Title

stata.com

bmacoefsample — Posterior samples of regression coefficients

Description

`bmacoefsample` simulates regression coefficients and other model parameters from their posterior distributions after `bmaregress`.

Quick start

Simulate regression coefficients and other model parameters from their posterior distributions after `bmaregress`, and use `bayesstats summary` to display posterior summaries for `{y:}`, all 12 regression coefficients, and the intercept, including their 90% highest posterior density (HPD) credible intervals (CrIs)

```
bmaregress y x1-x12, saving(bmamodelmcmc)
bmacoefsample
bayesstats summary {y:}, clevel(90) hpd
```

Same as above, but specify a random-number seed and the size of the simulated dataset, and save the simulated dataset in file `bmacoefmcmc.dta`

```
bmacoefsample, rseed(1234) mcmcsize(5000) saving(bmacoefmcmc)
```

Simulate results without saving first, but then later specify option `saving()` to save the previously simulated results in file `bmacoefmcmc1.dta`, and save a new set of results in `bmacoefmcmc2.dta`

```
bmacoefsample, rseed(1234)
bmacoefsample, saving(bmacoefmcmc1)
bmacoefsample, saving(bmacoefmcmc2)
```

Same as above, but save only one set of results, use option `simulate` to resimulate the results, and save the new ones in `bmacoefmcmc1`, replacing already existing file

```
bmacoefsample, rseed(1234)
bmacoefsample, saving(bmacoefmcmc1, replace) simulate
```

Menu

Statistics > Bayesian model averaging > Posterior samples of coefficients

Syntax

```
bmacoefsample [ , coefsimopts ]
```

coefsimopts	Description
Simulation	
mcmcsize(#)	Markov chain Monte Carlo (MCMC) sample size; default is mcmcsize(10000) or as specified with bmaregress
rseed(#)	random-number seed
Reporting	
saving(*filename*[, replace])	save simulation results to *filename*.dta
[no]dots	suppress or display dots every 1,000 iterations and iteration numbers every 5,000 iterations; default is dots
dots(#[, every(#)])	display dots as simulation is performed
simulate	simulate new results instead of saving existing ones

simulate does not appear in the dialog box.

Options

Simulation

mcmcsize(#) specifies the target MCMC sample size. The default is mcmcsize(10000) or as specified in bmaregress's mcmcsize() option.

rseed(#) sets the random-number seed. This option can be used to reproduce results. Option rseed(#) is equivalent to typing set seed # prior to calling the bmacoefsample command; see [R] **set seed**.

Reporting

saving(*filename*[, replace]) saves simulation results in *filename*.dta. The replace option specifies to overwrite *filename*.dta if it exists. If the saving() option is not specified, bmacoefsample saves simulation results in a temporary file for later access. This temporary file will be overwritten every time bmacoefsample or bmaregress is run and will be erased if the current estimation results are cleared. The simulation dataset has the same structure as described in option saving() of [BAYES] **bayesmh**. The simulation results include posterior samples of regression coefficients, the intercept, and the error variance.

If you first run bmacoefsample without option saving(), you can rerun it with this option later to save the current simulation results in a file without having to redo a potentially time-consuming simulation. If, for some reason, you need to simulate a new sample in this case, you should specify option simulate in addition to saving().

dots, nodots, dots(#), and dots(#, every(#)) specify to display or suppress dots during simulation. dots, the default, displays dots every 1,000 iterations and iteration numbers every 5,000 iterations; it is a synonym for dots(1000, every(5000)). dots(#) displays a dot every # iterations. If dots(..., every(#)) is specified, then an iteration number is displayed every #th iteration instead of a dot. dots(, every(#)) is equivalent to dots(1, every(#)).

The following option is available with bmacoefsample but is not shown in the dialog box:

simulate is used with option saving() to simulate a new set of results and save them in a file instead of saving the previously simulated results. This option is implied with bmacoefsample in all but one case—when option saving() is used and simulation results from a previous run of bmacoefsample have not been saved in a file already.

