Colombian Stata Conference

Introduction to Bayesian model averaging in Stata

Gustavo Sánchez

StataCorp LLC

September 12, 2023

Outline

1. Why model averaging
2. The Bayesian approach
   - Brief overview
   - The method
3. Bayesian model averaging in Stata
4. BMA commands
5. Example
   - BMA regression
   - Posterior inclusion probabilities
   - Posterior model probabilities
   - Variable inclusion dependence
   - Comparison with linear regression
   - Sensitivity analysis
   - Predictions

Why Model averaging

- In most cases, regression modeling relies on a theoretical framework that intends to derive the model that best describes the data generating process (DGP) for the outcome of interest.

- Researchers use a variety of statistical tools to find the model that is supposed to produce the best fit for the unknown DGP. For example
  - In terms of model specfication: AIC, BIC, Hannan-Quinn, among others.
  - In terms of predictive accuracy: MSE, MAE, MAPE, among others.

- However, those criteria may suggest different models. Then, what if we select the wrong model?

Why Model averaging

- Model averaging intends to address the model uncertainty and, therefore, reduce the risk of making inference and producing conclusions based on the wrong model.
- Let's consider, for example, the following model specifications (See, for example, Rizzo (2019) for an example on a model for life expectancy):

$$
\begin{aligned}
life\_exp &= \alpha_1 + \beta_{gdp} * gdp\_cap + \beta_{inf} * inflation + \epsilon_1 \\
life\_exp &= \alpha_2 + \beta_{pop} * pop\_growth + \beta_{urb} * urban + \epsilon_2 \\
life\_exp &= \alpha_3 + \beta_{co2} * co2 + \beta_{school} * schooling + \epsilon_3
\end{aligned}
$$

Where:

life_exp	:	Life expectancy at birth.
gdp_cap	:	Constant GDP per capita.
Inflation	:	Yearly inflation.
pop_growth	:	Population growth.
urban	:	Urban population.
co2	:	CO2 emissions.
schooling	:	Years of schooling.

Why Model averaging

- Instead of focusing the empirical analysis in just one model, this approach propose estimators that produce a weighted average of a number of potentially feasible models.
- Weigths are at then at the core of this approach, and both frequentists and Bayesians propose different ways for selecting those weights. Steel (2020) provides a broad description of the methods associated to both approaches.
- But frequentists and Bayesians approaches differ in a more fundamental theoretical modeling view of the model and the parameter, so let's just have a quick overview on those differences.

The Bayesian approach

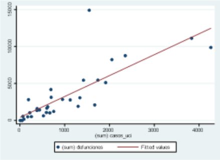
Frequentist

Data hypothetically repeatable

Theoretical Model

```
. list month defunciones casos_uci, abbreviate(12)
```

	month	defunciones	casos_uci
1.	2021m11	631	524
2.	2021m12	1912	1298
3.	2022m1	5453	1740
4.	2022m2	4183	691
5.	2022m3	1600	382
6.	2022m4	1422	436
7.	2022m5	1848	620
8.	2022m6	1663	601
9.	2022m7	3133	696
10.	2022m8	1046	219



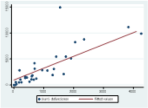
Bayesian

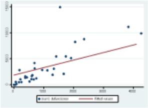
Data known

Theoretical Model

```
. list month defunciones casos_uci, abbreviate(12)
```

	month	defunciones	casos_uci
1.	2021m11	631	524
2.	2021m12	1912	1298
3.	2022m1	5453	1740
4.	2022m2	4183	691
5.	2022m3	1600	382
6.	2022m4	1422	436
7.	2022m5	1848	620
8.	2022m6	1663	601
9.	2022m7	3133	696
10.	2022m8	1046	219





Bayesian Analysis vs. Frequentist Analysis

Frequentist Analysis

- Estimates unknown fixed parameters.
- The data come from a random sample (hypothetical repeatable).
- Uses data to estimate unknown fixed parameters.
- *p*-values are conditional probability statements that assume Ho to be true.

"Conclusions are based on the distribution of statistics derived from random samples, assuming unknown but fixed parameters."

Bayesian Analysis

- Probability distributions for unknown random parameters.
- The data are assumed to be fixed.
- Combines data with prior beliefs to get updated probability distributions for the parameters.
- It allows formulating probabilistic statements for the hypothesis of interest.

"Bayesian analysis answers questions based on the distribution of parameters conditional on the observed sample."

The method

- Inverse law of probability (Bayes' Theorem):

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

Where:

$f(y;\theta)$: probability density function for y given θ.

$\pi(\theta)$: prior distribution for θ

- The marginal distribution of y, f(y), does not depend on θ; then we can write the fundamental equation for Bayesian analysis:

$$p(\theta|y) \propto L(\theta;y)\,\pi(\theta)$$

Where:

$L(\theta;y)$: likelihood function of the parameters given the data.

