------------------------------------------------------
  Random-effects Parameters  |   Estimate   Std. Err.     [95% Conf. Interval]
-----------------------------+------------------------------------------------
school: Identity             |
                  var(_cons) |   4.026853   1.189895      2.256545    7.186004
-----------------------------+------------------------------------------------
               var(Residual) |   28.12721    1.37289       25.5611    30.95094
------------------------------------------------------------------------------
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                  MCMC iterations  =     12,500
Metropolis-Hastings and Gibbs sampling          Burn-in          =      2,500
                                                MCMC sample size =     10,000
Group variable: school                          Number of groups =         48

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

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

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.
```

```
Bayesian multilevel regression                  MCMC iterations  =     12,500
Metropolis-Hastings and Gibbs sampling          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
Log marginal likelihood                                      max =      .6298
```

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

Note: Default priors are used for model parameters.

- Display estimates of the first 12 “random effects”:

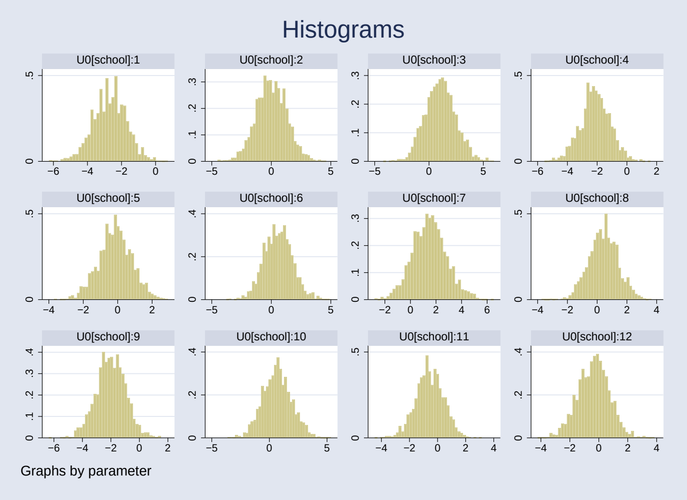
```
. bayes, showreffects({U0[1/12]})
```

	Mean	Std. Dev.	MCSE	Median	Equal-tailed [95% Cred.	Interval]
math5						
math3	.6087689	.0326552	.000436	.6087444	.5450837	.6729982
_cons	30.39202	.3597873	.01961	30.38687	29.67802	31.10252
U0[school]						
1	-2.685824	.9776969	.031227	-2.672364	-4.633162	-.7837494
2	.015465	1.290535	.03201	.0041493	-2.560203	2.556316
3	1.049006	1.401383	.033731	1.021202	-1.534088	3.84523
4	-2.123055	.9921679	.028859	-2.144939	-4.069283	-.1507593
5	-.1504003	.9650027	.033881	-.1468966	-2.093015	1.721503
6	.5833945	1.192379	.032408	.5918357	-1.660335	3.049718
7	1.490231	1.332917	.033846	1.481793	-1.095757	4.272903
8	.4198105	.9783772	.031891	.4579817	-1.496317	2.403908
9	-1.996105	1.02632	.035372	-2.001467	-4.037044	-.0296276
10	.6736806	1.249238	.031114	.660939	-1.70319	3.179273
11	-.5650109	.9926453	.031783	-.5839293	-2.646413	1.300388
12	-.3620733	1.090265	.033474	-.3203626	-2.550097	1.717532
school						
U0:sigma2	4.272626	1.299061	.039697	4.122282	2.247659	7.220809
e.math5						
sigma2	28.23014	1.37812	.017365	28.18347	25.63394	31.04375

Note: Default priors are used for model parameters.

- Posterior distributions of the first 12 “random effects”:

```
. bayesgraph histogram {U0[1/12]}, byparm
```



- 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 aaaaaaaaa1000aaaaaaaaa2000aaaaa 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                   MCMC iterations  =     12,500
Metropolis-Hastings and Gibbs sampling           Burn-in          =      2,500
                                                 MCMC sample size =     10,000
Group variable: school                           Number of groups =         48

                                                 Number of obs    =        887
                                                 Acceptance rate  =      .6985
                                                 Efficiency:  min =     .02935
                                                              avg =      .1559
Log marginal likelihood                                       max =      .5316
```

