Bayesian model averaging (BMA)

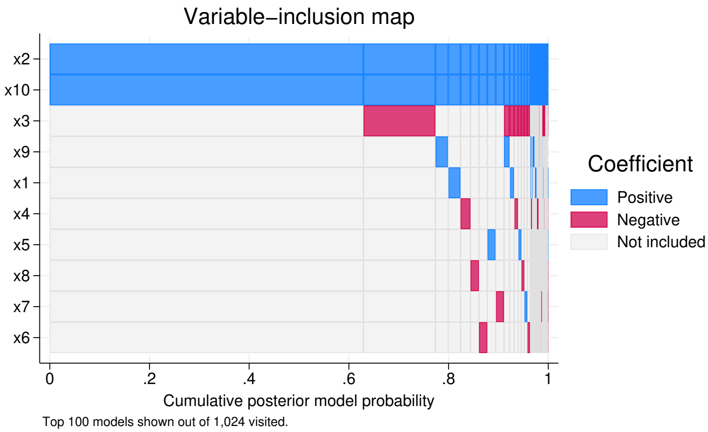
Yulia Marchenko

Vice President, Statistics and Data Science
StataCorp LLC

2023 Stata Conference

Outline

Teaser


Variable–inclusion map
x2
x10
x3
x9
x1
x4
x5
x8
x7
x6
0
.2
.4
.6
.8
1
Cumulative posterior model probability
Top 100 models shown out of 1,024 visited.
Coefficient
Positive
Negative
Not included

What is Bayesian model averaging (BMA)?

- The concept of *uncertainty* is fundamental to statistical analyses.
- We assess uncertainty about parameter estimates, predictions, hypothesis testing, etc.
- We often assume there is a true data-generating model (DGM), which we infer from the observed data.
- Traditionally, we select a model that fits the data well and proceed with our analysis. This typically does not incorporate uncertainty about the selected model.
- Model averaging accounts for *model* uncertainty in data analyses.
- BMA (Leamer 1978, Hoeting et al. 1999) uses the Bayesian principles, specifically the Bayes theorem, to account for model uncertainty.

Why BMA?

- Sometimes we may have a strong evidence for selecting a certain model for our data analysis.
- More often, however, there may be several plausible models that support our theory.
- In that case, choosing only one model may lead to overly optimistic or even wrong conclusions (if the selected model is drastically different from the true DGM).
- Model averaging considers a set of candidate models and accounts for model uncertainty by averaging the estimates across the models and weighting them according to how likely each model is.
- BMA uses posterior model probabilities (PMPs) as weights, which provide an intuitive and unified across analyses way to interpret models' importance.

- BMA also provides a way to assess a variable's importance by using posterior inclusion probabilities (PIPs) and interrelations between variables across the model space.
- BMA can be used for sensitivity analyses of the importance of different models and predictors.
- BMA can be used for model choice, prediction, and inference.
- See **[BMA] Intro** for details.
- Also see, for instance, Steel (2020) and Moral-Benito (2015) for a systematic review of BMA.

Brief review of Bayesian analysis

- 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** $p(\boldsymbol{\theta}) = \pi(\boldsymbol{\theta})$.
- 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** $p(\boldsymbol{\theta}|y)$:

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

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

- 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** (CrIs)—(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.
- Predictions and model checking are based on a **posterior predictive distribution**:

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

BMA for linear regression

- I'll focus on BMA in the context of a (*simpler*) linear regression:

$$y_i = \alpha + \beta_1 x_{1i} + \beta_2 x_{2i} + \epsilon_i,\ \epsilon_i \sim N(0,\sigma^2),\ i = 1, 2, \ldots, n$$

- Model uncertainty in the context of a linear regression with p predictors amounts to selecting predictors in a model.
- For instance, with $p = 2$ predictors, there are $2^p = 4$ possible models (ignoring potential interaction and nonlinear terms; see **Regression modeling and model space** in *Introduction to BMA linear regression* of **[BMA] bmaregress**):

$$\begin{aligned}
M_1\text{: } y_i &= \alpha + \epsilon_i^{(1)} \\
M_2\text{: } y_i &= \alpha + \beta_1^{(2)} x_{1i} + \epsilon_i^{(2)} \\
M_3\text{: } y_i &= \alpha + \beta_2^{(3)} x_{2i} + \epsilon_i^{(3)} \\
M_4\text{: } y_i &= \alpha + \beta_1^{(4)} x_{1i} + \beta_2^{(4)} x_{2i} + \epsilon_i^{(4)}
\end{aligned}$$

- By construction, $\beta_1^{(1)} = \beta_2^{(1)} = \beta_2^{(2)} = \beta_1^{(3)} = 0$.

