

Outline

Model  
averaging

The Bayesian  
approach

BMA

bma  
commands

Example

BMA regression

PMP

PIP

Jointness

Compare OLS

CRI

Sensitivity

Predictions

References

## Stata webinar

# Introduction to Bayesian model averaging in Stata

Gustavo Sánchez

StataCorp LLC

February 13, 2024

StataCorp LLC

STATA 18

## Outline

- 1 Why model averaging
- 2 Bayesian model averaging in Stata
- 3 BMA commands
- 4 Example
  - BMA regression estimates
    - Posterior model probabilities (PMP)
    - Posterior inclusion probabilities (PIP)
  - Postestimation
    - Variable inclusion dependence
    - Comparison with linear regression
    - Sensitivity analysis
    - Predictions

## Why Model averaging

- Regression modeling is mostly focussed on deriving one (best) model.
- Researchers use statistical tools to find that best model, in terms of:
  - Model specification: AIC, BIC, Hannan-Quinn, others.
  - Predictive accuracy: MSE, MAE, MAPE, among others.
- Then, what if we select the wrong model?

## Why Model averaging

- Model averaging addresses model uncertainty, which helps reduce the risk of making inference and getting conclusions based on the wrong model.
- Consider the following potential model specifications for life expectancy (See, for example, Rizzo (2019)):

$$life\_exp = \alpha_1 + \beta_{food} * food\_prod + \beta_{elect} * elect\_acc + \epsilon_1$$

$$life\_exp = \alpha_2 + \beta_{pop} * pop\_growth + \beta_{urb} * urban + \epsilon_2$$

$$life\_exp = \alpha_3 + \beta_{forest} * forest + \beta_{school} * schooling + \epsilon_3$$

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.
forest	: Forest area (% of total land).
schooling	: Years of schooling.

Outline

Model  
averaging

The Bayesian  
approach

BMA

bma  
commands

Example

BMA regression

PMP

PIP

Jointness

Compare OLS

CRI

Sensitivity

Predictions

References

## Why Model averaging

- Proposes estimators based on a weighted average of a number of potentially feasible models.
- Frequentists and Bayesians propose different ways for selecting those weights. Steel (2020) provides a broad description of the methods associated to both approaches.
- The Bayesian approach differs from the frequentist in a fundamental theoretical view of the model and the parameters, so let's just have a quick overview on those differences.

# The Bayesian approach

Outline

Model  
averaging

The Bayesian  
approach

BMA

bma  
commands

Example

BMA regression

PMP

PIP

Jointness

Compare OLS

CRI

Sensitivity

Predictions

References

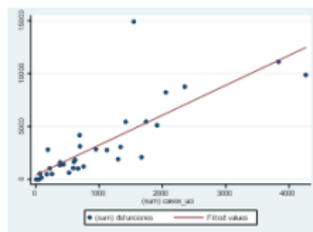
## Frequentist

Data hypothetically repeatable

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

	month	defunciones	casos_ucl
1.	2021e11	631	524
2.	2021e12	1912	1398
3.	2022e1	5453	1740
4.	2022e2	4183	691
5.	2022e3	1600	382
6.	2022e4	1422	436
7.	2022e5	1848	620
8.	2022e6	1663	691
9.	2022e7	3133	696
10.	2022e8	1046	219

Theoretical  
Model



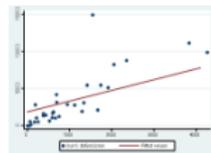
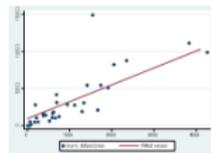
## Bayesian

Data known

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

	month	defunciones	casos_ucl
1.	2021e11	631	524
2.	2021e12	1912	1398
3.	2022e1	5453	1740
4.	2022e2	4183	691
5.	2022e3	1600	382
6.	2022e4	1422	436
7.	2022e5	1848	620
8.	2022e6	1663	691
9.	2022e7	3133	696
10.	2022e8	1046	219

Theoretical  
Model



# Bayesian Analysis vs. Frequentist Analysis

Outline

Model  
averaging

The Bayesian  
approach

BMA

bma  
commands

Example

BMA regression

PMP

PIP

Jointness

Compare OLS

CRI

Sensitivity

Predictions

References

## 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  $H_0$  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

- Fundamental equation for Bayesian analysis (based on the Bayes' Theorem):

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

Where:

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

$\pi(\theta)$ : Prior distribution for the parameters.

$p(\theta|y)$ : Posterior distribution for the parameters.

## 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>

## 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:

$\mathbf{y} = (y_1, y_2, \dots, y_n)'$	: (nx1) vector of outcome values.
$\mathbf{1}_n$	: vector of ones.
$\mathbf{X}_j$	: $n \times p_j$ design matrix.
$\beta_j$	: ( $p_j \times 1$ ) vector of coef. for model j
$\epsilon = (\epsilon_1, \epsilon_2, \dots, \epsilon_n)$	: (nx1) vector of error terms.