The method

- Some prior-likelihood combinations have closed form solution.
- What about the cases with non-closed solutions, or more complex distributions?
  - Integration is performed via simulation.
  - We need to use intensive computational simulation tools to find the posterior distribution in most cases.
  - Markov chain Monte Carlo (MCMC) methods are the current standard in most software. Stata implements two alternatives:
    - Metropolis–Hastings (MH) algorithm
    - Gibbs sampling

The method

- Links for Bayesian analysis and MCMC on our YouTube channel:
  - Introduction to Bayesian statistics, part 1: The basic concepts

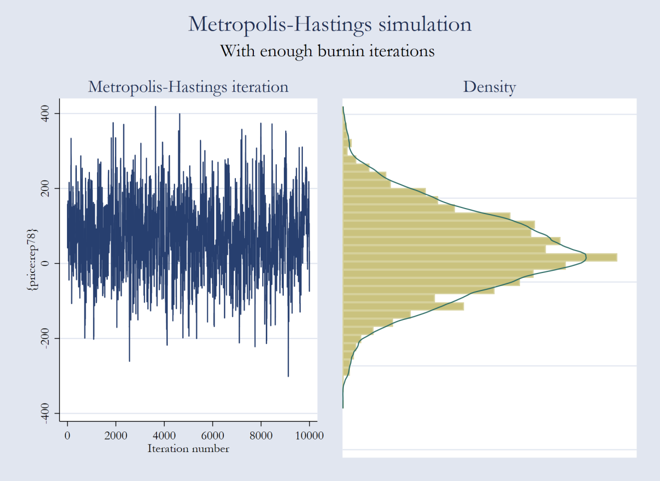
    https://www.youtube.com/watch?v=0F0QoMCSKJ4&feature=youtu.be

  - Introduction to Bayesian statistics, part 2: MCMC and the Metropolis–Hastings algorithm.

    https://www.youtube.com/watch?v=OTO1DygELpY&feature=youtu.be

The method

- Metropolis–Hastings simulation
  - The trace plot illustrates the sequence of accepted proposal states for a simulation with enough burnin iterations.



Bayesian model averaging in Stata (BMA)

- The current Stata implementation is focussed on linear regression:

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

Where:

$\mathbf{y} = (y1, y2, ..., y_n)'$	:	(nx1) vector of outcome values.
$\mathbf{1}_n$	:	vector of ones.
$\mathbf{X}_j$	:	nxp_j design matrix.
β_j	:	$(p_j x1)$ vector of coef. for model j
$\boldsymbol{\epsilon}=(\epsilon_1, \epsilon_2, ..., \epsilon_3)$	:	(nx1) vector of error terms.

- In addition to the standard posterior probability distributions for the regression coefficients, two probabilities are fundamental for the inference using the Bayesian approach for model averaging:
  - The posterior model probabilities (PMPs)
  - The posterior inclusion probabilities (PIPs)

BMA prior probability distributions

- Let's recall our linear model specification:

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

- Priors for a BMA linear regression:

$$\begin{aligned}
M_j &\sim P(M_j) \\
\beta_j | \alpha, \sigma, M_j &\sim N_{p_j}(0, g\sigma^2 (X_j' X_j)^{-1}) \\
\alpha | \sigma, M_j &\sim 1 \\
\sigma | M_j &\sim \sigma^{-1}
\end{aligned}$$

- Notice that in addition to the priors for the parameters $(\beta_j, \alpha, \sigma)$, BMA considers the models to be random, so a discrete model prior ($P(M_j)$) is specified over the models space $\mathbf{M}_F = M_1, M_2, ..., M_{2^p}$.
- Prior for g: fixed or random hyperprior p(g)

BMA posterior model and inclusion probabilities

- Posterior model probabilities conditional on the observed data (using Bayes theorem):

$$PMP = P\left(M_j|\boldsymbol{y}\right) \quad = \quad \frac{f\left(\boldsymbol{y}|M_j\right) P\left(M_j\right)}{p\left(\boldsymbol{y}\right)}$$

Where: $f\left(\boldsymbol{y}|M_j\right)$: Likelihood of **y** under model M_j.
$P(\boldsymbol{y})$: marginal probability of y over the model space $\mathbf{M}_F$

- We can then define the posterior inclusion probability (PIP) as:

$$PIP = \sum_{j \in J_F} I(\boldsymbol{X}_k \in M_j) P(M_j|\boldsymbol{y})$$

Where I(.) is the indicator function.