- In matrix notation,

$$\mathbf{y} = \alpha \mathbf{1}_n + \mathbf{X}_j \boldsymbol{\beta}_j + \boldsymbol{\epsilon}_j$$

where $\mathbf{X}_j$ and $\boldsymbol{\beta}_j$ are predictors and regression coefficients specific to model M_j.

- Priors for parameters:

$$\begin{aligned} \boldsymbol{\beta}_j | \alpha, \sigma, M_j &\sim N(\mathbf{0}, g\sigma^2(\mathbf{X}_j'\mathbf{X}_j)^{-1}) \\ \alpha | \sigma, M_j &\sim 1 \\ \sigma | M_j &\sim \sigma^{-1} \end{aligned}$$

- Priors for models: BMA treats model M_j as random with a discrete prior $\mathrm{P}(M_j)$ for $j = 1, 2, \ldots, p$.
- Priors for g: fixed value or random hyperprior $p(g)$.

BMA fundamentals

- Posterior distribution of $\boldsymbol{\beta}$ over the model space:

$$g(\boldsymbol{\beta}|\mathbf{y}) = \sum_{j=1}^{2^p} \mathrm{P}(M_j|\mathbf{y}) g(\boldsymbol{\beta}|\mathbf{y}, M_j)$$

- From the Bayes theorem applied to the model space, PMP is defined as

$$\mathrm{P}(M_j|\mathbf{y}) = \frac{f(\mathbf{y}|M_j)\mathrm{P}(M_j)}{p(\mathbf{y})}$$

where $f(\mathbf{y}|M_j)$ is the likelihood of $\mathbf{y}$ under model M_j and $p(\mathbf{y})$ is the marginal probability/likelihood over the model space.

- BMA linear regression coefficient estimates:

$$\widehat{\beta}_1^{\mathrm{BMA}} = \sum_{j=1}^{4} \widehat{\mathrm{P}}(M_j|y)\widehat{\beta}_1^{(j)}$$
$$\widehat{\beta}_2^{\mathrm{BMA}} = \sum_{j=1}^{4} \widehat{\mathrm{P}}(M_j|y)\widehat{\beta}_2^{(j)}$$

- $\widehat{\mathrm{P}}(M_j|y)$ is the estimate of the posterior probability of model M_j (probability of M_j given the observed data y).
- $\widehat{\beta}_1^{(j)}$ and $\widehat{\beta}_2^{(j)}$ are the posterior mean estimates of regression coefficients from model M_j.
- The above BMA estimates correspond to the estimates of posterior means of regression coefficients over the model space, $E(\boldsymbol{\beta}|\mathbf{y})$, based on $g(\boldsymbol{\beta}|\mathbf{y})$.

Toy example

- See **[BMA]** for various real-world BMA examples.
- Simulated data: $n = 200$; $p = 10$; x1 through x10 are independent standard normal.
- DGM:

$$\mathrm{y} = 0.5 + 1.2 \times \mathrm{x2} + 5 \times \mathrm{x10} + N(0,1)$$

```
. webuse bmaintro
(Simulated data for BMA example)

. summarize
```