- Two probabilities are fundamental for the inference using the Bayesian approach for model averaging:
  - Posterior model probabilities (PMPs)
  - Posterior inclusion probabilities (PIPs)

Outline

Model  
averaging

The Bayesian  
approach

BMA

bma  
commands

Example

BMA regression

PMP

PIP

Jointness

Compare OLS

CRI

Sensitivity

Predictions

References

## BMA prior probability distributions

- Let's recall our linear model specification:

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

- Priors for a BMA linear regression with fixed  $\mathbf{g}$ :

$$M_j \sim P(M_j)$$

$$\boldsymbol{\beta}_j | \alpha, \sigma, M_j \sim N_{p_j}(0, \mathbf{g} \sigma^2 (\mathbf{X}'_j \mathbf{X}_j)^{-1})$$

$$\alpha | \sigma, M_j \sim 1$$

$$\sigma | M_j \sim \sigma^{-1}$$

- In addition to the priors for the parameters  $(\boldsymbol{\beta}_j, \alpha, \sigma)$ , BMA specifies a discrete model prior  $(P(M_j))$  over the models space  $\mathbf{M}_F = M_1, M_2, \dots, M_{2^p}$ .
- Prior for  $\mathbf{g}$ : fixed or random hyperprior  $p(\mathbf{g})$

Outline

Model  
averaging

The Bayesian  
approach

BMA

bma  
commands

Example

BMA regression

PMP

PIP

Jointness

Compare OLS

CRI

Sensitivity

Predictions

References

## BMA posterior model and inclusion probabilities

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

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

Where:  $f(\mathbf{y} | M_j)$ : Likelihood of  $\mathbf{y}$  under model  $M_j$ .

$P(\mathbf{y})$ : marginal probability of  $\mathbf{y}$  over the model space  $\mathbf{M}_F$

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

$$PIP = \sum_{j \in \mathbf{J}_F} I(\mathbf{X}_k \in M_j) P(M_j | \mathbf{y})$$

Where  $I(\cdot)$  is the indicator function.

Outline

Model  
averaging

The Bayesian  
approach

BMA

bma  
commands

Example

BMA regression

PMP

PIP

Jointness

Compare OLS

CRI

Sensitivity

Predictions

References

## BMA posterior probability distributions

- Posterior distribution of  $\beta$  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  $\beta$   
for a Bayesian linear regression model  $M_j$

- BMA coefficient estimates for the linear model:

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

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

Outline

Model  
averagingThe Bayesian  
approach

BMA

bma  
commands

Example

BMA regression

PMP

PIP

Jointness

Compare OLS

CRI

Sensitivity

Predictions

References

## Stata's BMA suite consists of the following commands

### *Command*

### *Description*

#### **Setup**

`splitsample`

Split samples for training, validation and prediction

`v1`

Manage large variable lists

#### **Estimation**

`bmaregress`

BMA linear regression

`bmacoefsample`

Posterior samples of regression coefficients

#### **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

#### **Postestimation 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

#### **Predictions**

`bmapredict`

BMA predictions

Outline

Model  
averaging

The Bayesian  
approach

BMA

**bma**  
commands

Example

BMA regression

PMP

PIP

Jointness

Compare OLS

CRI

Sensitivity

Predictions

References

## Example: Life expectancy model for Chile

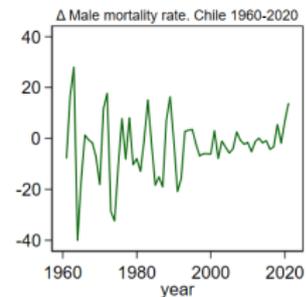
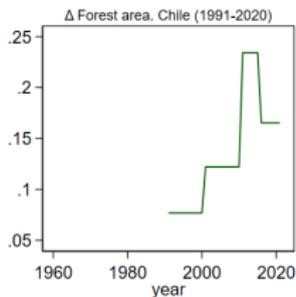
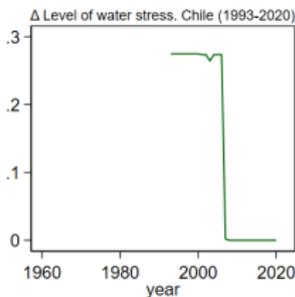
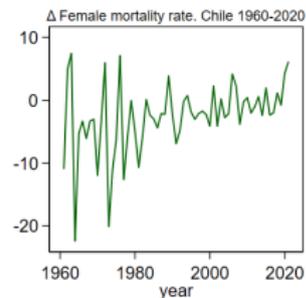
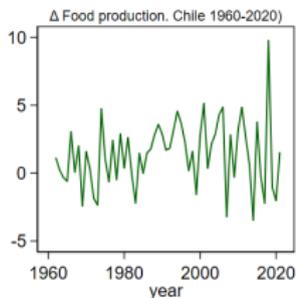
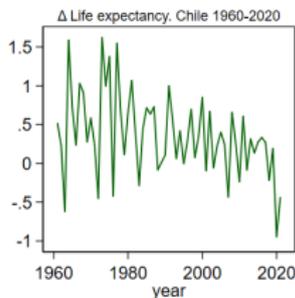
- Let's work with a model for life expectancy including the 10 explanatory variables described below:

```
. describe life_exp mrate_m mrate_f population gdp_cap ///
> forest_area electr_r urban_pop food_prd water
```

