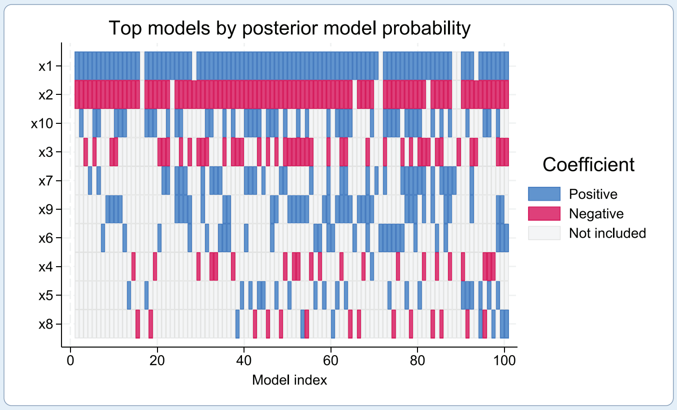
Stata Features

Bayesian model averaging

Uncertain which predictors to include in your regression model?
Would you like to account for this uncertainty in your anaysis?
Want to learn about influential models and important predictors?

Stata's Bayesian model averaging (BMA) suite can help.

- Model choice, inference, and prediction
- Influential models using posterior model probabilities (PMPs)
- Important predictors using posterior inclusion probabilities (PIPs)
- Uniform, binomial, and beta-binomial model priors
- Many *g*-priors: fixed, robust, hyper-*g*, ...
- Posterior inference, including credible intervals, for coefficients and predictions
- Jointness measures for pairs of predictors
- Variable-inclusion maps
- Predictive performance using log predictive-score (LPS)
- And more



Stata's **bma** suite performs BMA, which combines results from multiple candidate models weighted by models' probabilities given the observed data. This leads to more reliable inference and prediction that accounts for model uncertainty.

BMA workflow

Fit BMA linear regression

```
. bmaregress y x1-x100
```

Save BMA simulation results

```
. bmaregress, saving(bmamcmc)
```

Check BMA convergence

```
. bmagraph pmp
```

Identify influential models by using PMP

```
. bmastats models
```

Identify important predictors by using PIP

```
. bmastats pip x10-x20
```

Visually explore important predictors by producing variable-inclusion maps

```
. bmagraph varmap
```

Explore model-size distributions

```
. bmastats msize
. bmagraph msize
```

Simulate posterior distributions of model parameters

```
. bmacoefsample, saving(bmacoef)
```

Obtain posterior summaries of model parameters

```
. bayesstats summary
```

Plot posterior distributions of coefficients

```
. bmagraph coefdensity {x1} {x2}
```

Generate predictions and predictive credible intervals

```
. bmapredict ypmean, mean
. bmapredict y_cril y_criu, cri
```

Compare predictive performance of BMA models using LPS

```
. bmastats lps bma1 bma2 if testsample == 1
```

BMA linear regression

Fit a BMA linear regression of **y** on **x1** through **x40** to explore 2^{40} potential models. Use the default **Beta-binomial(1, 1)** model prior uniform over the model size and the **Hyper-g(3)** prior for the *g* parameter of Zellner's *g* prior. Save simulation results, specify a random-number seed for reproducibility, and display results only for predictors with a PIP of at least 0.1.

```
. bmaregress y x1-x40, gprior(hyperg 3) saving(bmamcmc) rseed(18)
  pipcutoff(0.1)

Burn-in ...
Simulation ...
Computing model probabilities ...

Bayesian model averaging                      No. of obs         =    200
Linear regression                             No. of predictors  =     40
MC3 and adaptive MH sampling                              Groups =     40
                                                          Always =      0
                                              No. of models      =    314
                                                 For CPMP >= .9  =    130
Priors:                                       Mean model size    =  5.123
  Models: Beta-binomial(1, 1)                 Burn-in            =  2,500
   Cons.: Noninformative                      MCMC sample size   = 10,000
   Coef.: Zellner's g                         Acceptance rate    = 0.4854
       g: Hyper-g(3)
  sigma2: Noninformative                      Mean sigma2        =  1.023

Sampling correlation = 0.9207

-----------------------------------------------------------------------------
           y |     Mean  Std. dev.                       Group          PIP
-------------+---------------------------------------------------------------
          x2 |  1.15429   .0728323                           2            1
         x10 | -4.730936   .185496                          10            1
         x37 |  .032751   .0145241                          37        .8906
         x35 | .0245493   .0189709                          35        .6767
         x22 | .0258066   .0331922                          22         .414
         x34 | .0092896   .0150387                          34        .3072
         x39 | .0060078   .0132044                          39        .1943
         x40 | .0026092   .0074414                          40        .1289
-------------+---------------------------------------------------------------
Always       |
       _cons | .5235217   .0787887                           0            1
-----------------------------------------------------------------------------
Note: Coefficient posterior means and std. dev. estimated from 314
      models.
Note: Default prior is used for models.
Note: 32 predictors with PIP less than .1 not shown.

-----------------------------------------------------------------------------
             |     Mean  Std. dev.      MCSE    Median  [95% cred. interval]
-------------+---------------------------------------------------------------
           g | 1436.719   4472.467   119.288  857.0709   300.7584   5156.588
   Shrinkage | .9986692    .000813   .000027  .9988346   .9966861   .9998061
-----------------------------------------------------------------------------
```