Variable	Obs	Mean	Std. dev.	Min	Max
y	200	.9944997	4.925052	-13.332	13.06587
x1	200	-.0187403	.9908957	-3.217909	2.606215
x2	200	-.0159491	1.098724	-2.999594	2.566395
x3	200	.080607	1.007036	-3.016552	3.020441
x4	200	.0324701	1.004683	-2.410378	2.391406
x5	200	-.0821737	.9866885	-2.543018	2.133524
x6	200	.0232265	1.006167	-2.567606	3.840835
x7	200	-.1121034	.9450883	-3.213471	1.885638
x8	200	-.0668903	.9713769	-2.871328	2.808912
x9	200	-.1629013	.9550258	-2.647837	2.472586
x10	200	.083902	.8905923	-2.660675	2.275681

```
. bmaregress y x1-x10
Enumerating models ...
Computing model probabilities ...

Bayesian model averaging                       No. of obs         =    200
Linear regression                              No. of predictors  =     10
Model enumeration                                          Groups =     10
                                                           Always =      0
Priors:                                        No. of models      =  1,024
  Models: Beta-binomial(1, 1)                      For CPMP >= .9 =      9
   Cons.: Noninformative                       Mean model size    =  2.479
   Coef.: Zellner´s g
       g: Benchmark, g = 200                   Shrinkage, g/(1+g) = 0.9950
  sigma2: Noninformative                       Mean sigma2        =  1.272

------------------------------------------------------------------------------
           y |      Mean   Std. dev.                          Group        PIP
-------------+----------------------------------------------------------------
          x2 |  1.198105   .0733478                               2          1
         x10 |   5.08343   .0900953                              10          1
          x3 | -.0352493   .0773309                               3     .21123
          x9 |   .004321   .0265725                               9    .051516
          x1 |  .0033937   .0232163                               1    .046909
          x4 | -.0020407   .0188504                               4    .039267
          x5 |  .0005972   .0152443                               5    .033015
          x8 | -.0005639   .0153214                               8    .032742
          x7 | -8.23e-06    .015497                               7    .032386
          x6 | -.0003648   .0143983                               6    .032361
-------------+----------------------------------------------------------------
Always       |
       _cons |  .5907923   .0804774                               0          1
------------------------------------------------------------------------------
Note: Coefficient posterior means and std. dev. estimated from 1,024 models.
Note: Default priors are used for models and parameter g.
```

- Estimation: Model enumeration (few predictors, fixed g); $2^{10} = 1{,}024$ considered models.
- Default priors: Beta-binomial(1,1) for models and fixed $g = 200$.
- Little shrinkage: $g/(1+g) = 0.995$ close to 1.
- Mean model size is 2.48.
- Important predictors: Estimated PIPs of `x2` and `x10` are 1; others are small.
- BMA coefficient estimates for `x2` and `x10` (1.2 and 5.1) are close to the true values.
- BMA estimates of other coefficients are close to zero.
- BMA estimates are based on 1,024 models; see *Interpretation of BMA regression coefficients* in **[BMA] bmaregress**.

- Store BMA estimation results for later use:

```
. bmaregress, saving(bmareg)
note: file bmareg.dta saved.
. estimates store bmareg
```

- As with other Bayesian commands, we save the BMA MCMC simulation file first by using `bmaregress`'s `saving()` option (available on replay).
- We then use `estimates store` to save the BMA estimation results.

Classical linear regression

```
. regress y x1-x10
```

Source	SS	df	MS
Model	4607.24837	10	460.724837
Residual	219.723235	189	1.1625568
Total	4826.9716	199	24.2561387

Number of obs	=	200
F(10, 189)	=	396.30
Prob > F	=	0.0000
R-squared	=	0.9545
Adj R-squared	=	0.9521
Root MSE	=	1.0782

y	Coefficient	Std. err.	t	P>|t|	[95% conf.	interval]
x1	.0753537	.0781737	0.96	0.336	-.0788513	.2295587
x2	1.18854	.0716658	16.58	0.000	1.047172	1.329907
x3	-.1871012	.0789484	-2.37	0.019	-.3428344	-.0313679
x4	-.0459335	.0785503	-0.58	0.559	-.2008813	.1090144
x5	.0343498	.0793095	0.43	0.665	-.1220956	.1907953
x6	-.0149194	.0767357	-0.19	0.846	-.1662879	.136449
x7	.007174	.0831239	0.09	0.931	-.1567958	.1711437
x8	-.0384917	.0810626	-0.47	0.635	-.1983953	.1214119
x9	.0968948	.0817218	1.19	0.237	-.0643093	.2580989
x10	5.13251	.0877447	58.49	0.000	4.959426	5.305595
_cons	.617996	.0791152	7.81	0.000	.4619337	.7740582