Variable name	Storage type	Display format	Value label	Variable label
life_exp	double	%10.0g		Life expectancy at birth, total (years)
mrate_m	double	%10.0g		Mortality rate, adult, male (per 1,000 male adults)
mrate_f	double	%10.0g		Mortality rate, adult, female (per 1,000 female adults)
population	long	%10.0g		Population, total
gdp_cap	double	%10.0g		GDP per capita (constant 2015 US\$)
forest_area	double	%10.0g		Forest area (% of land area)
electr_r	double	%10.0g		Access to electricity,rural (% of rural population)
urban_pop	double	%10.0g		Urban population (% of total population)
food_prd	double	%10.0g		Food prod. index (2014-16 = 100)
water	double	%10.0g		Level of water stress: freshwater withdrawal as a proportion of available freshw

- 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/chile>

Outline

Model  
averaging

The Bayesian  
approach

BMA

bma  
commands

Example

BMA regression

PMP

PIP

Jointness

Compare OLS

CRI

Sensitivity

Predictions

References

# BMA regression

```
. bmaregress dlife_exp dmrates_m dmrates_f dpopulation dgdp_cap ///
> dforest_area delectr_r durban_pop dfood_prd dwater, ///
> saving($simul_dir\bma_enumerate_c11,replace)
```

Enumerating models ...  
Computing model probabilities ...

```
Bayesian model averaging          No. of obs          =          28
Linear regression                 No. of predictors   =           9
Model enumeration                  Groups              =           9
                                   Always                 =           0
Priors:
Models: Beta-binomial(1, 1)
Cons.: Noninformative
Coef.: Zellner's g
      g: Benchmark, g = 81
sigma2: Noninformative
                                   No. of models       =          512
                                   For CPMP >= .9         =           29
                                   Mean model size      =          1.965
                                   Shrinkage, g/(1+g)   = 0.9878
                                   Mean sigma2           =          0.069
```

dlife_exp	Mean	Std. dev.	Group	PIP
dmrate_m	-.0483797	.0283438	1	.81719
dmrate_f	-.0421563	.0467395	2	.52719
dfood_prd	.0066501	.0148556	8	.21246
durban_pop	.048554	.2168773	7	.089491
dgdp_cap	.0000133	.0000674	4	.081959
delectr_r	.0008296	.0047437	6	.073272
dwater	.0086067	.129569	9	.05864
dpopulation	-1.91e-08	2.89e-07	3	.053395
dforest_area	.0070917	.26385	5	.05132
<b>Always</b>				
_cons	.0566042	.1048209	0	1

Note: Coefficient posterior means and std. dev. estimated from 512 models.  
Note: Default priors are used for models and parameter g.

```
file C:\Users\gas\Documents\webinars\2024\chile\simul\bma_enumerate_c11.dta sav
> ed.
. estimates store bmareg_enum1
```

## Regression output

- Estimation default
  - Model enumeration (<12 predictors) ( $2^9 = 512$  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  $g = 81$
- Results
  - Little shrinkage:  $81 / (1 + 81) = .9878$
  - Mean model size: 1.965
  - Top three predictors: dmrates\_male, dmrates\_female, dfood\_prod (PIPs>.2)
  - No other predictors seem relevant (with PIPs<.10)
  - BMA estimates based on  $2^9 = 512$  models. 29 of those models contribute to .9 of the cumulative PMP.
  - Estimation stored for some of the postimation analysis