Bayesian Model Averaging

Outline
Model averaging
The Bayesian approach
BMA
bma commands
Example
BMA regression
PIP
PMP
Jointness
Compare OLS
CRI
Sensitivity
Predictions
References

BMA posterior probability distributions

- Posterior distribution of β over all models:

$$g(\beta|\mathbf{y}) = \sum_{j \in J_F} g(\beta|\mathbf{y}, M_j) P(M_j|\mathbf{y})$$

Where: $g(\beta|\mathbf{y}, M_j)$ is the posterior distribution of β for a Bayesian linear regression model M_j

- BMA coefficient estimates for the linear model:

$$\hat{\beta}_{BMA} = E[\beta|y] = \sum_{j=1}^{2^p} P(M_j|y)\hat{\beta}_j$$

Where $\hat{\beta}_j'$ is the vector of posterior mean estimates of regression coefficients based on model M_j

Stata's BMA suite consists of the following commands

Command	*Description*
Setup	
`splitsample`	Split samples for training,validation and prediction
`vl`	Manage large variable lists
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 summaries
`bmastats models`	Posterior model and variable-inclusion summaries
`bmastatspip`	Posterior inclusion probabilities for predictors
`bmastats jointness`	Jointness measures for predictors
`bmastats lps`	Log predictive-score
Predictions	
`bmapredict`	BMA predictions

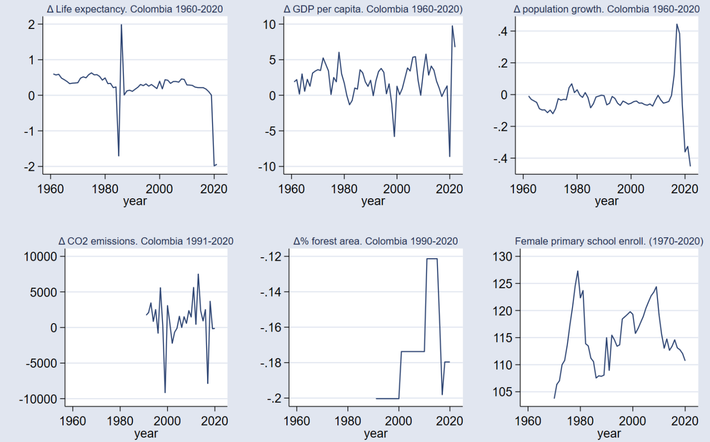
Example: Life expectancy model for Colombia

- Let's work with a model for life expectancy including the 10 explanatory variables described below:

```
. describe life_exp fertility gdp_cap_gr inflation urban pop_growth ///
>          primary_enrol_f forest_area co2_tot mrate_inf remittances

Variable      Storage   Display    Value
    name         type    format    label      Variable label
------------------------------------------------------------------------------
life_exp        double   %10.0g               Life expectancy at birth (years)
fertility       double   %10.0g               Fertility rate (births per woman)
gdp_cap_gr      double   %10.0g               GDP per capita growth (annual %)
inflation       double   %10.0g               Inflation (annual %)
urban           double   %10.0g               Urban population (% of total)
pop_growth      double   %10.0g               Population growth (annual %)
primary_enrol_f double   %10.0g               Female primary school enrol.(%)
forest_area     double   %10.0g               Forest area (% of land area)
co2_tot         double   %10.0g               CO2 emissions (kt)
mrate_inf       double   %10.0g               Infant Mortal.rate (per 1,000
                                                live births)
remittances     double   %10.0g               Personal remittances (% of GDP)
```

- Annual change in a variable is specified with d as a prefix.

Plot some of the series



Source: The World Bank
https://data.worldbank.org/country/colombia?view=chart

BMA regression

```
. bmaregress dlife_exp dfertility gdp_cap_gr inflation durban dpop_growth ///
>          primary_enrol_f  dforest_area dco2_tot dmrate_inf remittances, ///
>          saving($simul_dir\bma_enumerate,replace)
Enumerating models ...
Computing model probabilities ...
Bayesian model averaging                           No. of obs         =      28
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  =      65
   Cons.: Noninformative                           Mean model size    =   5.487
   Coef.: Zellner´s g
       g: Benchmark, g = 100                       Shrinkage, g/(1+g) = 0.9901
  sigma2: Noninformative                           Mean sigma2        =   0.039

------------------------------------------------------------------------------
      dlife_exp |      Mean   Std. dev.                          Group      PIP
----------------+-------------------------------------------------------------
    dpop_growth |  1.567063   .4491715                               5   .98144
primary_enrol_f |  .0424761   .0168248                               6   .94227
     gdp_cap_gr |  .0520123    .026503                               2   .90293
         durban |  4.922548   3.313974                               4    .7885
   dforest_area |  4.603837   3.553353                               7   .73745
     dmrate_inf | -.2613446    .579727                               9   .29843
       dco2_tot | -4.77e-06   .0000115                               8   .25203
     dfertility |  1.397038   4.145878                               1   .23616
      inflation | -.0013962   .0088316                               3   .20076
    remittances |  .0002999   .0371939                              10   .14746
----------------+-------------------------------------------------------------
Always          |
          _cons | -6.026107   1.915335                               0        1
------------------------------------------------------------------------------
Note: Coefficient posterior means and std. dev. estimated from 1,024 models.
Note: Default priors are used for models and parameter g.
file C:\Users\gas\Documents\conferences\Colombia\simul\bma_enumerate.dta saved.
. estimates store bmareg_enum
```