- Compare the estimates:

	regress	bmaregress
y		
x1	0.075	0.003
	(0.078)	(0.023)
x2	1.189	1.198
	(0.072)	(0.073)
x3	-0.187	-0.035
	(0.079)	(0.077)
x4	-0.046	-0.002
	(0.079)	(0.019)
x5	0.034	0.001
	(0.079)	(0.015)
x6	-0.015	-0.000
	(0.077)	(0.014)
x7	0.007	-0.000
	(0.083)	(0.015)
x8	-0.038	-0.001
	(0.081)	(0.015)
x9	0.097	0.004
	(0.082)	(0.027)
x10	5.133	5.083
	(0.088)	(0.090)
_cons	0.618	0.591
	(0.079)	(0.080)
Number of observations	200	200

- BMA coefficients for "unimportant" predictors are shrunk toward zero.
- Let's continue with our BMA analysis:

```
. estimates restore bmareg
(results bmareg are active now)
```

Credible intervals (CrIs)

- For computational reasons, `bmaregress` does not compute CrIs by default.
- For fixed g, analytical closed-form formulas are available for BMA posterior means and standard deviations.
- The formulas for CrIs are not as straightforward; `bmaregress` computes them from the posterior sample of parameters.
- Obtaining the posterior sample of parameters requires a potentially time-consuming simulation and may not always be needed, depending on a BMA analysis objective.
- But this sample can be generated by using `bmacoefsample` following `bmaregress`.
- Many standard Bayesian postestimation commands such as `bayesstats summary` can then be used.

```
. bmacoefsample, rseed(18) mcmcsize(1000)
Simulation (1000): . done
. bayesstats summary
Posterior summary statistics                      MCMC sample size =     1,000
```

	Mean	Std. dev.	MCSE	Median	Equal-tailed [95% cred.	interval]
y						
x1	.0017904	.0176576	.000549	0	0	.0230942
x2	1.201273	.0695129	.00224	1.201107	1.06427	1.337961
x3	-.0361735	.0755013	.002435	0	-.2537224	0
x4	-.0010145	.0156635	.000495	0	0	0
x5	.0003393	.0114519	.000383	0	0	0
x6	-.0003742	.0145684	.000478	0	0	0
x7	.0002788	.0156012	.000423	0	0	0
x8	-.0003383	.0152805	.000483	0	0	0
x9	.0048314	.0291115	.000906	0	0	.0924737
x10	5.08115	.0841466	.002581	5.079152	4.913381	5.247999
_cons	.5879177	.0841129	.002632	.5879514	.4153159	.7560713
sigma2	1.273245	.1288853	.003956	1.266904	1.045943	1.55155
g	200	0	0	200	200	200

Influential models

- Compute PMPs to identify influential models:

```
. bmastats models
Computing model probabilities ...
Model summary          Number of models:
                                Visited = 1,024
                               Reported =     5

─────────────┬────────────────────────────────
             │   Analytical PMP     Model size
─────────────┼────────────────────────────────
Rank         │
           1 │            .6292              2
           2 │            .1444              3
           3 │            .0258              3
           4 │            .0246              3
           5 │           .01996              3
─────────────┴────────────────────────────────

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
          x3 │                 x
          x9 │                            x
          x1 │                                       x
          x4 │                                                  x
─────────────┴────────────────────────────────────────────────────────
Legend:
  x - estimated
```

- Cumulative PMPs (CPMPs):

```
. bmastats models, cumulative
Computing model probabilities ...
Model summary          Number of models:
                                Visited = 1,024
                               Reported =     5
```

	Analytical CPMP	Model size
Rank		
1	.6292	2
2	.7736	3
3	.7994	3
4	.824	3
5	.844	3