Outline

Model  
averaging

The Bayesian  
approach

BMA

bma  
commands

Example

BMA regression

PMP

PIP

Jointness

Compare OLS

CRI

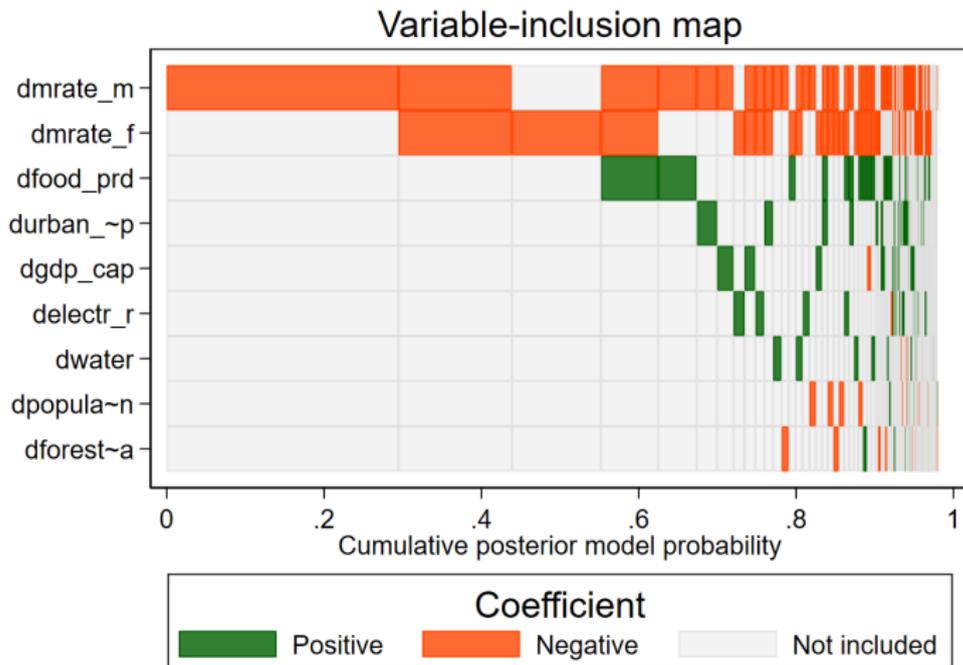
Sensitivity

Predictions

References

# Variable inclusion map `bmagraph` `varmap`

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



Top 100 models shown out of 512 visited.

Outline

Model  
averaging

The Bayesian  
approach

BMA

bma  
commands

Example

BMA regression

PMP

PIP

Jointness

Compare OLS

CRI

Sensitivity

Predictions

References

## BMA regression (always option)

```
. bmaregress dlife_exp dmrte_m dmrte_f dpopulation dgdp_cap ///
> dforest_area delectr_r durban_pop dwater (dfood_prd,always), ///
> saving($simul_dir\bma_enumerate_c12,replace)
```

Enumerating models ...

Computing model probabilities ...

Bayesian model averaging	No. of obs	=	28
Linear regression	No. of predictors	=	9
Model enumeration	Groups	=	8
	Always	=	1
Priors:	No. of models	=	256
Models: Beta-binomial(1, 1)	For CPMP >= .9	=	14
Cons.: Noninformative	Mean model size	=	2.853
Coef.: Zellner's g			
g: Benchmark, g = 81	Shrinkage, g/(1+g)	=	0.9878
sigma2: Noninformative	Mean sigma2	=	0.059

dlife_exp	Mean	Std. dev.	Group	PIP
dmrate_m	-.0554959	.0244436	1	.91523
dmrate_f	-.0410957	.0423503	2	.56898
durban_pop	.0454242	.2107874	7	.089166
delectr_r	.0003983	.0035752	6	.06221
dwater	-.0047232	.1245774	8	.056169
dpopulation	-1.30e-08	2.62e-07	3	.054345
dforest_area	.0113007	.2522723	5	.054344
dgdp_cap	1.31e-06	.0000428	4	.05231
<b>Always</b>				
dfood_prd	.0312108	.0164328	0	1
_cons	.0047405	.0967019	0	1

Note: Coefficient posterior means and std. dev. estimated from 256 models.

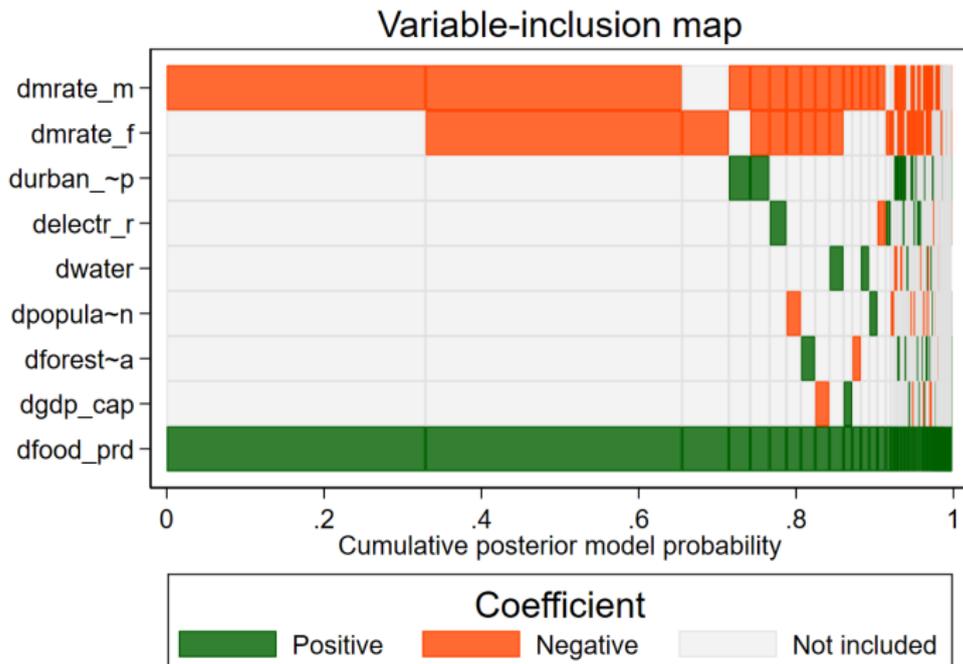
Note: Default priors are used for models and parameter g.

```
file C:\Users\gas\Documents\webinars\2024\chile\simul\bma_enumerate_c12.dta sav
> ed.
```

```
. estimates store bmareg_enum2
```

# Variable inclusion map `bmagraph` `varmap`

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



Top 100 models shown out of 256 visited.