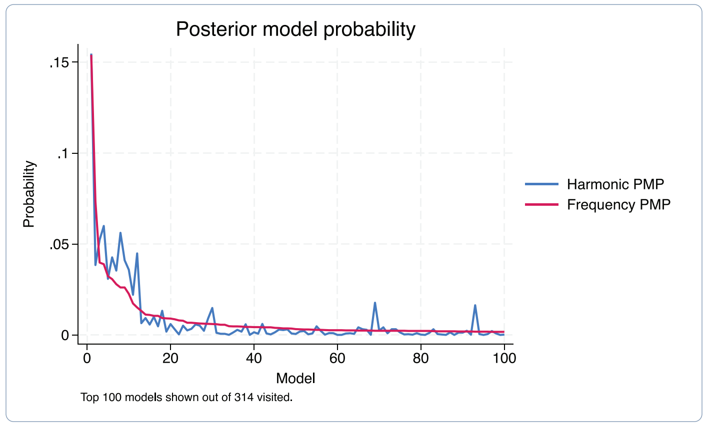
bmaregress explored 314 models with an average model size of 5.123. Predictors **x2** and **x10** with PIPs of 1 were included in essentially all considered models. And roughly 90% of the models included predictor **x37**. There are also 32 predictors with a PIP less than 0.1, which are not shown in the output.

Because we specified a random prior for the *g* parameter, we also see the posterior summaries for it and the shrinkage parameter, which is $g/(g + 1)$. The shrinkage is close to 1, so there is little shrinkage of coefficients toward 0 in this model.

Posterior model probabilities (PMPs)

Check BMA convergence, and explore the number of models with high PMP

```
. bmagraph pmp
```



Explore influential models ranked by their PMPs

Viewer - view bmamodels.smcl

view bmamodels.smcl

Dialog Also see Jump to

```
. bmastats models

Computing model probabilities ...

Model summary            Number of models:
                                  Visited = 314
                                 Reported =   5

--------------------------------------------
             |  Frequency PMP     Model size
-------------+------------------------------
Rank         |
           1 |           .154              4
           2 |          .0734              4
           3 |          .0398              5
           4 |          .0389              5
           5 |          .0323              6
--------------------------------------------

Variable-inclusion summary

--------------------------------------------------------------
             |   Rank     Rank     Rank     Rank     Rank
             |     1        2        3        4        5
-------------+------------------------------------------------
          x2 |     x        x        x        x        x
         x10 |     x        x        x        x        x
         x35 |     x                 x        x        x
         x37 |     x        x        x        x        x
         x22 |              x                 x        x
         x39 |                       x
         x34 |                                         x
--------------------------------------------------------------
Legend:
  x - estimated
```

CAP NUM INS

By default, **bmastats models** shows the top 5 models ranked by their PMP, but we can specify the **top()** option to see more models. The model with the highest PMP of 0.154 includes the predictors **x2**, **x10**, **x35**, and **x37**. The model with the next-highest PMP of 0.0734 includes all the same predictors, except that **x22** is included instead of **x35**. The remaining listed models have similar PMPs below 0.05.