Variable-inclusion summary

	Rank 1	Rank 2	Rank 3	Rank 4	Rank 5
x2	x	x	x	x	x
x10	x	x	x	x	x
x3		x			
x9			x		
x1				x	
x4					x

Legend:
  x - estimated

- Specify a CPMP cutoff:

```
. bmastats models, cumulative(0.75)
Computing model probabilities ...
Model summary              Number of models:
                                    Visited = 1,024
                                   Reported =     2
```

	Analytical CPMP	Model size
Rank		
1	.6292	2
2	.7736	3

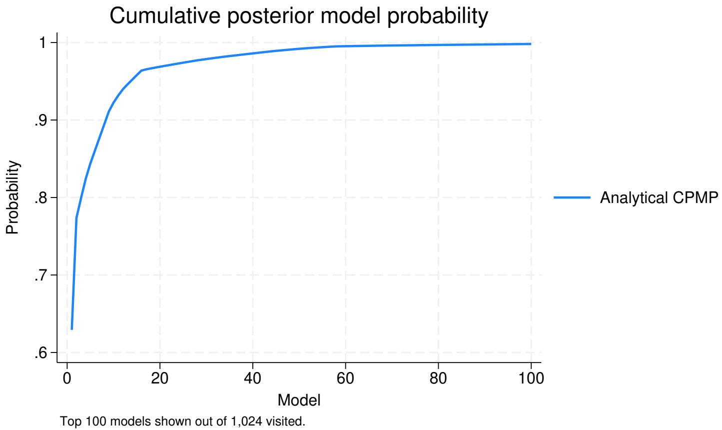
Variable-inclusion summary

	Rank 1	Rank 2
x2	x	x
x10	x	x
x3		x

```
Legend:
  x - estimated
```

- Plot CPMPs:

```
. bmagraph pmp, cumulative
note: frequency estimates not available with model enumeration; option
      nofreqline implied.
```



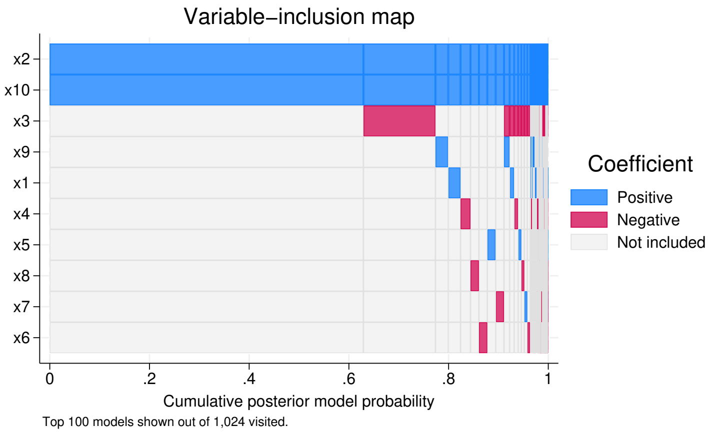
Important predictors

- Report PIPs:

```
. bmastats pip

Posterior inclusion probability (PIP)

No. of obs          =    200
No. of predictors   =     10
            Groups  =     10
            Always  =      0
          Reported  =     10
No. of models       =  1,024
Mean model size     =  2.479

-------------------------------------
             |        PIP       Group
-------------+-----------------------
          x2 |          1           2
         x10 |          1          10
          x3 |     .21123           3
          x9 |    .051516           9
          x1 |    .046909           1
          x4 |    .039267           4
          x5 |    .033015           5
          x8 |    .032742           8
          x7 |    .032386           7
          x6 |    .032361           6
-------------+-----------------------
Always       |
       _cons |          1           0
-------------------------------------
Note: Using analytical PMPs.
```

- Variable-inclusion map:

```
. bmagraph varmap
Computing model probabilities ...
```