	Mean	Std. Dev.	MCSE	Median	Equal-tailed [95% Cred.	Interval]
math5						
math3	.6234197	.0570746	.002699	.6228624	.5144913	.7365849
_cons	30.34691	.3658515	.021356	30.34399	29.62991	31.07312
school						
U:Sigma_1_1	4.527905	1.363492	.046275	4.345457	2.391319	7.765521
U:Sigma_2_1	-.322247	.1510543	.004913	-.3055407	-.6683891	-.0679181
U:Sigma_2_2	.0983104	.0280508	.000728	.0941222	.0556011	.1649121
e.math5						
sigma2	26.8091	1.34032	.018382	26.76549	24.27881	29.53601

Note: 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)
------------------------------------------------------------------------------
(1) Parameters are elements of the linear form xb__t.
```

```
Bayesian exponential PH regression                MCMC iterations  =     12,500
Random-walk Metropolis-Hastings sampling          Burn-in          =      2,500
                                                  MCMC sample size =     10,000
No. of subjects =         148                     Number of obs    =        206
No. of failures =          37
No. at risk     =        1703
                                                  Acceptance rate  =      .1927
                                                  Efficiency:  min =     .05694
                                                               avg =     .07511
Log marginal likelihood = -106.19703                           max =       .086

------------------------------------------------------------------------------
             |                                                Equal-tailed
          _t | Haz. Ratio   Std. Dev.     MCSE     Median  [95% Cred. Interval]
-------------+----------------------------------------------------------------
     protect |   .1279039   .0447223   .001525   .1189394   .0616285   .2328919
         age |   1.086308   .0372036   .001559   1.085883   1.018374   1.159326
       _cons |   .0043577   .0352772   .001229   .0002529   2.05e-06   .0224516
------------------------------------------------------------------------------
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(0,10000)
------------------------------------------------------------------------------
(1) Parameters are elements of the linear form xb__t.
```

```
Bayesian Weibull PH regression                  MCMC iterations  =     12,500
Random-walk Metropolis-Hastings sampling        Burn-in          =      2,500
                                                MCMC sample size =     10,000
No. of subjects =         148                   Number of obs    =        206
No. of failures =          37
No. at risk     =        1703
                                                Acceptance rate  =       .368
                                                Efficiency:  min =     .05571
                                                             avg =     .09994
Log marginal likelihood = -107.88854                         max =      .1767
```

					Equal-tailed	
	Haz. Ratio	Std. Dev.	MCSE	Median	[95% Cred.	Interval]
_t						
protect	.0956023	.0338626	.001435	.0899154	.0463754	.1787249
age	1.103866	.0379671	.001313	1.102685	1.033111	1.180283
_cons	.0075815	.0411427	.000979	.000567	4.02e-06	.0560771
ln_p	.4473869	.1285796	.004443	.4493192	.1866153	.6912467

```
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
```

- Bayesian Weibull model with group-specific shape parameters:

```
. 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                     MCMC iterations  =     12,500
Random-walk Metropolis-Hastings sampling           Burn-in          =      2,500
                                                   MCMC sample size =     10,000
No. of subjects =          148                     Number of obs    =        206
No. of failures =           37
No. at risk     =         1703
                                                   Acceptance rate  =       .136
                                                   Efficiency:  min =    .006093
                                                                avg =     .02061
Log marginal likelihood =    -102.48                            max =     .03044
```

	Mean	Std. Dev.	MCSE	Median	Equal-tailed [95% Cred.	Interval]
_t						
protect	-2.108707	.3616945	.024969	-2.078421	-2.870089	-1.437823
age	.0920509	.0330708	.001896	.0944527	.0324366	.1559498
_cons	-9.881823	2.472154	.152612	-9.976053	-14.53088	-5.076762
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
```

	DIC	log(ML)	log(BF)
weib_anc	147.9772	-102.48	.
exp	171.2604	-106.197	-3.717029
weib	162.7683	-107.8885	-5.408532

```
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(BF)`s are negative and $|2 \times \texttt{log}(\texttt{BF})| > 6$.

- 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/
- See **[BAYES] bayes** and www.stata.com/new-in-stata/bayes-prefix/ for more examples.

References

Harvey, A. C. 1989. *Forecasting, Structural Time Series Models, and the Kalman Filter*. Cambridge: Cambridge University Press.

Mortimore, P., P. Sammons, L. Stoll, D. Lewis, and R. Ecob. 1988. *School Matters: The Junior Years*. Wells, Somerset, UK: Open Books.