Outline

Model  
averaging

The Bayesian  
approach

BMA

bma  
commands

Example

BMA regression

PMP

PIP

Jointness

Compare OLS

CRI

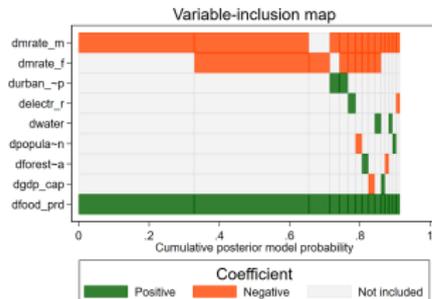
Sensitivity

Predictions

References

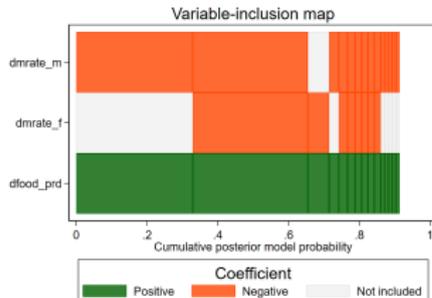
## Variable inclusion map `bmagraph` `varmap`

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



Top 15 models shown out of 256 visited.

```
. bmagraph varmap, top(15) pipcutoff(.5) legend(rows(1))
Computing model probabilities ...
```



Top 15 models shown out of 256 visited. 6 predictors with PIP less than .5 not shown.

Outline

Model  
averaging

The Bayesian  
approach

BMA

bma  
commands

Example

BMA regression

PMP

PIP

Jointness

Compare OLS

CRI

Sensitivity

Predictions

References

```
. bmastats models
```

```
Computing model probabilities ...
```

```
Model summary          Number of models:
                        Visited = 256
                        Reported =  5
```

	Analytical PMP	Model size
<b>Rank</b>		
1	.329	2
2	.3258	3
3	.0596	2
4	.02725	3
5	.02448	4

```
Variable-inclusion summary
```

	Rank 1	Rank 2	Rank 3	Rank 4	Rank 5
<b>dmrate_m</b>	<b>x</b>	<b>x</b>		<b>x</b>	<b>x</b>
<b>dfood_prd</b>	<b>x</b>	<b>x</b>	<b>x</b>	<b>x</b>	<b>x</b>
<b>dmrate_f</b>		<b>x</b>	<b>x</b>		<b>x</b>
<b>durban_pop</b>				<b>x</b>	<b>x</b>

```
Legend:
```

```
x - estimated
```

Outline

Model  
averaging

The Bayesian  
approach

BMA

`bma`  
commands

Example

BMA regression

PMP

PIP

Jointness

Compare OLS

CRI

Sensitivity

Predictions

References

## Cumulative posterior model probability `bmastats models, cumulative`

```
. bmastats models, cumulative(.80) novartable
Computing model probabilities ...
```

```
Model summary          Number of models:
                        Visited = 256
                        Reported =  7
```

	Analytical CPMP	Model size
<b>Rank</b>		
1	.329	2
2	.6548	3
3	.7144	2
4	.7417	3
5	.7661	4
6	.7877	4
7	.8061	4

Outline

Model  
averaging

The Bayesian  
approach

BMA

bma  
commands

Example

BMA regression

PMP

PIP

Jointness

Compare OLS

CRI

Sensitivity

Predictions

References

## Cumulative posterior model probability `bmagraph pmp, cumulative`

```
. bmagraph pmp, cumulative xline(7 100) yline(.80 1) xlabel(7, add)
note: frequency estimates not available with model enumeration; option
nofreqline implied.
```

