Mexican Stata Conference

Introduction to Bayesian model averaging in Stata

Gustavo Sánchez

StataCorp LLC

October 26, 2023
Hermosillo, México

STATA 18

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):

$$
\begin{aligned}
\textit{life\_exp} &= \alpha_1 + \beta_{\textit{food}} * \textit{food\_prod} + \beta_{\textit{elect}} * \textit{elect\_acc} + \epsilon_1 \\
\textit{life\_exp} &= \alpha_2 + \beta_{\textit{pop}} * \textit{pop\_growth} + \beta_{\textit{urb}} * \textit{urban} + \epsilon_2 \\
\textit{life\_exp} &= \alpha_3 + \beta_{\textit{co2}} * \textit{co2} + \beta_{\textit{school}} * \textit{schooling} + \epsilon_3
\end{aligned}
$$

Where:

life_exp	:	Life expectancy at birth.
food_prod	:	Food production index (2014-16 = 100).
elect_acc	:	Electric access (% of population).
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

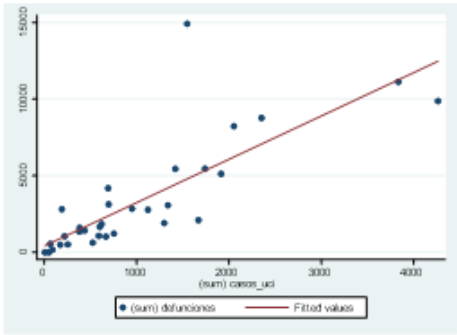
Frequentist

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

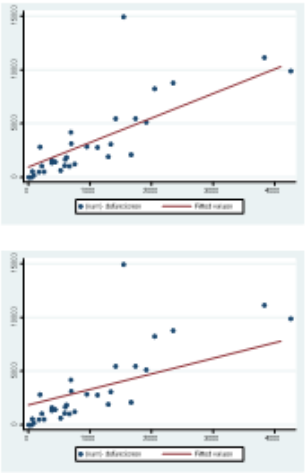
Data known

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

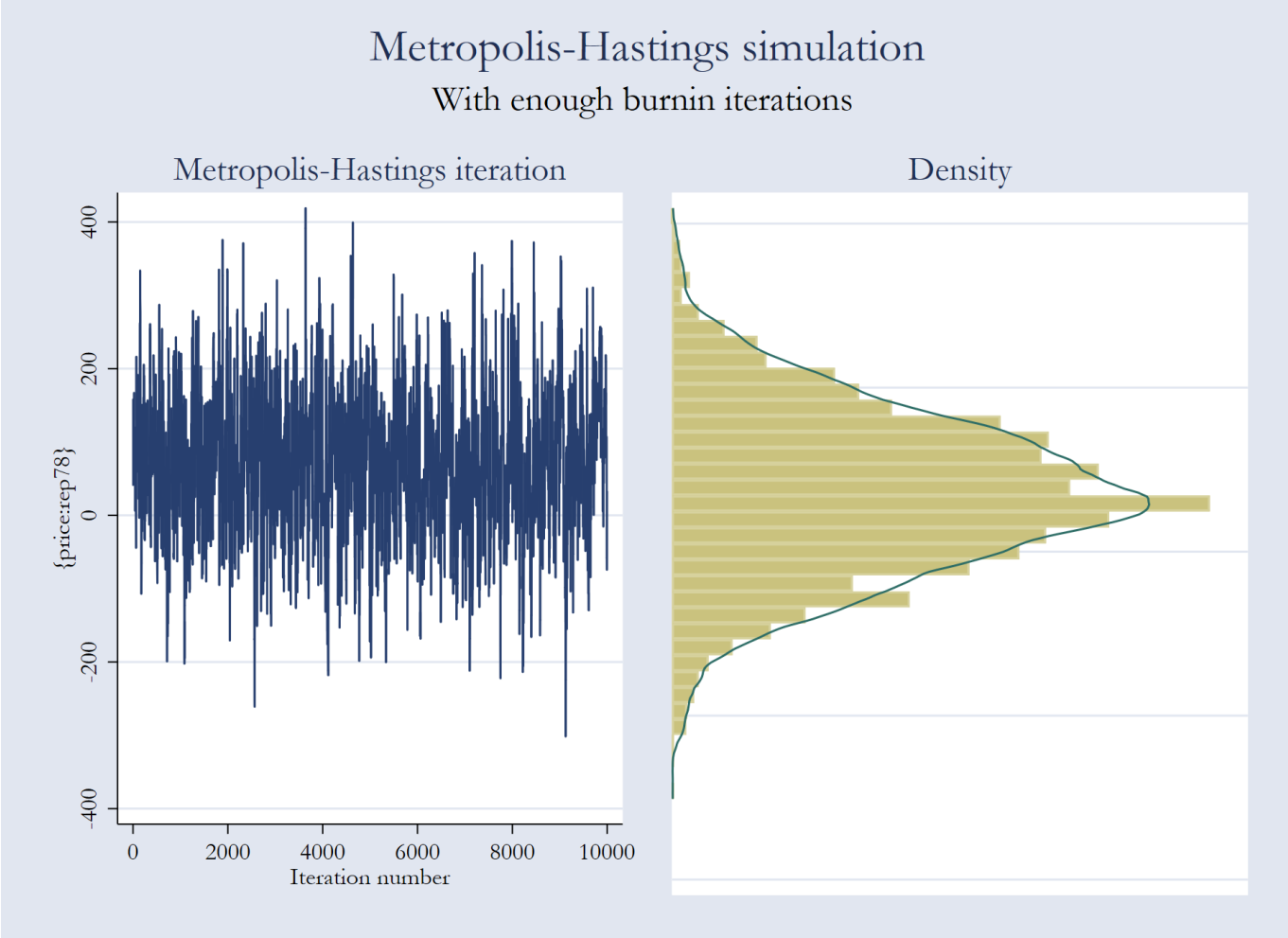
    https://www.youtube.com/watch?v=0F0QoMCSKJ4&feature=youtu.be

  - 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:

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

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	:	$(p_j x 1)$ vector of coef. for model j
ϵ=$(\epsilon_1, \epsilon_2, ..., \epsilon_3)$	:	(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 with fixed g:

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

- 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 $\mathbf{y}$ 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.