Regression output

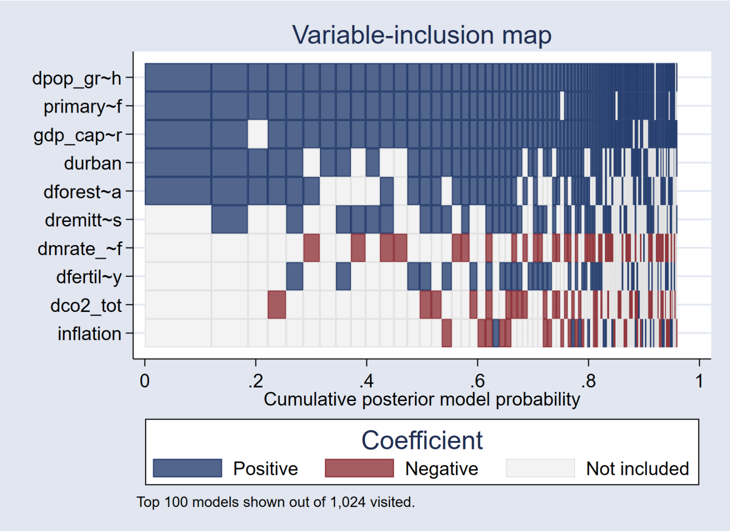
- Estimation default
  - Model enumeration (<12 predictors) ($2^{10} = 1024$ models)
  - Priors: Beta-binomial(1,1) for models (binomial model prior with an inclusion probability (IP) and a beta prior on the IP) and fixed $g = 100$
- Results
  - Little shrinkage: $100/(1 + 100) = .9901$
  - Mean model size: 5.487
  - Top three predictors: dpop_growth, primary_enrol_f, gdp_cap_gr. (PIPs>.9)
  - Other predictors seem relevant too (with PIPs>.30)
  - BMA estimates based on $2^{10} = 1024$ models. 65 of those models contribute to .9 of the cumulative PMP.
  - Estimation stored for some of the postimation analysis

Predictors with highest probability of inclusion `bmastats pip`

```
. bmastats pip, cutoff(.5)

Posterior inclusion probability (PIP)

No. of obs          =     28
No. of predictors   =     10
            Groups  =     10
            Always  =      0
          Reported  =      5
No. of models       =  1,024
Mean model size     =  5.487

---------------------------------------
                |       PIP       Group
----------------+----------------------
    dpop_growth |    .98144           5
primary_enrol_f |    .94227           6
     gdp_cap_gr |    .90293           2
         durban |     .7885           4
   dforest_area |    .73745           7
----------------+----------------------
Always          |
          _cons |         1           0
---------------------------------------

Note: Using analytical PMPs.
Note: 5 predictors with PIP less than .5
      not shown.
```

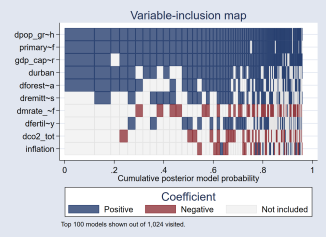
Variable inclusion map `bmagraph varmap`

```
. bmagraph varmap, top(100) legend(rows(1))
Computing model probabilities ...
```

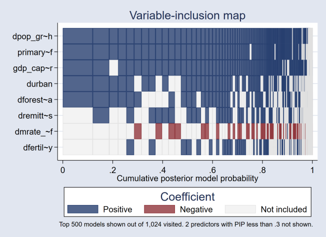


Variable inclusion map `bmagraph varmap`

```
. bmagraph varmap, top(100) legend(rows(1))
Computing model probabilities ...
```




```
. bmagraph varmap, top(500) pipcutoff(.3) legend(rows(1))
Computing model probabilities ...
```



High posterior model probabilites `bmastats pmp`

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

	Analytical PMP	Model size
Rank		
1	.1993	5
2	.05914	4
3	.05505	6
4	.04956	5
5	.04777	4

Variable-inclusion summary

	Rank 1	Rank 2	Rank 3	Rank 4	Rank 5
gdp_cap_gr	x		x	x	x
durban	x	x	x		x
dpop_growth	x	x	x	x	x
primary_enrol_f	x	x	x	x	x
dforest_area	x	x	x	x	
dco2_tot			x		
dmrate_inf				x	

```
Legend:
  x - estimated
```