Remarks and examples

stata.com

Bayesian model averaging (BMA) simulation consists of two steps: 1) simulation of a model space and, optionally, 2) simulation of model parameters, which include regression coefficients, the intercept, and the error variance. The first, main step is performed by bmaregress, during which a posterior sample of models is obtained. In addition, a posterior sample of parameter g of a Zellner's g-prior, which affects the simulation of the model space, is obtained. Because bmaregress reports only posterior means and standard deviations of model parameters, which are available analytically, it does not automatically simulate posterior samples of model parameters to save time. But these samples may be needed if we want to obtain CrIs, functions of regression coefficients, or certain predictions.

The bmacoefsample command simulates model parameters of a BMA linear regression fit by bmaregress from their posterior distributions. The simulated data include posterior samples of regression coefficients, the intercept, and the error variance. For convenience, the data also include the posterior sample of the g parameter simulated by bmaregress. To use bmacoefsample after bmaregress, you must first save the BMA model simulation results by using the saving() option with bmaregress either during estimation or on replay.

bmacoefsample differs from a standard Stata postestimation command—it is a hybrid between an estimation and a postestimation command. Unlike other Stata postestimation commands, bmacoefsample actually modifies or, rather, augments the estimation results stored by bmaregress. But it is still a postestimation command because it can run only after bmaregress is run. You can think of it as a "replay option" of the bmaregress command, which continues the simulation process started by bmaregress. But it is more convenient as a stand-alone command.

In addition to storing results in e(), bmacoefsample has its own concept of "replay". Because it performs a potentially time-consuming simulation, it allows you to save your simulation results later, if you did not do so the first time you ran the command. That is,

```
. bmaregress ..., saving(bmamodelmcmc)
. bmacoefsample
```

produces a simulation sample of model parameters but saves them temporarily. If you run bmacoefsample or bmaregress again, these results will be lost. You can use bmacoefsample's saving() option to save the current model parameter simulation results permanently in a file:

```
. bmacoefsample, saving(mycoef)
```

Once the current results are permanently saved, the next run of bmacoefsample will automatically produce a new sample. In fact, if you use any option with bmacoefsample other than saving(), it will automatically trigger the generation of a new sample. You can also force the bmacoefsample command to resample and save a new set of simulation results by specifying the simulate option together with saving():

```
. bmacoefsample
. bmacoefsample, saving(mycoef) simulate
```

For a fixed g, the samples generated by bmaregress and bmacoefsample are independent. For a random g, these MCMC samples are dependent.

Inference based on MCMC simulations depends on the variation of the MCMC sample. Specifically, the size of the MCMC sample, that is the number of draws from parameters' posterior distributions, affects the precision of the parameter estimates. The larger the MCMC sample size, the more precise the estimates. The MCMC sample size is specified in the `mcmcsize()` option. The default is 10,000 or the value specified in `bmaregress`'s `mcmcsize()` option.

`bmacoefsample` is a stochastic command, so you need to specify the `rseed()` option to reproduce results.

After `bmacoefsample`, you can use the following standard Bayesian postestimation commands to explore posterior summaries of model parameters: [BAYES] **bayesstats summary**, [BAYES] **bayesstats ess**, [BAYES] **bayesstats ppvalues**, [BAYES] **bayesgraph**, and [BAYES] **bayespredict**. You can also use the `estimates` command, except `estimates table`, `estimates stats`, and `estimates selected`; see [BMA] **BMA postestimation**.

For examples of how to use this command, see, for instance, example 5, example 16, and example 19 of [BMA] **bmaregress**.

Stored results

In addition to the estimation results stored by `bmaregress`, `bmacoefsample` stores the following in `e()`:

Scalars

`e(mcmcsize2)`	MCMC sample size

Macros

`e(cmd2)`	`bmacoefsample`
`e(filename)`	name of the file with simulation results
`e(parnames)`	names of model parameters
`e(postvars)`	variable names corresponding to model parameters in `e(parnames)`
`e(scparams)`	scalar model parameters
`e(pareqmap)`	model parameters in display order
`e(rngstate2)`	random-number state at the time of simulation

Methods and formulas

Methods and formulas are presented under the following headings:

Enumerated model space
Simulated model space

Conditional on a model M and parameter g, the posterior distribution of the error variance σ^2 is an inverse-gamma distribution, and the posterior distribution of regression coefficients and the intercept is a multivariate location-scale t-distribution; see *Conditional posterior distribution of model parameters* in *Methods and formulas* of [BMA] **bmaregress**.