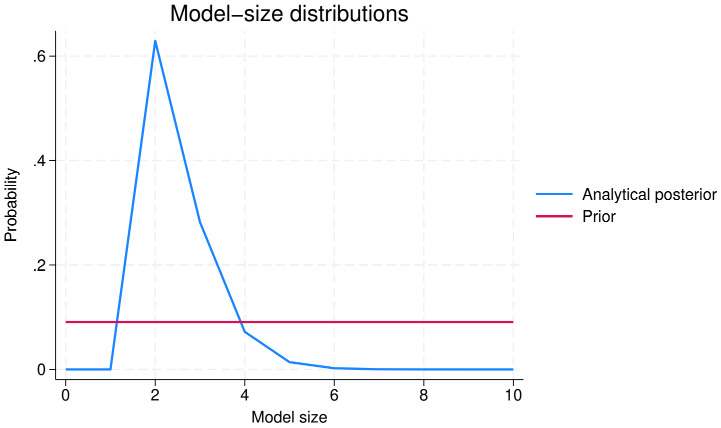
Model-size distribution

```
. bmastats msize

Model-size summary

Number of models = 1,024
Model size:
  Minimum =  0
  Maximum = 10
```

	Mean	Median
Prior		
Analytical	5.0000	5
Posterior		
Analytical	2.4794	2

```
Note: Frequency summaries not available.
```

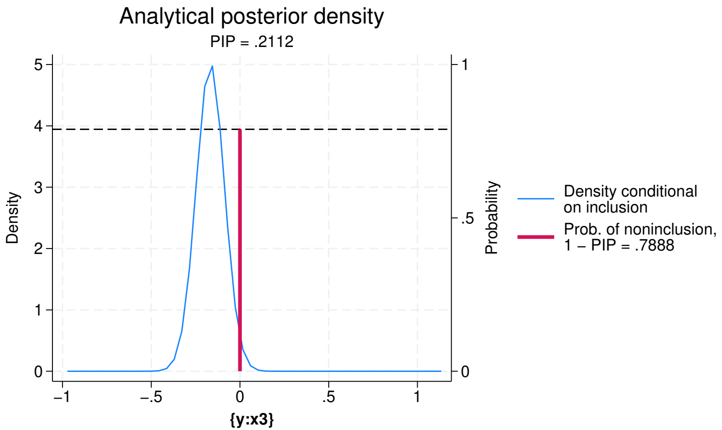
```
. bmagraph msize
note: frequency posterior model-size distribution not available.
```



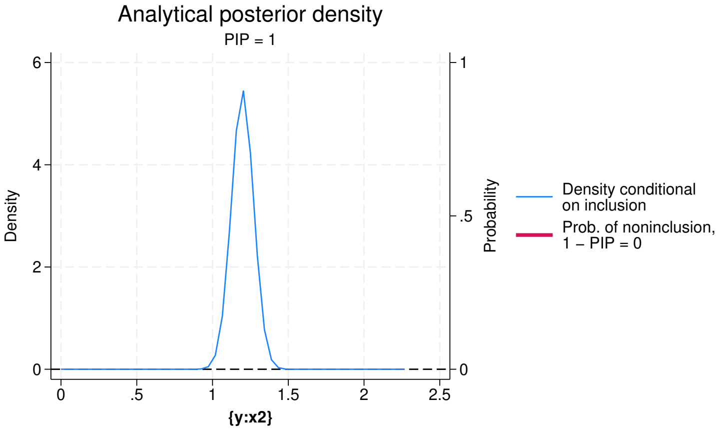
Posterior distribution of coefficients

- Mixture of a point mass at zero with $1 - PIP$ and a continuous density conditional on inclusion:

```
. bmagraph coefdensity {x3}
```



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



Jointness

- Tendency of the predictors to appear together, separately, or independently in the models:

```
. bmastats jointness x2 x10
Variables: x2 x10

──────────────────────────────────────
                      │   Jointness
──────────────────────┼───────────────
      Doppelhofer-Weeks │      75.947
       Ley-Steel type 1 │           1
       Ley-Steel type 2 │    3.59e+35
               Yule´s Q │           1
──────────────────────────────────────
Notes: Using analytical PMPs. See
       thresholds.
```

- x2 and x10 are strong *complements*—they tend to be included in the models together.
- Strong or decisive jointness; see **[BMA] bmastats jointness** for the thresholds or click on blue "thresholds" in the Stata output.