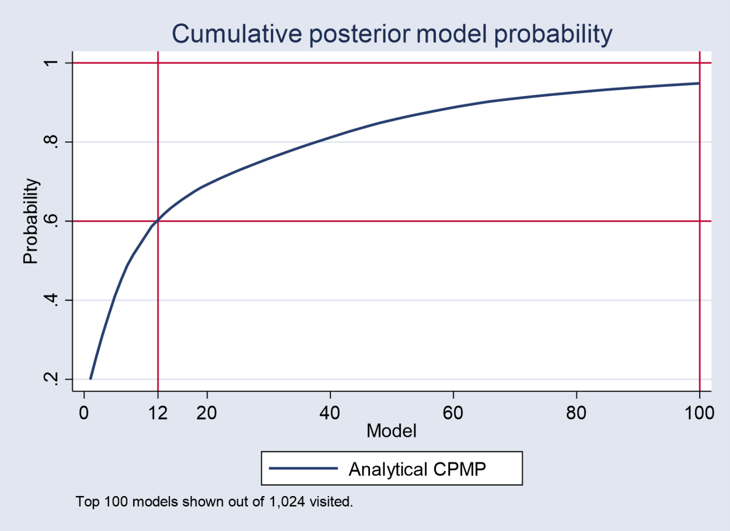
Cumulative posterior model probability `bmastats pmp,cumulative`

```
. bmastats models, cumulative(.60) novartable
Computing model probabilities ...
Model summary          Number of models:
                                Visited = 1,024
                               Reported =    12
```

Rank	Analytical CPMP	Model size
1	.1993	5
2	.2585	4
3	.3135	6
4	.3631	5
5	.4108	4
6	.451	4
7	.4879	6
8	.5158	6
9	.5397	7
10	.5635	6
11	.5873	6
12	.6031	6

Cumulative posterior model probability `bmagraph pmp,cumulative`

```
. bmagraph pmp, cumulative xline(12 100) yline(.60 1) xlabel(12, add)
note: frequency estimates not available with model enumeration; option
      nofreqline implied.
```



Model size distribution `bmastats msize`

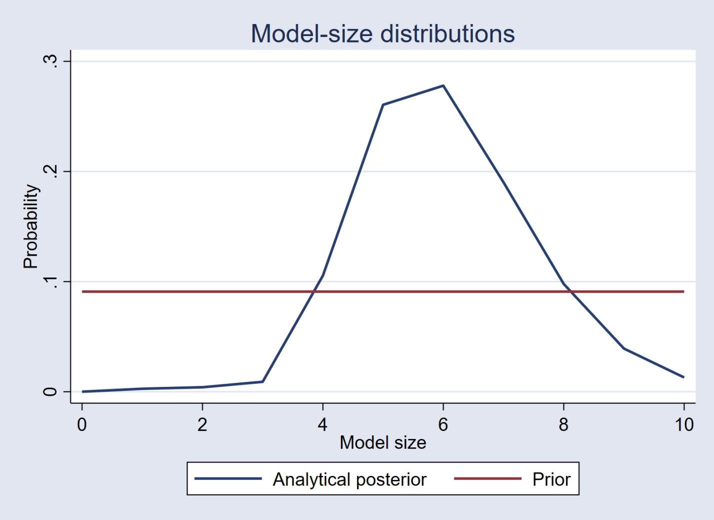
```
. bmastats msize
Model-size summary
Number of models = 1,024
Model size:
  Minimum =  0
  Maximum = 10
```

	Mean	Median
Prior		
Analytical	5.0000	5
Posterior		
Analytical	5.4874	5

Note: Frequency summaries not available.

Model size distribution `bmagraph msize`

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

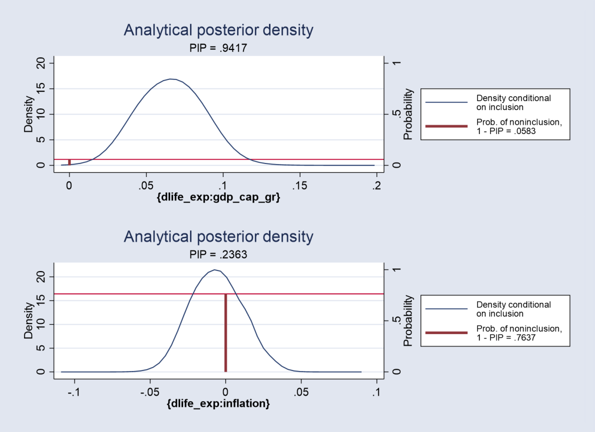


Bayesian Model Averaging

Outline
Model averaging
The Bayesian approach
BMA
bma commands
Example
BMA regression
PIP
PMP
Jointness
Compare OLS
CRI
Sensitivity
Predictions
References

Posterior density for betas `bmagraph coefdensity`

```
. bmagraph coefdensity {gdp_cap_gr},name(coefd_gdpgr,replace) ///
>         legend(size(small) rows(2) pos(3))
. bmagraph coefdensity {inflation},name(coefd_inflation,replace) ///
>         legend(size(small) rows(2) pos(3))
. graph combine coefd_gdpgr coefd_inflation,rows(2)
```



Variable inclusion dependence `bmastats jointness`

- Explore inclusion pattern for predictors using bivariate jointness measures from the joint posterior distribution of inclusion of predictors over the model space.
  - Doppelhofer –Weeks measure (DW)
  - Ley –Steel type 1 (LS1)
  - Ley –Steel type 2 (LS2)
  - Yule's Q
- Look at the threshold values for each measure in the manual entry for `bmastats jointness` (or click on the blue link for the thresholds in the output). Treshold values for DW:

DW	Interpretation
$(-\infty, -2)$	Strong disjointness
(-2,-1)	Significant disjointness
(-1,1)	Independent inclusion
(1,2)	Significant jointness
$(2, \infty)$	Strong jointness

Variable inclusion dependence `bmastats jointness`

- Explore inclusion pattern for predictors using bivariate jointness measures from the joint posterior distribution of inclusion of predictors over the model space.
  - Doppelhofer –Weeks measure (DW)
  - Ley –Steel type 1 (LS1)
  - Ley –Steel type 2 (LS2)
  - Yule's Q
- Look at the threshold values for each measure in the manual entry for `bmastats jointness` (or click on the blue link for the thresholds in the output). Treshold values for DW:

DW	Interpretation
$(-\infty, -2)$	Strong disjointness
(-2,-1)	Significant disjointness
(-1,1)	Independent inclusion
(1,2)	Significant jointness
$(2, \infty)$	Strong jointness

Variable inclusion dependence `bmastats jointness`

```
. bmastats jointness gdp_cap_gr durban dco2_tot,dw
Doppelhofer-Weeks jointness
```

	gdp_cap_gr	durban	dco2_tot
gdp_cap_gr	.	-2.283519	1.279352
durban	-2.283519	.	.1044019
dco2_tot	1.279352	.1044019	.

```
Notes: Using analytical PMPs. See thresholds.
```

- gdp_cap_gr and durban are strong substitutes: when one of them is included in the model, the other does not add significant explanatory power for change in life expectancy.
- How about dco2_tot and durban, or dco2_tot and gdp_cap_gr?

Compare `bmaregress` vs. `regress`

- Let's use the suite of `collect` commands to generate a table with the results from OLS and BMA:

```
 . collect clear
. collect create bma_compare
. collect _r_b:regress dlife_exp dfertility gdp_cap_gr inflation ///
>                      durban dpop_growth primary_enrol_f       ///
>                      dforest_area dco2_tot dmrate_inf remittances
.
. collect _r_b=e(b_bma):                                          ///
>           bmaregress dlife_exp dfertility gdp_cap_gr inflation  ///
>                     durban dpop_growth primary_enrol_f          ///
>                     dforest_area dco2_tot dmrate_inf remittances
. collect dims
. collect label levels program_class eclass "ols" nclass "bma_reg"
. collect style cell, nformat(%5.2f)
. collect style header result, level(hide)
. collect style column, extraspace(2)
. collect style row stack,spacer
```

Compare bmaregress vs. regress

```
. collect layout (colname#result) (program_class)
Collection: bma_compare
      Rows: colname#result
   Columns: program_class
   Table 1: 21 x 2
```

	ols	bma_reg
dfertility	9.03	1.40
GDP per capita growth (annual %)	0.06	0.05
Inflation (annual %)	-0.01	-0.00
durban	6.85	4.92
dpop_growth	1.51	1.57
Female primary school enrol.(%)	0.06	0.04
dforest_area	4.85	4.60
dco2_tot	-0.00	-0.00
dmrate_inf	-0.56	-0.26
Personal remittances (% of GDP)	0.05	0.00
Intercept	-8.08	-6.03

Compare `bmaregress` vs. `regress`

- Some beta estimates are pretty close, particularly the ones that were present in most models with `bmaregress`.
- Do the reported betas represent point estimates or summary statistics from a posterior distribution?
- Does any of the two sets of estimates correspond to the true model?
- How do you determine whether the included variables are relevant to explain the outcome variable?

How about credible intervals for the BMA estimation

- The regression output with fixed g reports analytical means and standard deviations.
- However, analytical formulas for the credible intervals are much more involved, and they are not currently implemented.
- The credible interval limits can be estimated from a sample of the posterior distributions of the coefficients. The sample is generated with `bmacoefsample`
- Then `bayestats summary` can be used to get the credible interval limits.

Compare bmaregress vs. regress

```
. estimates restore bmareg_enum
(results bmareg_enum are active now)
. bmacoefsample,rseed(123)
Simulation (10000): ....5000....10000 done
. bayesstats summary
Posterior summary statistics                      MCMC sample size =     10,000
```

	Mean	Std. dev.	MCSE	Median	Equal-tailed [95% cred.	interval]
dlife_exp						
dfertility	1.370008	4.109684	.040139	0	-1.96701	13.97248
gdp_cap_gr	.0521469	.0264181	.000264	.0539244	0	.1005381
inflation	-.0015345	.0088047	.000088	0	-.0297957	.0141918
durban	4.955787	3.295118	.032951	5.617096	0	10.87977
dpop_growth	1.568373	.4528216	.004528	1.585967	.5864631	2.410423
primary_en~f	.0421991	.0172291	.000172	.0433484	0	.0738838
dforest_area	4.584972	3.569635	.035081	5.028764	0	11.30876
dco2_tot	-4.88e-06	.0000115	1.1e-07	0	-.0000384	1.17e-06
dmrate_inf	-.2557381	.5723385	.005723	0	-1.78755	.3027644
remittances	.0000893	.0365022	.00036	0	-.0804401	.085957
_cons	-6.006948	1.948707	.019736	-6.065546	-9.858821	-1.161631
sigma2	.0387002	.0160531	.000161	.0353625	.0197374	.0786804
g	100	0	0	100	100	100