BMA posterior probability distributions

- Posterior distribution of β over all models:

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

Where: $g(\beta|\boldsymbol{y}, M_j)$ 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

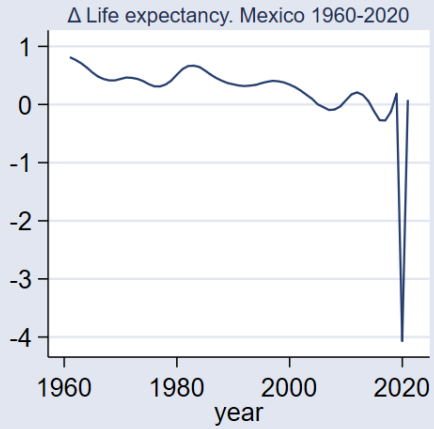
- Let's work with a model for life expectancy including the 10 explanatory variables described below:

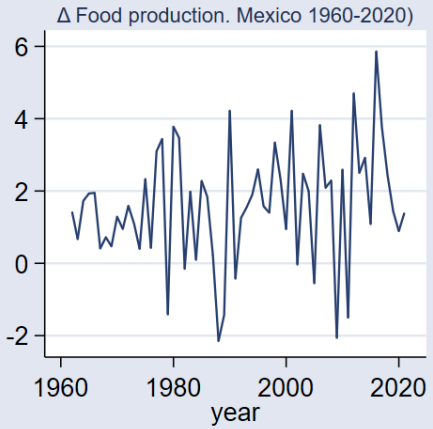
```
. describe life_exp food_prod elect_acc co2_transp forest_area ///
>     urban pop_growth fertility enrol_secnd enrol_prim physicians
```

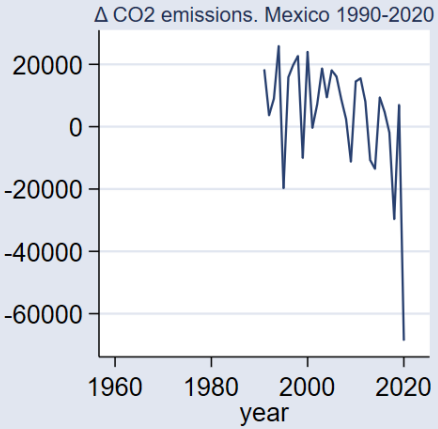
Variable name	Storage type	Display format	Value label	Variable label
life_exp	double	%10.0g		Life expectancy at birth (years)
food_prod	double	%10.0g		Food prod. index (2014-16 = 100)
elect_acc	double	%10.0g		Access to electricity (% of population)
co2_transp	double	%10.0g		CO2 emiss.transp. (% of tot fuel)
forest_area	double	%10.0g		Forest area (% of land area)
urban	double	%10.0g		Urban population (% of total)
pop_growth	double	%10.0g		Population growth (annual %)
fertility	double	%10.0g		Fertility rate (births per woman)
enrol_secnd	double	%10.0g		School enrol. secondary (% gross)
enrol_prim	double	%10.0g		School enroll. primary (% gross)
physicians	double	%10.0g		Physicians (per 1,000 people)

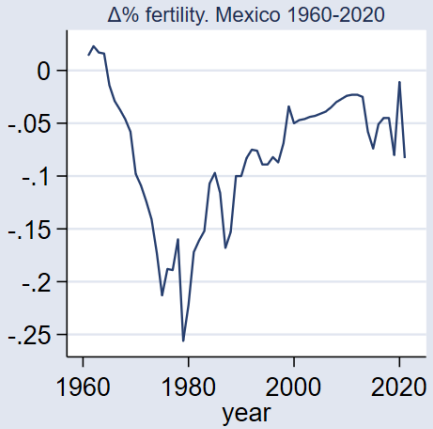
- Annual change in a variable is specified with d as a prefix.

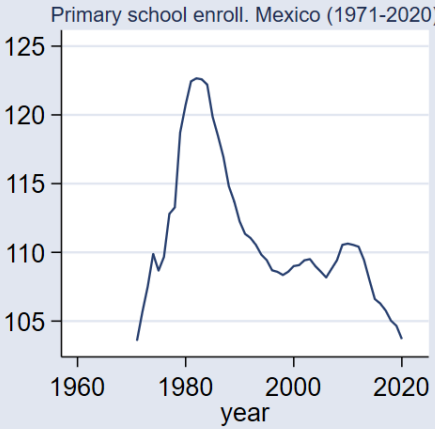
Plot some of the series











Source: The World Bank
https://data.worldbank.org/country/mexico

BMA regression

```
. bmaregress dlife_exp dfood_prod delect_acc co2_transp forest_area     ///
>       durban dpop_growth dfertility denrol_secnd enrol_prim physicians, ///
>       saving($simul_dir\bma_enumerate_mx,replace)
Enumerating models ...
Computing model probabilities ...
Bayesian model averaging                        No. of obs          =     22
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  =     66
   Cons.: Noninformative                        Mean model size     =  6.143
   Coef.: Zellner´s g
       g: Benchmark, g = 100                    Shrinkage, g/(1+g)  = 0.9901
  sigma2: Noninformative                        Mean sigma2         =  0.003
```