BMA predictions

- Posterior predictive means:

```
. bmapredict pmean, mean
note: computing analytical posterior predictive means.
```

- Predictive CrIs:

```
. bmacoefsample, saving(bmacoef)
note: saving existing MCMC simulation results without resampling; specify
      option simulate to force resampling in this case.
note: file bmacoef.dta saved.

. bmapredict cri_l cri_u, cri rseed(18)
note: computing credible intervals using simulation.

Computing predictions ...
```

- Summary:

```
. summarize y pmean cri*
```

Variable	Obs	Mean	Std. dev.	Min	Max
y	200	.9944997	4.925052	-13.332	13.06587
pmean	200	.9944997	4.783067	-13.37242	12.31697
cri_l	200	-1.24788	4.787499	-15.66658	10.03054
cri_u	200	3.227426	4.779761	-11.06823	14.58301

Sensitivity analysis: Random g-prior

- Random prior (hyperprior) for g instead of treating it as fixed.
- Hyperpriors are often suggested for robustness.
- Specify a hyper-g prior with hyperparameter 4 for g:

```
. bmaregress y x1-x10, gprior(hyperg 4) rseed(18)
Burn-in ...
Simulation ...
Computing model probabilities ...

Bayesian model averaging                          No. of obs        =    200
Linear regression                                 No. of predictors =     10
MC3 and adaptive MH sampling                                 Groups =     10
                                                             Always =      0
                                                  No. of models     =     27
                                                     For CPMP >= .9 =      2
Priors:                                           Mean model size   =  2.175
  Models: Beta-binomial(1, 1)                     Burn-in           =  2,500
   Cons.: Noninformative                          MCMC sample size  = 10,000
   Coef.: Zellner's g                             Acceptance rate   = 0.3838
       g: Hyper-g(4)
  sigma2: Noninformative                          Mean sigma2       =  1.184

Sampling correlation = 0.9985
```

y	Mean	Std. dev.	Group	PIP
x2	1.205111	.0706146	2	1
x10	5.101085	.0869608	10	1
x3	-.0153289	.0534981	3	.0921
x4	-.00075	.0112903	4	.0151
x9	.0010838	.0132084	9	.0137
x1	.0008948	.0118064	1	.0124
x5	.0002045	.008905	5	.0121
x6	-.0001291	.00818	6	.0111
Always				
_cons	.5871921	.0774449	0	1

Note: Coefficient posterior means and std. dev. estimated from 27 models.
Note: Default prior is used for models.
Note: 2 predictors with PIP less than .01 not shown.

	Mean	Std. dev.	MCSE	Median	Equal-tailed [95% cred.	interval]
g	1991.648	9547.263	186.39	1129.102	330.1158	7337.703
Shrinkage	.9989299	.0007563	.000016	.9991151	.9969799	.9998637

- Estimation: MC3 and adaptive MH sampling.
- Only 27 models explored compared with the total of 1,024.
- Mean model size is 2.18.
- The header now reports some standard MCMC summaries.
- The sampling correlation is also reported. (More about this later.)
- Analytical formulas are not available.
- BMA results are similar, but PIPs for all but the `x2` and `x10` coefficients are smaller.
- Parameter g (and shrinkage) are now random, and thus the posterior summaries are reported for them.
- Let's store these BMA results for later comparison:

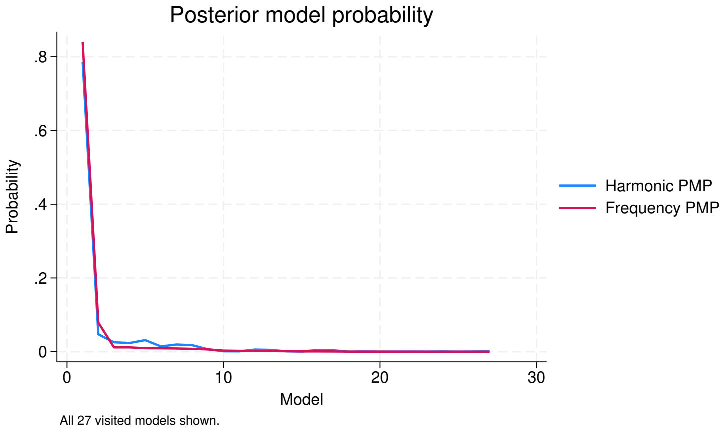
```
. bmaregress, saving(bmareg_hyperg)
note: file bmareg_hyperg.dta saved.
. estimates store bmareg_hyperg
```

