Title

BMA commands — Introduction to commands for Bayesian model averaging

Description

This entry describes a suite of commands to perform Bayesian model averaging (BMA). BMA is a statistical procedure that performs inference and computes predictions by combining results from multiple models according to Bayesian principles. It accounts for model uncertainty and thus can provide more reliable inference and prediction than traditional methods that choose one model. See *Brief motivation* in *Remarks and examples* of [BMA] **Intro** for a quick overview of BMA.

Setup

splitsample	split data into random samples for training, validation, and prediction
vl	manage large variable lists conveniently

Estimation

bmaregress	BMA linear regression
bmacoefsample	Posterior samples of regression coefficients

Graphical commands

bmagraph	Graphical summaries
bmagraph pmp	Model-probability plots
bmagraph varmap	Variable-inclusion maps
bmagraph msize	Model-size distribution plots
bmagraph coefdensity	Coefficient density plots

Postestimation statistics

bmastats	Posterior summaries
bmastats msize	Model-size summary
bmastats models	Posterior model and variable-inclusion summaries
bmastats pip	Posterior inclusion probabilities for predictors
bmastats jointness	Jointness measures for predictors
bmastats lps	Log predictive-score

Predictions

bmapredict	BMA predictions

Remarks and examples

Here we provide a brief overview of the BMA commands implemented in Stata. See [BMA] **Intro** for an introduction to the BMA methodology and *Usage of BMA*, in particular, for various applications of BMA.

The `bmaregress` estimation command implements BMA for linear regression. In the regression context, the BMA model space consists of 2^p regression models formed by all possible combinations of the inclusion or exclusion of each of the p predictors. The model is considered a discrete random parameter with a prior distribution over the model space. You can choose from a variety of prior model distributions. The model parameters—the regression coefficients, intercept, and error variance—are also random, as they are in standard Bayesian analysis. The intercept and error variance are assumed to have noninformative priors, and the regression coefficients are assumed to have a Zellner's g-prior. The g parameter controls the shrinkage of coefficients toward zero and can be fixed or random. `bmaregress` supports many fixed choices and hyperpriors for this parameter.

`bmaregress` explores the model space either by enumerating all possible regression models or by sampling from the posterior model distribution. The sampling is based on specialized Markov chain Monte Carlo (MCMC) algorithms. When sampling is used, you must check convergence of the results before continuing with inference and prediction. This can be done visually with the `bmagraph pmp` command; [BMA] **bmagraph pmp**.

`bmaregress` also offers various modeling options such as always including certain predictors and keeping groups of predictors together in the models. It supports factor variables and offers several ways of handling main effects and interactions during model building. See [BMA] **bmaregress** for details about the command.

`bmaregress` reports posterior summaries (over the model space) for regression coefficients as well as posterior inclusion probabilities (PIPs) for the predictors. The PIP measures the importance of a predictor in explaining the outcome over the model space. You can use `bmastats pip` to report PIPs for a subset of predictors; [BMA] **bmastats pip**.

After `bmaregress`, you can use postestimation commands described in [BMA] **bmastats** and [BMA] **bmagraph** for inference about models and predictors.

`bmastats models` summarizes models with high posterior model probabilities, PMPs, and the predictors they include; see [BMA] **bmastats models**. `bmagraph pmp` and `bmagraph varmap` are its graphical counterparts; see [BMA] **bmagraph pmp** and [BMA] **bmagraph varmap**.

`bmastats msize` and `bmagraph msize` provide the descriptive and graphical summaries, respectively, for the posterior model-size distribution. These are useful to explore the complexity of the BMA model; see [BMA] **bmastats msize** and [BMA] **bmagraph msize**.

The concept of jointness is of particular interest in BMA. It describes the tendency of a pair of variables to be included in the models together, separately, or independently. You can use the `bmastats jointness` command to compute various measures of jointness; see [BMA] **bmastats jointness**.

The `bmaregress` command computes and reports the posterior means and standard deviations for regression coefficients. To compute other posterior summaries such as credible intervals, you need a posterior sample of regression coefficients. You can use the `bmacoefsample` command to generate this sample of coefficients (and other model parameters); see [BMA] **bmacoefsample**. Once the sample is available, many standard Bayesian postestimation commands can be used to summarize this sample such as `bayesstats summary`. See [BMA] **BMA postestimation** for the full list.

The posterior distribution of the regression coefficients can be visualized by `bmagraph coefdensity`. With a fixed g, you can use this command directly after `bmaregress`. With a random g, an MCMC sample of model parameters must be simulated first by using `bmacoefsample`. See [BMA] **bmagraph coefdensity** for details.

Another important application of BMA is prediction. You can use `bmapredict` to compute various predictions after `bmaregress`; see [BMA] **bmapredict**. As with some other postestimation commands, certain predictions rely on an MCMC sample of model parameters. For those, you need to run

bmacoefsample first. If the sample is available, you can also use the bayespredict command to compute more complicated predictions; [BAYES] **bayespredict**.

Finally, you can use the bmastats lps command to compare model fit and predictive performance of BMA models by using the log predictive-score (LPS). The fit of models is typically compared by computing LPS for the estimation sample. Predictive performance is evaluated by computing LPS for the out-of-sample observations, which were not used during estimation. See [BMA] **bmastats lps** for details.

For examples of BMA commands, see the following.

Getting started. *Motivating examples* in *Remarks and examples* of [BMA] **Intro** provides a quick introduction to and motivation for the use of bmaregress. *Getting started examples* in *Remarks and examples* of [BMA] **bmaregress** provides a tour of various BMA analysis in Stata.

Model choice and inference. *BMA analysis of cross-country economic growth data* in *Remarks and examples* of [BMA] **bmaregress** demonstrates the use of BMA for inference and model choice, including the investigation of jointness of predictors.

Prediction. The use of BMA for prediction is described in *BMA predictive performance for the USA crime rate data* in *Remarks and examples* of [BMA] **bmaregress** and example of [BMA] **bmapredict**. The evaluation of a model's predictive performance using the LPS is demonstrated in *Remarks and examples* in [BMA] **bmastats lps** with the application of predicting the systolic blood pressure.

Acknowledgments

Reference

De Luca, G., and J. R. Magnus. 2011. Bayesian model averaging and weighted-average least squares: Equivariance, stability, and numerical issues. *Stata Journal* 11: 518–544.

Also see

[BMA] **Intro** — Introduction to Bayesian model averaging

[BMA] **Glossary**

[BAYES] **Bayesian commands** — Introduction to commands for Bayesian analysis