dlife_exp	Mean	Std. dev.	Group	PIP
forest_area	.2237174	.0737425	4	.95386
denrol_secnd	.0509517	.0206289	8	.95088
enrol_prim	.0616976	.0307517	9	.87058
co2_transp	-.0475698	.0280115	3	.83225
physicians	.3157756	.2381506	10	.73753
dfood_prod	.0084947	.0095266	1	.55099
durban	-.4312485	.666253	5	.42783
delect_acc	.009009	.0149244	2	.37956
dfertility	-.2195602	.9680006	7	.23576
dpop_growth	.0173422	.2024709	6	.20332
Always				
_cons	-13.43001	3.930782	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_mx.dta sav
> ed.
. 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: 6.143
  - Top four predictors: forest_area, denrol_secnd, enrol_prim, co2_transp (PIPs>.8)
  - Other predictors seem relevant too (with PIPs>.30)
  - BMA estimates based on $2^{10} = 1024$ models. 66 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`

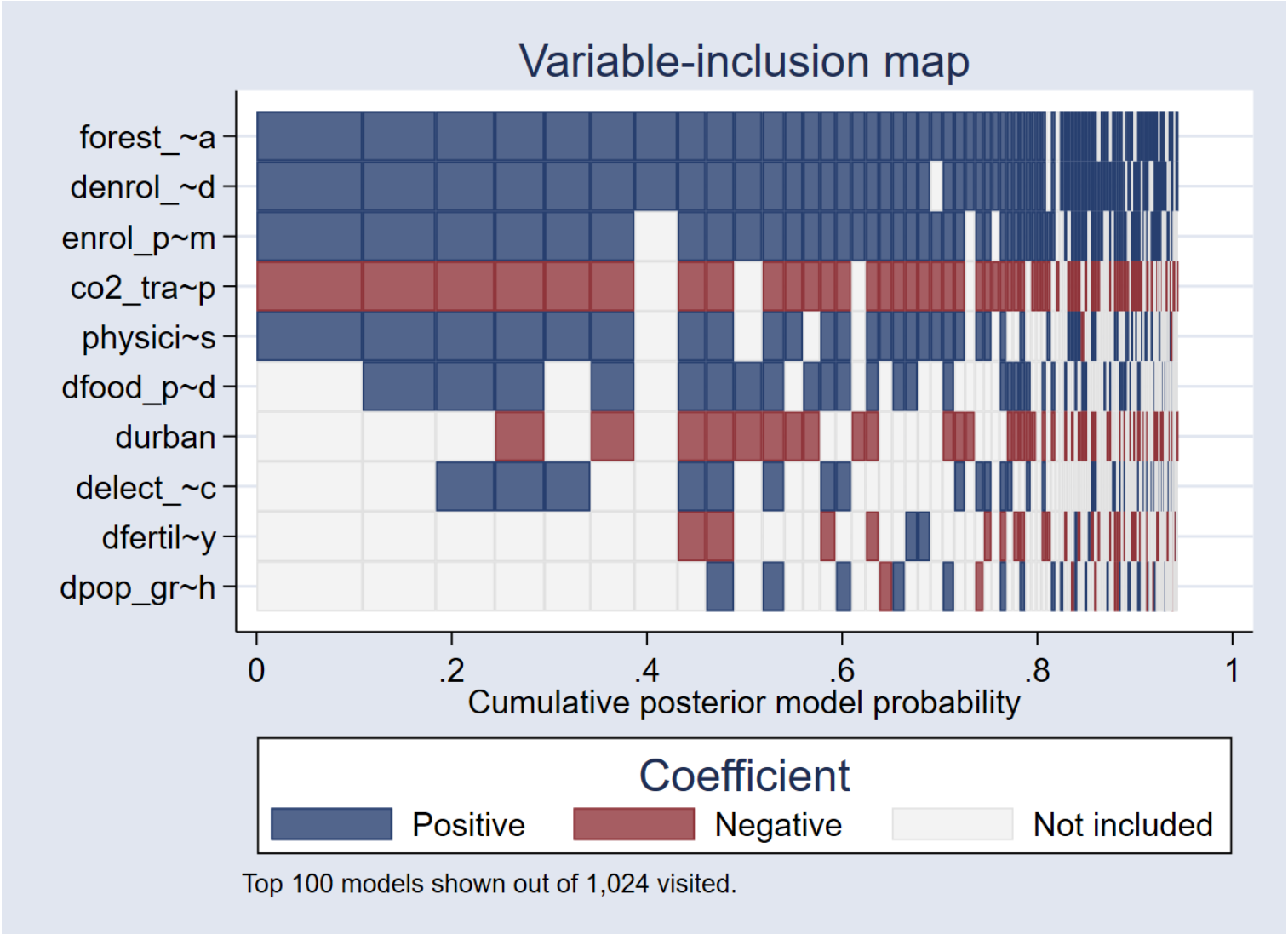
```
. bmastats pip, cutoff(.75)
Posterior inclusion probability (PIP)
No. of obs          =     22
No. of predictors   =     10
            Groups  =     10
            Always  =      0
          Reported  =      4
No. of models       = 1,024
Mean model size     = 6.143
```