Model convergence

- Sampling correlation is used to evaluate the MCMC convergence of the BMA model.
- This is the correlation between the analytical (whenever available) and frequency PMPs.
- The estimated sampling correlation of 0.9985 does not indicate any convergence issues.
- See *Convergence of BMA* in **[BMA] bmaregress** for details.

- We can also explore the BMA convergence visually:

```
. bmagraph pmp
```



Sensitivity analysis: Informative prior

- We can consider a more informative prior for the model space:

```
. bmaregress y x1-x10, mprior(binomial x2 x10 0.5 x1 x3-x9 0.05) saving(bmareg_
> inf)
Enumerating models ...
Computing model probabilities ...
Bayesian model averaging                          No. of obs          =    200
Linear regression                                 No. of predictors   =     10
Model enumeration                                             Groups  =     10
                                                              Always  =      0
Priors:                                           No. of models       =  1,024
  Models: Binomial, IP varies                         For CPMP >= .9  =      1
   Cons.: Noninformative                          Mean model size     =  2.062
   Coef.: Zellner´s g
       g: Benchmark, g = 200                      Shrinkage, g/(1+g)  = 0.9950
  sigma2: Noninformative                          Mean sigma2         =  1.277
```

y	Mean	Std. dev.	Group	PIP
x2	1.201574	.0729557	2	1
x10	5.080061	.0899387	10	1
x3	-.0051795	.0320662	3	.031299
Always				
_cons	.5879401	.0803296	0	1

```
Note: Coefficient posterior means and std. dev. estimated from 1,024 models.
Note: Default prior is used for parameter g.
Note: 7 predictors with PIP less than .01 not shown.
file bmareg_inf.dta saved.
. estimates store bmareg_inf
```

Log predictive-score (LPS)

- LPS is the negative of the log of the posterior predictive density evaluated at an observation.
- The smaller the LPS value, the better the model fit.
- We can use LPS to compare the model fit of different BMA models:

```
. bmastats lps bmareg bmareg_hyperg bmareg_inf, compact
Log predictive-score (LPS)
Number of observations = 200
```

LPS	Mean	Minimum	Maximum
bmareg	1.485701	1.040332	6.110174
bmareg_hyp~g	1.484734	1.004092	6.480865
bmareg_inf	1.489453	1.041369	6.272715

```
Notes: Results using analytical and frequency PMPs.
       Result bmareg_hyperg has the smallest mean LPS.
```

- The `hyperg` model is reported to have the smallest LPS value, but all considered models have similar LPS values.
- We can use LPS to compare in-sample and out-of-sample predictive performance of models; see **[BMA] bmastats lps**.
- We can also use prediction mean squared error and empirical coverage of CrIs to compare predictive performance of BMA models; see **[BMA]**.

Summary

- BMA may not be your final solution to every regression analysis, but, at the very least, it is definitely a beneficial exploratory tool!
- You can use BMA for prediction and for inference to account for model uncertainty.
- If you need to choose a model, you can use BMA's PMPs to guide your decision in a principled and unified way.
- You can use BMA to learn about interrelations between predictors across the model space.
- You can use BMA to explore the sensitivity of your results to various assumptions about the importance of different models and predictors.

References

Hoeting, J. A., D. Madigan, A. E. Raftery, and C. T. Volinsky. 1999. Bayesian model averaging: A tutorial. *Statistical Science* 14: 382–417.

Leamer, E. E. 1978. *Specification Searches: Ad Hoc Inference with Nonexperimental Data*. New York: Wiley.

Moral-Benito, E. 2015. Model averaging in economics: An overview. *Journal of Economic Surveys* 29: 46–75.

Steel, M. F. J. 2020. Model averaging and its use in economics. *American Economic Review* 58: 644–719.