Outline

Model  
averaging

The Bayesian  
approach

BMA

bma  
commands

Example

BMA regression

**PMP**

PIP

Jointness

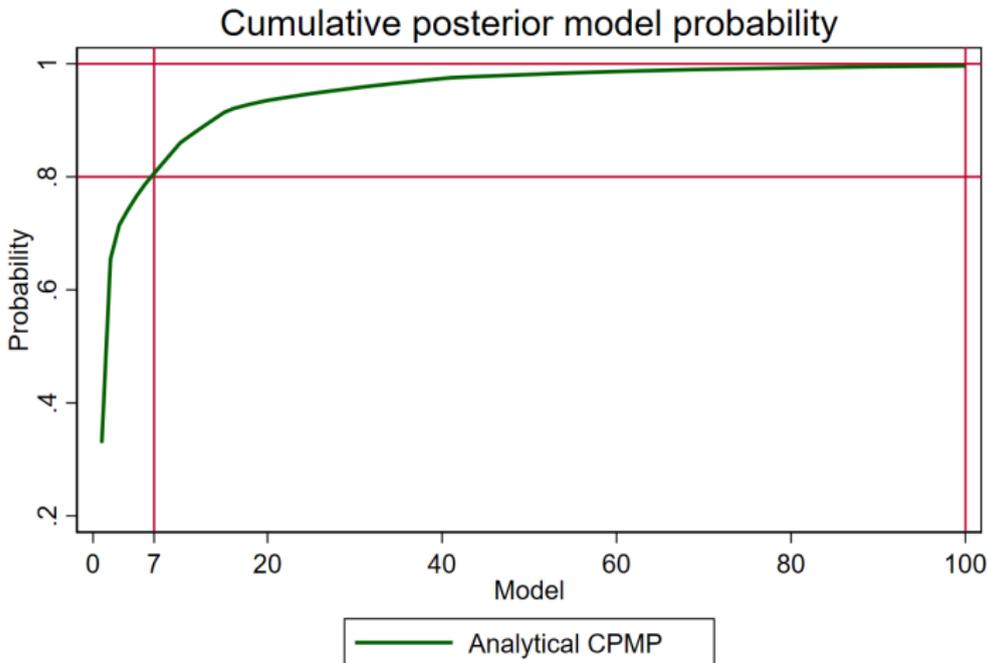
Compare OLS

CRI

Sensitivity

Predictions

References



Top 100 models shown out of 256 visited.

## Model size distribution `bmastats msize`

```
. bmastats msize
```

**Model-size summary**

**Number of models = 256**

**Model size:**

**Minimum = 1**

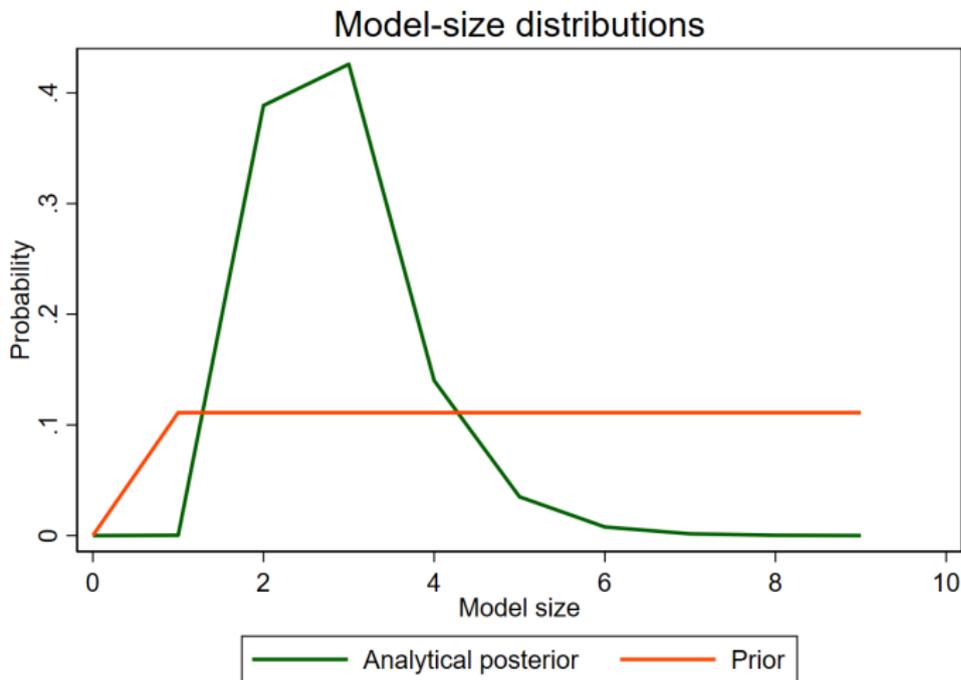
**Maximum = 9**

	Mean	Median
<b>Prior</b>		
<b>Analytical</b>	5.0000	5
<b>Posterior</b>		
<b>Analytical</b>	2.8527	3

**Note: Frequency summaries not available.**

## Model size distribution `bmagraph` `msize`

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



Outline

Model  
averaging

The Bayesian  
approach

BMA

`bma`  
commands

Example

BMA regression

PMP

PIP

Jointness

Compare OLS

CRI

Sensitivity

Predictions

References

## Predictors with highest probability of inclusion `bmastats pip`

```
. bmastats pip, cutoff(.50)
Posterior inclusion probability (PIP)
No. of obs          =    28
No. of predictors   =    9
                   Groups =    8
                   Always =    1
                   Reported =   3
No. of models       =   256
Mean model size     =  2.853
```

	PIP	Group
<code>dmrate_m</code>	.91523	1
<code>dmrate_f</code>	.56898	2
<b>Always</b>		
<code>dfood_prd</code>	1	0
<code>_cons</code>	1	0