	PIP	Group
forest_area	.95386	4
denrol_secnd	.95088	8
enrol_prim	.87058	9
co2_transp	.83225	3
Always		
_cons	1	0

```
Note: Using analytical PMPs.
Note: 6 predictors with PIP less than
      .75 not shown.
```

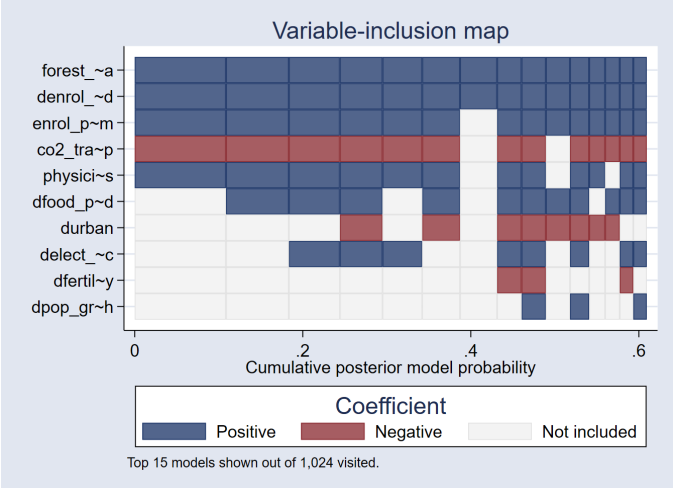
Variable inclusion map `bmagraph varmap`

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

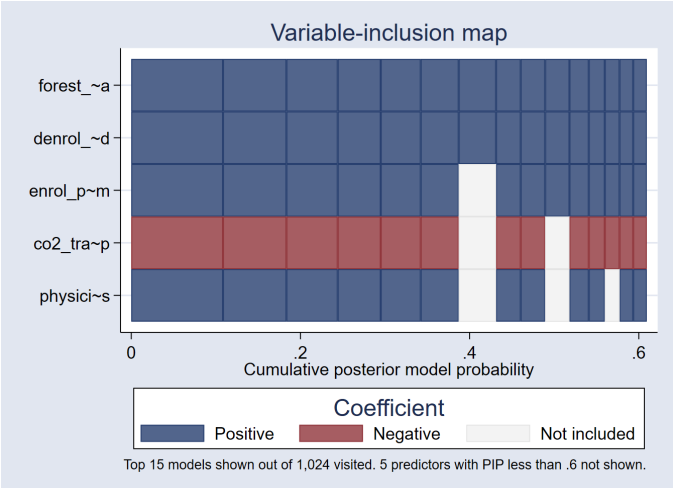


Variable inclusion map `bmagraph varmap`

```
. bmagraph varmap, top(15) legend(rows(1))
Computing model probabilities ...
```




```
. bmagraph varmap, top(15) pipcutoff(.6) legend(rows(1))
Computing model probabilities ...
```



High posterior model probabilites `bmastats models`

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

	Analytical PMP	Model size
Rank		
1	.1085	5
2	.07494	6
3	.06061	7
4	.05085	8
5	.04724	6

Variable-inclusion summary

	Rank 1	Rank 2	Rank 3	Rank 4	Rank 5
co2_transp	x	x	x	x	x
forest_area	x	x	x	x	x
denrol_secnd	x	x	x	x	x
enrol_prim	x	x	x	x	x
physicians	x	x	x	x	x
dfood_prod		x	x	x	
delect_acc			x	x	x
durban				x	