Sensitivity analysis: Random g-prior (Header)

- Let's specify a robust random prior for the g parameter:

```
 . bmaregress dlife_exp dfertility gdp_cap_gr inflation durban dpop_growth ///
>     primary_enrol_f dforest_area dco2_tot dmrate_inf remittances,       ///
>     gprior(robust) rseed(123) saving(bma_robust, replace) notable
Burn-in ...
Simulation ...
Computing model probabilities ...
Bayesian model averaging                          No. of obs         =     28
Linear regression                                 No. of predictors =     10
MC3 and adaptive MH sampling                                 Groups =     10
                                                             Always =      0
                                                  No. of models      =    284
                                                     For CPMP >= .9 =     88
Priors:                                           Mean model size    =  6.947
  Models: Beta-binomial(1, 1)                     Burn-in            =  2,500
   Cons.: Noninformative                          MCMC sample size   = 10,000
   Coef.: Zellner´s g                             Acceptance rate    = 0.5888
       g: Robust
  sigma2: Noninformative                          Mean sigma2        =  0.049
Sampling correlation = 0.9499
file bma_robust.dta saved.
```

- The sampling correlation can be checked as a indicator for convergence. It measures the correlation between the analytical posterior model probabilities (PMPs) and their MCMC estimates based on sampling frequencies.

Sensitivity analysis: Random g-prior (Estimation)

```
. bmaregress dlife_exp dfertility gdp_cap_gr inflation durban dpop_growth ///
>     primary_enrol_f dforest_area dco2_tot dmrate_inf remittances,       ///
>     gprior(robust) rseed(123) saving(bma_robust, replace) noheader
Burn-in ...
Simulation ...
Computing model probabilities ...
```

dlife_exp	Mean	Std. dev.	Group	PIP
dpop_growth	1.421714	.4270151	5	.9954
gdp_cap_gr	.0515116	.024519	2	.9399
primary_enrol_f	.0419044	.0194081	6	.9394
durban	5.015494	3.75865	4	.8211
dforest_area	4.224761	3.395383	7	.7988
dfertility	3.220716	6.191173	1	.5171
dco2_tot	-8.97e-06	.0000151	8	.5118
inflation	-.0031942	.0136722	3	.4964
dmrate_inf	-.3022117	.7107796	9	.4802
remittances	.0075522	.0734045	10	.4471
Always				
_cons	-5.999691	2.27732	0	1

Note: Coefficient posterior means and std. dev. estimated from 284 models.
Note: Default prior is used for models.

	Mean	Std. dev.	MCSE	Median	Equal-tailed [95% cred.	interval]
g	17.04498	16.28341	.416768	12.48753	3.546497	58.19242
Shrinkage	.9150258	.0521987	.001453	.9258575	.7800504	.9831059

```
file bma_robust.dta saved.
. estimates store bma_robust
```

Sensitivity analysis: Model prior (Header)

- Let's specify a binomial prior for the inclusion probabilities for some of the coefficients:

```
 . bmaregress dlife_exp dfertility gdp_cap_gr inflation durban      ///
>     dpop_growth primary_enrol_f dforest_area dco2_tot dmrate_inf ///
>     remittances, mprior(betabinomial 2) gprior(hyperg 3)        ///
>     rseed(123) saving(bma_mprior, replace) notable
Burn-in ...
Simulation ...
Computing model probabilities ...
Bayesian model averaging                         No. of obs         =     28
Linear regression                                No. of predictors =     10
MC3 and adaptive MH sampling                                Groups =     10
                                                           Always =      0
                                                 No. of models      =    369
                                                    For CPMP >= .9 =    133
Priors:                                          Mean model size    =  5.476
  Models: Beta-binomial, mean = 2                Burn-in            =  2,500
   Cons.: Noninformative                         MCMC sample size   = 10,000
   Coef.: Zellner´s g                            Acceptance rate    = 0.6445
       g: Hyper-g(3)
  sigma2: Noninformative                         Mean sigma2        =  0.054
Sampling correlation = 0.8052
file bma_mprior.dta saved.
. estimates store bma_hyperg
```