**Note:** Using analytical PMPs.

**Note:** 6 predictors with PIP less than .5 not shown.

Outline

Model  
averaging

The Bayesian  
approach

BMA

bma  
commands

Example  
BMA regression

PMP

PIP

Jointness

Compare OLS

CRI

Sensitivity

Predictions

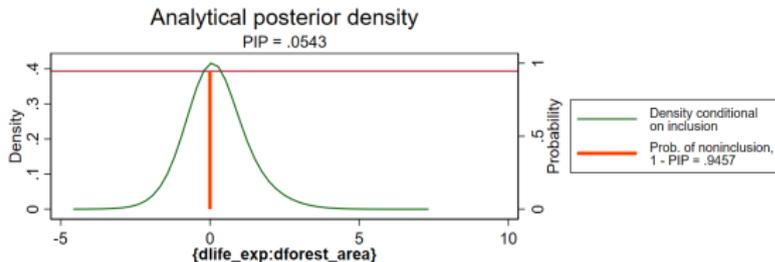
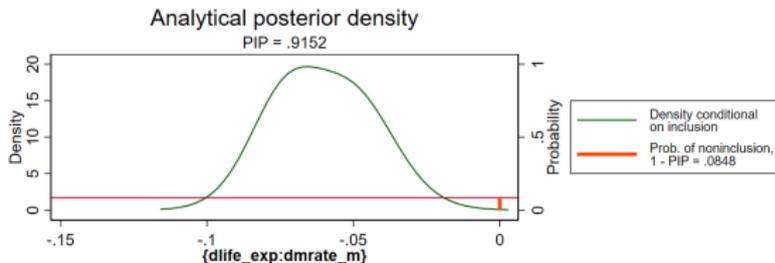
References

## Posterior density for betas `bmagraph` `coefdensity`

```
. bmagraph coefdensity {dmrate_m},name(coefd_dmrate_m,replace) ///
> legend(size(small) rows(2) pos(3))

. bmagraph coefdensity {dforest_area},name(coefd_dforest_area,replace) ///
> legend(size(small) rows(2) pos(3))

. graph combine coefd_dmrate_m coefd_dforest_area,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

Outline

Model  
averaging

The Bayesian  
approach

BMA

`bma`  
commands

Example

BMA regression

PMP

PIP

**Jointness**

Compare OLS

CRI

Sensitivity

Predictions

References

## 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

Outline

Model  
averaging

The Bayesian  
approach

BMA

`bma`  
commands

Example

BMA regression

PMP

PIP

Jointness

Compare OLS

CRI

Sensitivity

Predictions

References

## Variable inclusion dependence `bmastats jointness`

```
. bmastats jointness dmrate_m dmrate_f durban_pop dwater ,dw
Doppelhofer-Weeks jointness
```

	dmrate_m	dmrate_f	durban_pop	dwater
dmrate_m	.	-5.178394	.9036705	.1656873
dmrate_f	-5.178394	.	-.1499009	.4048622
durban_pop	.9036705	-.1499009	.	1.174585
dwater	.1656873	.4048622	1.174585	.

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

- `dmrate_m` and `dmrate_f` 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.
- `dwater` and `durban_pop` are significant complements: Each of them add relevant information when they are both included as predictors in the same model.

Outline

Model  
averaging

The Bayesian  
approach

BMA

bma  
commands

Example

BMA regression

PMP

PIP

Jointness

Compare OLS

CR1

Sensitivity

Predictions

References

# 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 dmrte_m dmrte_f dpopulation dgdp_cap    ///
>          dforest_area delectr_r durban_pop dwater dfood_prd
. collect _r_b=e(b_bma):                             ///
>          bmaregress dlife_exp dmrte_m dmrte_f dpopulation dgdp_cap    ///
>          dforest_area delectr_r durban_pop dwater (dfood_prd,always)
.
. 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
```

Outline

Model  
averagingThe Bayesian  
approach

BMA

bma  
commands

Example

BMA regression

PMP

PIP

Jointness

Compare OLS

CRI

Sensitivity

Predictions

References

Compare `bmaregress` vs. `regress`

```
. collect layout (colname#result) (program_class)
```

```
Collection: bma_compare
```

```
Rows: colname#result
```

```
Columns: program_class
```

```
Table 1: 19 x 2
```

	ols	bma_reg
<b>dmrate_m</b>	-0.05	-0.06
<b>dmrate_f</b>	-0.06	-0.04
<b>dpopulation</b>	-0.00	-0.00
<b>dgdpcap</b>	0.00	0.00
<b>dforest_area</b>	0.79	0.01
<b>delectr_r</b>	0.01	0.00
<b>durban_pop</b>	0.86	0.05
<b>dwater</b>	-0.38	-0.00
<b>dfood_prd</b>	0.03	0.03
<b>Intercept</b>	-0.10	0.00