Legend:
x - estimated

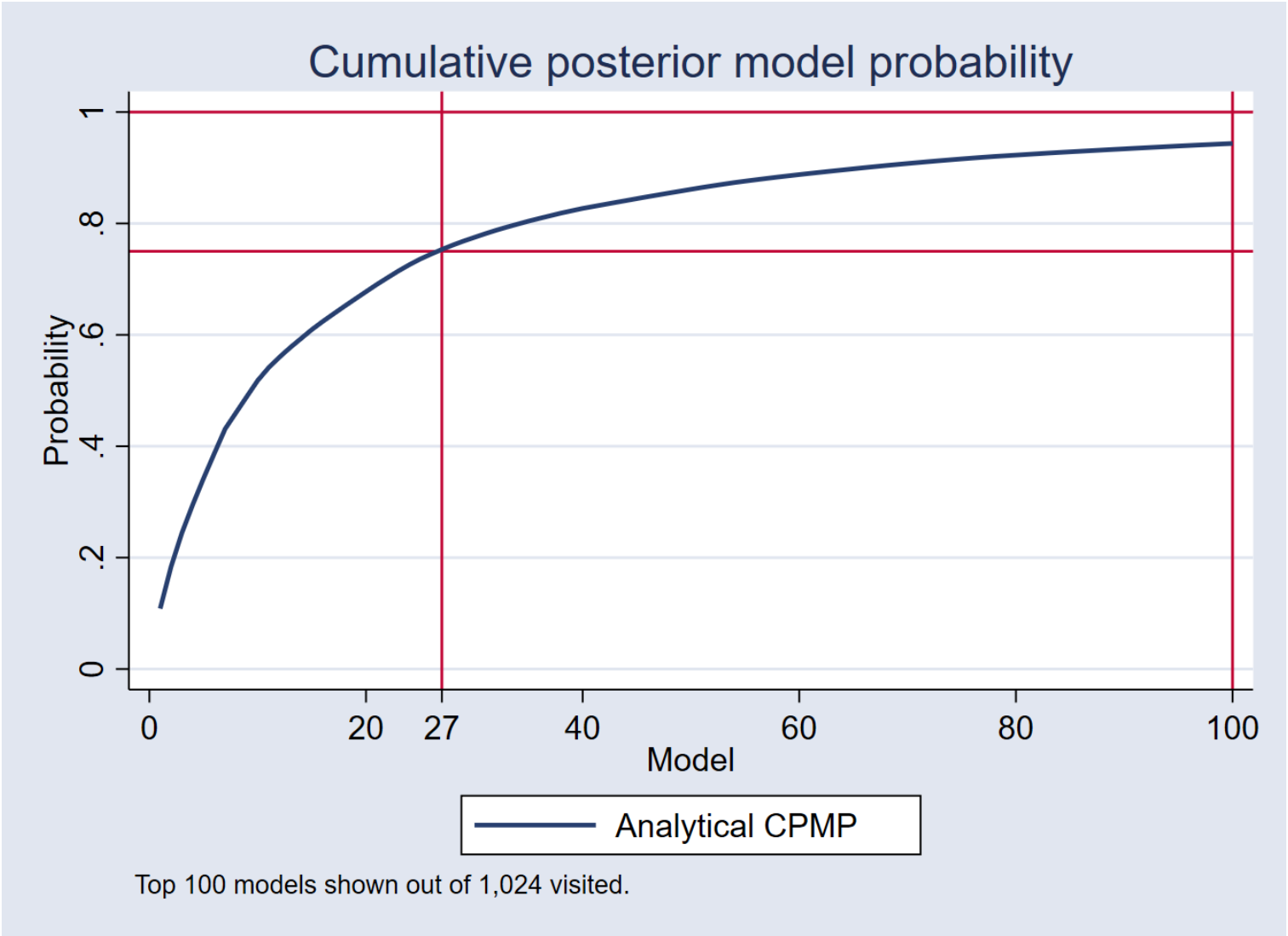
Cumulative posterior model probability `bmastats models,cumulative`

```
. bmastats models, cumulative(.75) novartable
Computing model probabilities ...
Model summary          Number of models:
                                Visited = 1,024
                               Reported =    27
```

Rank	Analytical CPMP	Model size
1	.1085	5
2	.1834	6
3	.2441	7
4	.2949	8
5	.3421	6
6	.3871	7
7	.4312	2
8	.4603	9
9	.4893	10
10	.518	5
11	.5408	9
12	.5598	6
13	.5772	6
14	.5934	8
15	.6095	8
16	.6239	4
17	.6377	8
18	.6512	6
19	.6644	7
20	.6775	7
21	.6905	6
22	.7027	4
23	.7148	8
24	.7259	7
25	.736	3
26	.7449	7
27	.7533	7

Cumulative posterior model probability `bmagraph pmp,cumulative`

```
. bmagraph pmp, cumulative xline(27 100) yline(.75 1) xlabel(27, add)
note: frequency estimates not available with model enumeration; option
      nofreqline implied.
```



Model size distribution `bmastats msize`

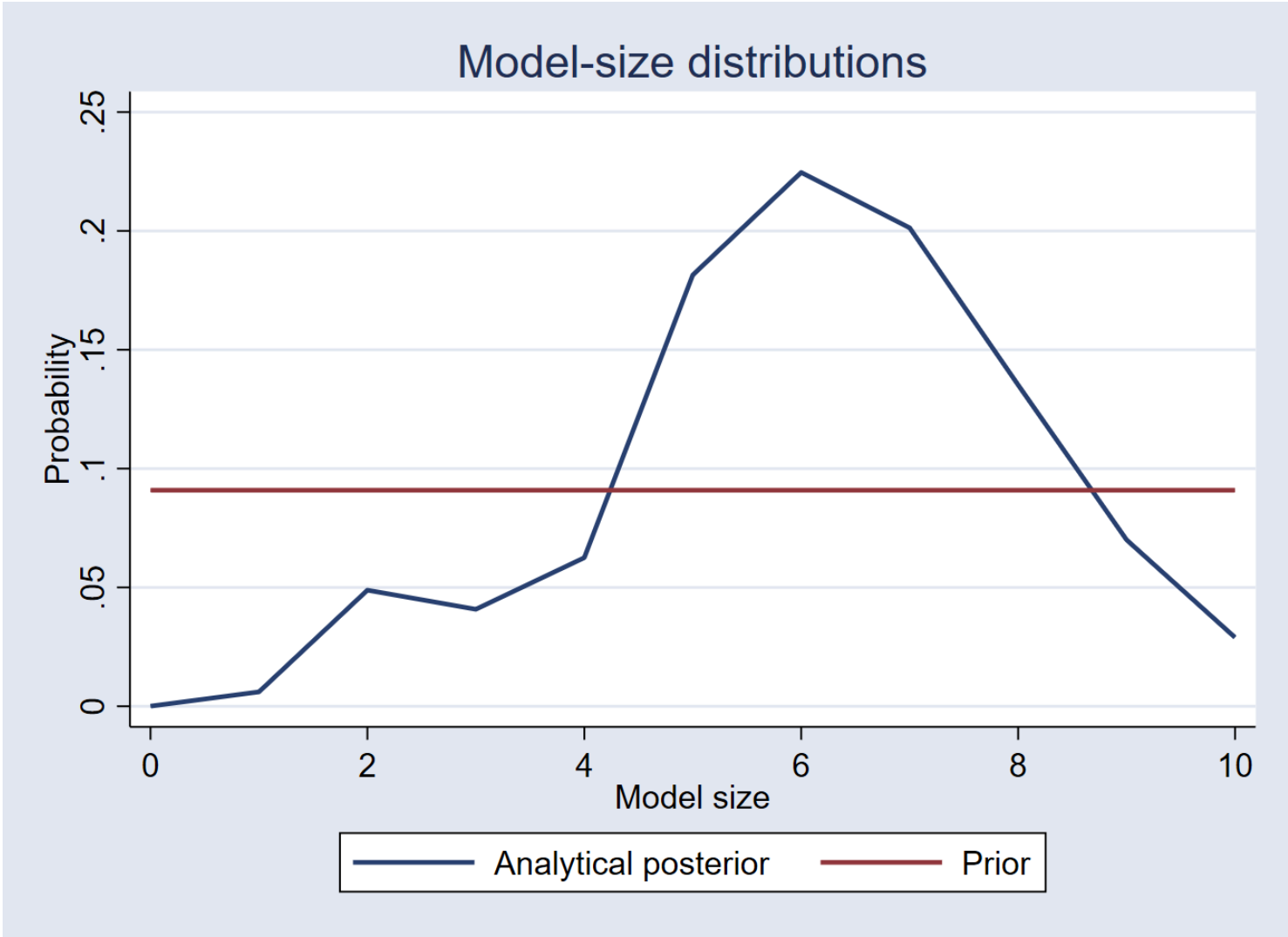
```
. bmastats msize
Model-size summary
Number of models = 1,024
Model size:
  Minimum =  0
  Maximum = 10
```