`bmacoefsample` simulates model parameters from their posterior distributions in two ways, which depend on whether `bmaregress` enumerated or sampled the model space.

Enumerated model space

When all models in the model space are visited by `bmaregress`, the posterior model distribution is known, and PMPs can be computed exactly. Also, Zellner's g parameter is fixed with model enumeration. In this case, `bmacoefsample` generates an independent sample of size T for model parameters as follows.

For $t = 1, 2, \ldots, T$:

1. Draw a model M_t from the model space according to the PMP distribution, which is described in *Posterior model probability* in *Methods and formulas* of [BMA] **bmaregress**.

2. Draw model parameters according to

$$\begin{aligned} \sigma_t^2|g, M_t &\sim \text{InverseGamma}\left(\frac{n-1}{2}, \frac{s_{t,\delta}^2}{2}\right) \\ \boldsymbol{\theta}_{t,p_t}|\sigma_t^2, g, M_t &\sim N_{p_t+1}(\boldsymbol{\mu}_t, \sigma_t^2\boldsymbol{\Sigma}_t) \end{aligned}$$

where n is the sample size and $s_{t,\delta}^2$, $\boldsymbol{\mu}_t$, and $\boldsymbol{\Sigma}_t$ are computed according to (1), (3), and (4) in *Conditional posterior distribution of model parameters* in *Methods and formulas* of [BMA] **bmaregress** (with index j substituted for t, that is, conditional on model M_t instead of M_j). $\boldsymbol{\theta}_{t,p_t} = (\boldsymbol{\beta}_{t,p_t}', \alpha)'$ is a $(p_t + 1) \times 1$ vector of the p_t regression coefficients included in model M_t and the intercept. Regression coefficients that are not included in M_t are assigned 0s. Then a $(p + 1) \times 1$ vector $\boldsymbol{\theta}_t$ includes all p regression coefficients and the intercept from iteration t.

The result is a sample $\{\boldsymbol{\theta}_t, \sigma_t^2\}_{t=1}^T$ from the BMA posterior distribution of model parameters.

Simulated model space

For a random g, the exact posterior model distribution is not available, and `bmaregress` provides an MCMC sample (m_t, g_t), for $t = 1, 2, \ldots, T$, from the BMA posterior distribution of models and parameter g. `bmacoefsample` command reuses this sample to simulate model parameters.

For each t, regression coefficients that are not included in m_t are assigned 0s. The remaining p_t regression coefficients and the intercept, $\boldsymbol{\theta}_{t,p_t}$, and variance σ_t^2 are drawn according to

$$\begin{aligned} \sigma_t^2|g_t, m_t &\sim \text{InverseGamma}\left(\frac{n-1}{2}, \frac{s_{t,\delta}^2}{2}\right) \\ \boldsymbol{\theta}_{t,p_t}|\sigma_t^2, g_t, m_t &\sim N_{p_t+1}(\boldsymbol{\mu}_t, \sigma_t^2\boldsymbol{\Sigma}_t) \end{aligned}$$

where n is the sample size and $s_{t,\delta}^2$, $\boldsymbol{\mu}_t$, and $\boldsymbol{\Sigma}_t$ are computed according to (1), (3), and (4) in *Conditional posterior distribution of model parameters* in *Methods and formulas* of [BMA] **bmaregress** (with index j substituted for t, that is, conditional on model m_t instead of M_j).

Let $\boldsymbol{\theta}_t$ be a $(p + 1) \times 1$ vector that includes all p regression coefficients and the intercept from iteration t. Then the result is a sample $\{\boldsymbol{\theta}_t, \sigma_t^2\}_{t=1}^T$ from the BMA posterior distribution of model parameters.

Also see

[BMA] **bmaregress** — Bayesian model averaging for linear regression

[BMA] **BMA postestimation** — Postestimation tools for Bayesian model averaging

[BMA] **Glossary**

[BAYES] **Bayesian postestimation** — Postestimation tools for bayesmh and the bayes prefix

Stata, Stata Press, and Mata are registered trademarks of StataCorp LLC. Stata and Stata Press are registered trademarks with the World Intellectual Property Organization of the United Nations. Other brand and product names are registered trademarks or trademarks of their respective companies. Copyright © 1985–2023 StataCorp LLC, College Station, TX, USA. All rights reserved.