Sensitivity analysis: Model prior (Estimation)

```
. bmaregress dlife_exp dfertility gdp_cap_gr inflation durban      ///
>     dpop_growth primary_enrol_f dforest_area dco2_tot dmrate_inf ///
>     remittances, mprior(betabinomial) gprior(hyperg 3)           ///
>     rseed(123) saving(bma_mprior, replace) noheader
Burn-in ...
Simulation ...
Computing model probabilities ...
```

dlife_exp	Mean	Std. dev.	Group	PIP
dpop_growth	1.384748	.4673228	5	.979
gdp_cap_gr	.0520796	.0254548	2	.9359
primary_enrol_f	.0406559	.0199708	6	.9239
durban	4.95577	3.774566	4	.8265
dforest_area	4.009261	3.451226	7	.7648
dco2_tot	-9.00e-06	.0000155	8	.5171
dfertility	3.091163	6.19397	1	.5131
dmrate_inf	-.2851908	.7108143	9	.4942
inflation	-.0034508	.0133842	3	.4696
remittances	.0046889	.0734949	10	.4194
Always				
_cons	-5.85823	2.349691	0	1

Note: Coefficient posterior means and std. dev. estimated from 351 models.

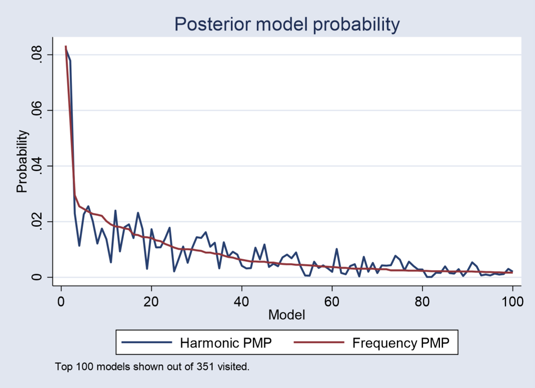
	Mean	Std. dev.	MCSE	Median	Equal-tailed [95% cred.	interval]
g	17.21842	18.33173	.466986	12.43851	2.772172	58.22623
Shrinkage	.9089842	.0649853	.001823	.925587	.7349008	.9831154

```
file bma_mprior.dta saved.
. estimates store bma_mprior
```

Let's check for convergence for the MCMC simulation

- Another tool to check convergence corresponds to the plot for the posterior model probability (pmp)

```
. estimates restore bma_mprior
(results bma_mprior are active now)
. bmagraph pmp
```



Sensitivity analysis: Comparison with `bmastats lps`

- As stated in the manual, LPS corresponds to the negative of the log of the posterior predictive density evaluated at an observation.
- This measure can be used to evaluate the out of sample predictive performance, and also to evaluate model fit when making in sample comparisons for different models.
- The model with the smallest LPS should be selected. In the result below, the default model (bmareg_enum) would be the best alternative.

```
. bmastats lps bmareg_enum bma_robust bma_mprior,compact
Log predictive-score (LPS)
Number of observations = 63

-----------------------------------------------------------
          LPS |       Mean     Minimum     Maximum
-------------+---------------------------------------------
  bmareg_enum |   2.117462   -.6741457    13.60309
   bma_robust |   2.658486   -.5509572    12.75069
   bma_mprior |   2.416428   -.5398606    11.81978
-----------------------------------------------------------

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

BMA predictions

- Let's finish our exercise obtaining the predictions for the mean of the outcome variable.
  - Analytic mean prediction.

```
 . estimates restore bmareg_enum
(results bmareg_enum are active now)
. bmapredict pmean,mean
note: computing analytical posterior predictive means.
```

  - Use `bmacoefsample` to produce the simulated mcmc data.

```
 . bmacoefsample,saving(bma_coef,replace)
Simulation (10000): ....5000....10000 done
file bma_coef.dta saved.
```

BMA predictions

- We can now obtain the predicted mean and its credible intervals:

```
 . bmapredict pmean_simul, mean mcmcsample rseed(123)
note: computing posterior predictive means using simulation.
Computing predictions ...
. bmapredict cri_l cri_u, cri rseed(123)
note: computing credible intervals using simulation.
Computing predictions ...
. summarize dlife_exp pmean* cri*
```

Variable	Obs	Mean	Std. dev.	Min	Max
dlife_exp	61	.257377	.5532763	-1.983	1.984
pmean	28	.1928571	.3870608	-1.519764	.541859
pmean_simul	28	.192392	.3868875	-1.520528	.5404659
cri_l	28	-.2613104	.4079078	-2.041368	.0975296
cri_u	28	.6439852	.3539703	-.9345627	.9883481

Summary

- Model averaging intends to account for model uncertainty.
- BMA provides the tools to perform this kind of analysis based on posterior probability distributions.
- BMA can be helpful in determining the most important predictors for your model.
- Even if you plan to work with just one model, BMA can be used as an exploratory tool. For example, you can look at the interrelations across predictors.
- BMA can be used for inference and prediction.
- Just like with any other Bayesian estimation, sensitivity analysis should be performed.

References

- Rizzo, Taylor. (2019), "A panel data study of the determinants of life expectancy in low income countries". Honor Thesis. Bryant University.
- Mark F. J. Steel, 2020. "Model Averaging and Its Use in Economics," Journal of Economic Literature, American Economic Association, vol. 58(3), pages 644-719, September