	Mean	Median
Prior		
Analytical	5.0000	5
Posterior		
Analytical	6.1426	6

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

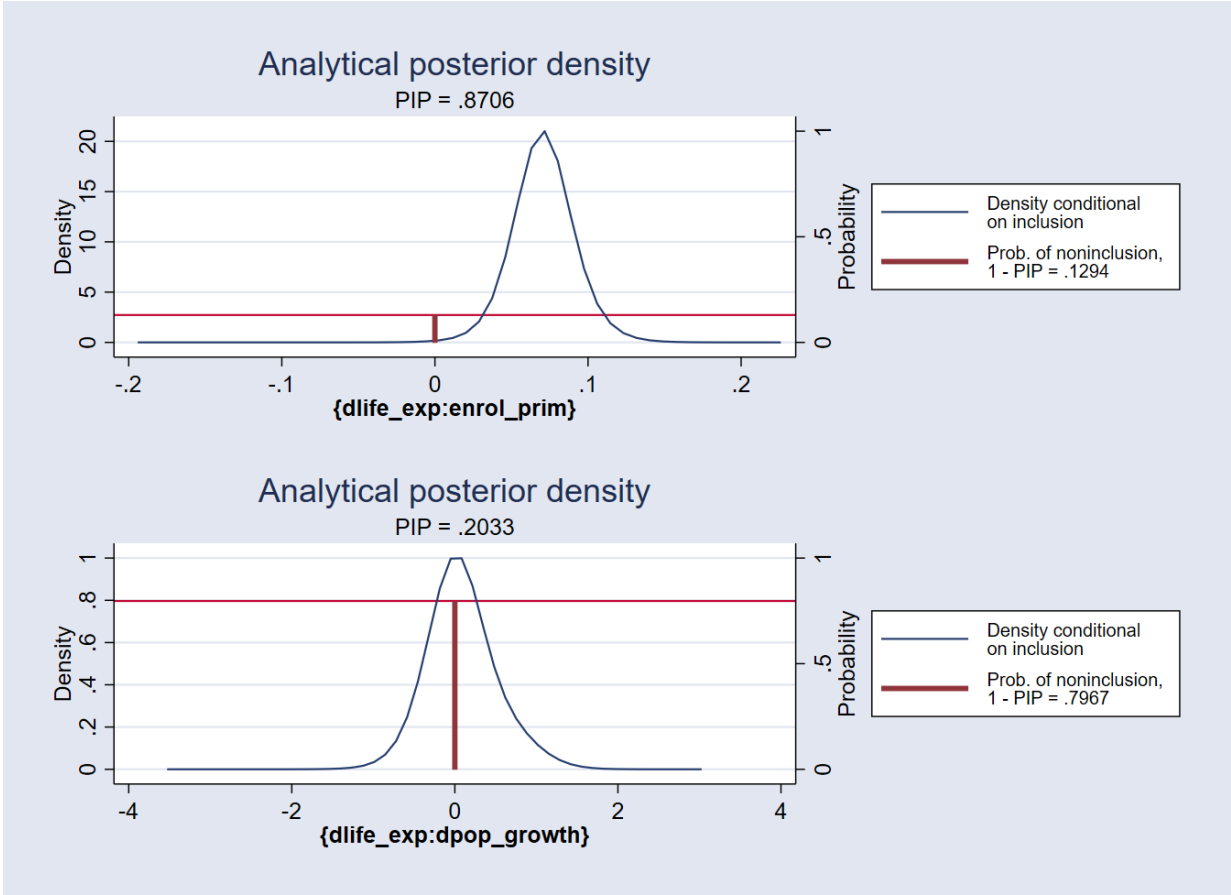
Model size distribution `bmagraph msize`

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



Posterior density for betas `bmagraph coefdensity`

```
. bmagraph coefdensity {enrol_prim},name(coefd_enrol_prim,replace) ///
>    legend(size(small) rows(2) pos(3))
. bmagraph coefdensity {dpop_growth},name(coefd_dpop_growth,replace) ///
>    legend(size(small) rows(2) pos(3))
. graph combine coefd_enrol_prim coefd_dpop_growth,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 durban co2_transp forest_area dfood_prod,dw
Doppelhofer-Weeks jointness
```

	durban	co2_transp	forest_area	dfood_prod
durban	.	-.7318593	.4996025	1.523982
co2_transp	-.7318593	.	-2.381376	.7966998
forest_area	.4996025	-2.381376	.	1.088441
dfood_prod	1.523982	.7966998	1.088441	.