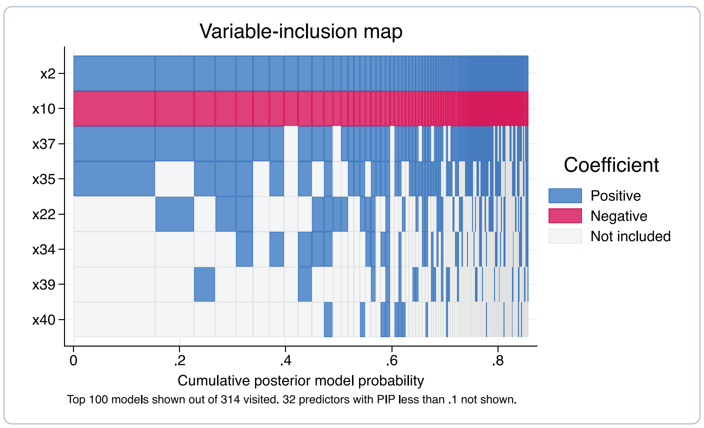
Posterior inclusion probabilities (PIPs)

Identify important predictors

```
. bmastats pip, pipcutoff(0.5)

Posterior inclusion probability (PIP)

No. of obs         =    200
No. of predictors  =     40
           Groups  =     40
           Always  =      0
         Reporting =      4
No. of models      =    314
Mean model size    =  5.123

------------------------------
             |    PIP    Group
-------------+----------------
          x2 |      1        2
         x10 |      1       10
         x37 |  .8906       37
         x35 |  .6767       35
-------------+----------------
Always       |
       _cons |      1        0
------------------------------
Note: Using frequency PMPs.
Note: 36 predictors with PIP less than
      .5 not shown.
```

Variable-inclusion maps

. bmagraph varmap, pipcutoff(0.1)



Jointness measures

Explore jointness for pairs of predictors

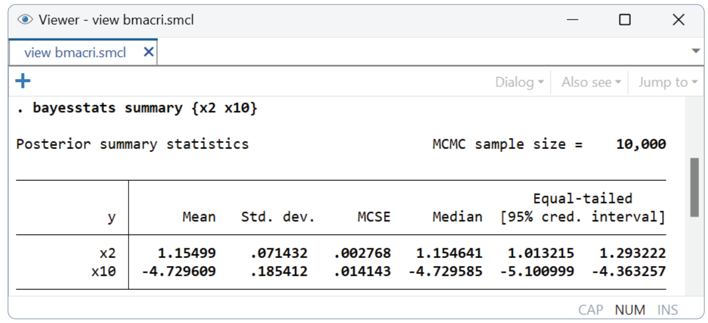
```
. bmastats jointness x37 x22

Variables: x37 x22

--------------------------------
                  |   Jointness
------------------+-------------
  Doppelhofer–Weeks |  .8903163
     Ley–Steel type 1 |  .4228378
     Ley–Steel type 2 |  .7326153
            Yule's Q |  .4179109
   Modified Yule's Q |  .4173874
--------------------------------
Notes: Using frequency PMPs. See
       thresholds
```

Posterior summaries: Credible intervals

Simulate posterior distributions of model parameters, including regression coefficients

. bmacoefsample, saving(coefmcmc) rseed(18)

Obtain posterior summaries, including posterior means and credible intervals

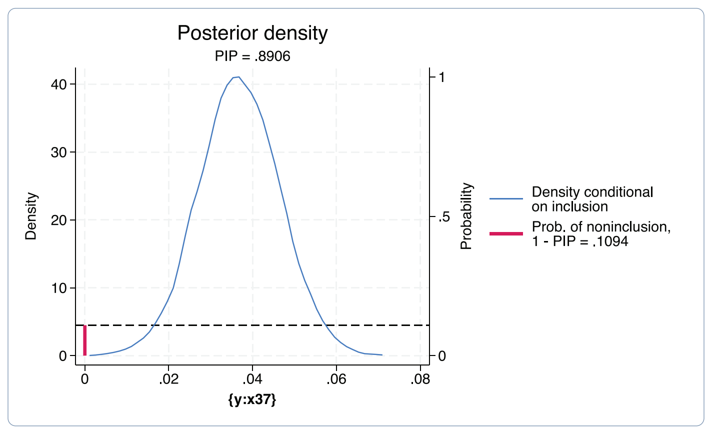



```
. bayesstats summary {x2 x10}

Posterior summary statistics                      MCMC sample size =    10,000

---------------------------------------------------------------------------------
             |                                                 Equal-tailed
           y |      Mean   Std. dev.     MCSE     Median  [95% cred. interval]
-------------+-------------------------------------------------------------------
          x2 |   1.15499    .071432   .002768   1.154641   1.013215   1.293222
         x10 | -4.729609    .185412   .014143  -4.729585  -5.100999  -4.363257
---------------------------------------------------------------------------------
```

Coefficient posterior density plots

Plot posterior density for coefficient of predictor **x37**

. bmagraph coefdensity {x37}



And much more

Generate posterior predictive means

. bmapredict ypmean, mean

Generate posterior predictive credible intervals

. bmapredict y_cril y_criu, cri

Check the model's performance

. bmastats lps if testsample == 1

stata.com/bma