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

- co2_transp and forest_area 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.
- dfood_prod and durban are significant complements: Each of them add relevant information when they are both included as predictors in the same model.

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 dfood_prod delect_acc co2_transp           ///
>                 forest_area durban dpop_growth dfertility        ///
>                 denrol_secnd enrol_prim physicians
. collect _r_b=e(b_bma):                                           ///
>     bmaregress dlife_exp dfood_prod delect_acc co2_transp        ///
>                 forest_area durban dpop_growth dfertility        ///
>                 denrol_secnd enrol_prim physicians
.
. 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
dfood_prod	0.02	0.01
delect_acc	0.02	0.01
CO2 emiss.transp. (% of tot fuel)	-0.05	-0.05
Forest area (% of land area)	0.21	0.22
durban	-0.82	-0.43
dpop_growth	0.01	0.02
dfertility	-0.80	-0.22
denrol_secnd	0.05	0.05
School enroll. primary (% gross)	0.08	0.06
Physicians (per 1,000 people)	0.33	0.32
Intercept	-14.86	-13.43

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.

Credible Intervals with `bayestats summary`

```
. 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						
dfood_prod	.008471	.0094742	.000095	.0058945	0	.0283774
delect_acc	.0090061	.0149449	.000149	0	0	.0467779
co2_transp	-.0476273	.0281324	.000281	-.0540539	-.0928139	0
forest_area	.2228425	.0751102	.000751	.2329875	0	.3397013
durban	-.4340334	.66432	.006643	0	-2.073307	.0172046
dpop_growth	.018259	.2066398	.002066	0	-.398241	.6208839
dfertility	-.2271458	.9803636	.009804	0	-3.429898	1.131774
denrol_secnd	.0509952	.0209893	.000208	.0510589	0	.0921971
enrol_prim	.0619294	.0309827	.00031	.0676031	0	.1119484
physicians	.3146744	.2391575	.002392	.3617826	0	.7245469
_cons	-13.42049	3.968817	.039688	-14.05391	-19.80799	-4.735401
sigma2	.00349	.0023631	.000024	.0027768	.0013113	.0097628
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 dfood_prod delect_acc co2_transp forest_area    ///
>     durban dpop_growth dfertility denrol_secnd enrol_prim physicians, ///
>     gprior(robust) rseed(123) saving(bma_robust, replace) notable
Burn-in ...
Simulation ...
Computing model probabilities ...
Bayesian model averaging                          No. of obs        =     22
Linear regression                                 No. of predictors =     10
MC3 and adaptive MH sampling                                 Groups =     10
                                                             Always =      0
                                                  No. of models     =    386
                                                     For CPMP >= .9 =    126
Priors:                                           Mean model size   =  6.607
  Models: Beta-binomial(1, 1)                     Burn-in           =  2,500
   Cons.: Noninformative                          MCMC sample size  = 10,000
   Coef.: Zellner´s g                             Acceptance rate   = 0.5987
       g: Robust
  sigma2: Noninformative                          Mean sigma2       =  0.005
Sampling correlation = 0.9540
file bma_robust.dta saved.
```

- The sampling correlation can be checked as an 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 dfood_prod delect_acc co2_transp forest_area     ///
>       durban dpop_growth dfertility denrol_secnd enrol_prim physicians, ///
>       gprior(robust) rseed(123) saving(bma_robust, replace) noheader
Burn-in ...
Simulation ...
Computing model probabilities ...
```

dlife_exp	Mean	Std. dev.	Group	PIP
denrol_secnd	.0478065	.0235847	8	.9199
forest_area	.1976912	.0960191	4	.9139
co2_transp	-.0427626	.0284823	3	.8318
enrol_prim	.0571541	.0364731	9	.8228
physicians	.2614008	.2477664	10	.7025
dfood_prod	.0086104	.0102062	1	.581
durban	-.4821421	.7148029	5	.5252
delect_acc	.0105903	.0174731	2	.4774
dfertility	-.4020217	1.398836	7	.4264
dpop_growth	.0343917	.3326273	6	.4056
Always				
_cons	-12.06564	4.653832	0	1

```
Note: Coefficient posterior means and std. dev. estimated from 386 models.
Note: Default prior is used for models.
```

	Mean	Std. dev.	MCSE	Median	Equal-tailed [95% cred.	interval]
g	30.43109	142.2239	3.71803	18.19193	4.112701	94.26857
Shrinkage	.9353532	.0479047	.001194	.9478948	.8044086	.9895033

```
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 dfood_prod delect_acc co2_transp forest_area     ///
>       durban dpop_growth dfertility denrol_secnd enrol_prim physicians, ///
>       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        =     22
Linear regression                               No. of predictors =     10
MC3 and adaptive MH sampling                              Groups  =     10
                                                          Always  =      0
                                                No. of models     =    426
                                                   For CPMP >= .9 =    160
Priors:                                         Mean model size   =  4.932
  Models: Beta-binomial, mean = 2               Burn-in           =  2,500
   Cons.: Noninformative                        MCMC sample size  = 10,000
   Coef.: Zellner´s g                           Acceptance rate   = 0.6194
       g: Hyper-g(3)
  sigma2: Noninformative                        Mean sigma2       =  0.007
Sampling correlation = 0.8889
file bma_mprior.dta saved.
```

Sensitivity analysis: Model prior (Estimation)

```
. bmaregress dlife_exp dfood_prod delect_acc co2_transp forest_area     ///
>       durban dpop_growth dfertility denrol_secnd enrol_prim physicians, ///
>       mprior(betabinomial 2) gprior(hyperg 3)                         ///
>       rseed(123) saving(bma_mprior, replace) noheader
Burn-in ...
Simulation ...
Computing model probabilities ...
```

dlife_exp	Mean	Std. dev.	Group	PIP
forest_area	.1902372	.1014379	4	.8571
denrol_secnd	.0486305	.0282031	8	.8548
enrol_prim	.0437142	.0390415	9	.6587
co2_transp	-.0348878	.0323326	3	.6542
physicians	.1915191	.2527384	10	.4847
durban	-.3961861	.706564	5	.3739
dfood_prod	.0050693	.0092194	1	.3379
dfertility	-.3701969	1.364219	7	.2748
delect_acc	.005283	.0143608	2	.2329
dpop_growth	.0427503	.2770385	6	.2026
Always				
_cons	-10.47968	4.994059	0	1

Note: Coefficient posterior means and std. dev. estimated from 426 models.

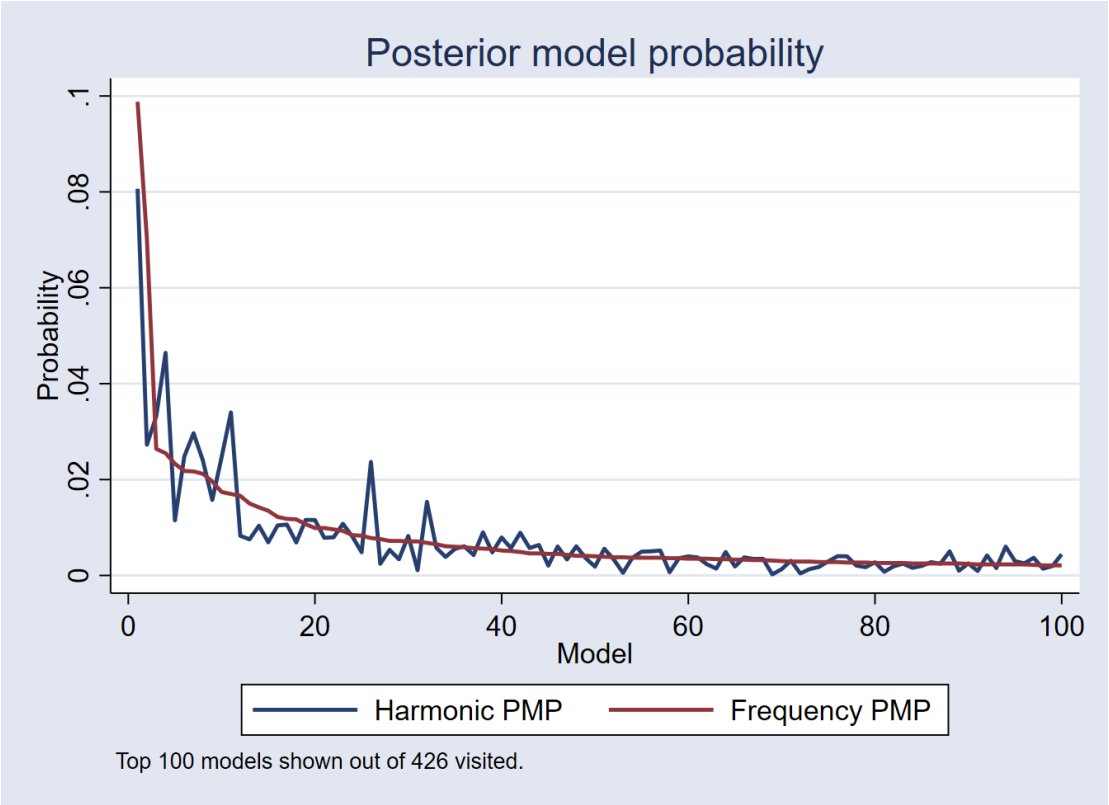
	Mean	Std. dev.	MCSE	Median	Equal-tailed [95% cred.	interval]
g	31.81657	79.80838	1.84292	17.48761	3.479734	126.1176
Shrinkage	.9323111	.0564595	.001458	.9459097	.7767722	.9921333

```
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	3.427504	-1.858614	36.77494
bma_robust	1.908033	-1.661124	37.65425
bma_mprior	2.77926	-1.546922	36.7759

```
Notes: Results using analytical and frequency PMPs.
       Result bma_robust 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 with the robust gprior.

```
. estimates restore bma_robust
(results bma_robust are active now)
. 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	.2491148	.6138077	-4.069	.814
pmean	22	.1900455	.1577893	-.1105771	.4442372
pmean_simul	22	.1899677	.1481466	-.0882326	.4322272
cri_l	22	.0119365	.1436502	-.2576388	.2457577
cri_u	22	.3675655	.1402754	.1082546	.6014318

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