Outline

Model  
averagingThe Bayesian  
approach

BMA

bma  
commands

Example

BMA regression

PMP

PIP

Jointness

Compare OLS

CRI

Sensitivity

Predictions

References

## 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_enum2
(results bmareg_enum2 are active now)
. bmacroefsample, rseed(123)
Simulation (10000): ....5000....10000 done
. bayestats summary
Posterior summary statistics                                MCMC sample size =    10,000
```

	Mean	Std. dev.	MCSE	Median	Equal-tailed [95% cred. interval]	
<b>dlife_exp</b>						
dmrate_m	-.0552308	.024515	.000241	-.0584593	-.0944329	0
dmrate_f	-.0414628	.0423996	.000424	-.0400205	-.1296837	0
dpopulation	-1.18e-08	2.63e-07	2.6e-09	0	-2.91e-07	0
dgdg_cap	6.79e-07	.0000446	4.5e-07	0	-1.77e-06	.0000217
dforest_area	.0115317	.2602391	.002632	0	0	.2983952
delectr_r	.0004227	.0035767	.000036	0	0	.0093204
durban_pop	.0452585	.2089718	.002128	0	0	.7810504
dwater	-.0047322	.1254232	.001254	0	-.1245275	.0076195
dfood_prd	.0315437	.0165329	.000165	.0314189	-.0010433	.0648134
_cons	.0041732	.0980046	.000098	.0049199	-.1823419	.1767347
<b>sigma2</b>	.0597448	.0197926	.000198	.0558336	.0319951	.1075204
g	81	0	0	81	81	81

Outline

Model  
averaging

The Bayesian  
approach

BMA

bma  
commands

Example

BMA regression

PMP

PIP

Jointness

Compare OLS

CRI

Sensitivity

Predictions

References

## Sensitivity analysis: Random g-prior (Header)

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

```
. bmaregress dlife_exp dmrate_m dmrate_f dpopulation dgdp_cap      ///
>   dforest_area delectr_r durban_pop dwater (dfood_prd,always), ///
>   gprior(robust) rseed(123) saving(bma_robust, replace) notable

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

Bayesian model averaging
Linear regression
MC3 and adaptive MH sampling

                                No. of obs      =      28
                                No. of predictors =      9
                                Groups           =      8
                                Always           =      1
                                No. of models    =     180
                                For CPMP >= .9 =      71
                                Mean model size  =   4.154
                                Burn-in         =   2,500
                                MCMC sample size = 10,000
                                Acceptance rate  =   0.5602

Priors:
  Models: Beta-binomial(1, 1)
  Cons.: Noninformative
  Coef.: Zellner's g
         g: Robust
  sigma2: Noninformative
                                Mean sigma2    =   0.064

Sampling correlation = 0.9924
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)

Outline

Model  
averaging

The Bayesian  
approach

BMA

bma  
commands

Example

BMA regression

PMP

PIP

Jointness

Compare OLS

CRI

Sensitivity

Predictions

References

```
. bmaregress dlife_exp dmrte_m dmrte_f dpopulation dgdp_cap ///
> dforest_area delectr_r durban_pop dwater (dfood_prd,always), ///
> gprior(robust) rseed(123) saving(bma_robust, replace) noheader
```

Burn-in ...

Simulation ...

Computing model probabilities ...

dlife_exp	Mean	Std. dev.	Group	PIP
dmrate_m	-.0479737	.0234835	1	.9169
dmrate_f	-.0455784	.038259	2	.7214
durban_pop	.1684573	.4190427	7	.3059
dwater	-.0476847	.3236723	8	.2657
delectr_r	.0014442	.006575	6	.2378
dgdp_cap	3.69e-06	.0000947	4	.237
dforest_area	.1116288	.6398553	5	.235
dpopulation	-6.28e-08	5.57e-07	3	.2342
<b>Always</b>				
dfood_prd	.0284937	.0177335	0	1
_cons	.0003427	.1808627	0	1

Note: Coefficient posterior means and std. dev. estimated from 180 models.

Note: Default prior is used for models.

	Mean	Std. dev.	MCSE	Median	Equal-tailed [95% cred. interval]	
g	20.15981	50.69094	1.13525	11.08187	3.202799	88.17337
Shrinkage	.9075516	.0588137	.004334	.9172313	.7620633	.9887859

file bma\_robust.dta saved.

```
. estimates store bma_robust
```

## Sensitivity analysis: Model prior (Header)

- Let's specify a betabinomial prior with a model size equal to 3, and a random g parameter with a hyper-g prior distribution with hyperparameter 3:

```

. bmaregress dlife_exp dmrte_m dmrte_f dpopulation dgdp_cap      ///
>   dforest_area delectr_r durban_pop dwater (dfood_prd,always), ///
>   mprior(betabinomial 3) 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 =      9
MC3 and adaptive MH sampling                               Groups          =      8
                                                           Always          =      1
                                                           No. of models   =     197
                                                           For CPMP >= .9 =      62
                                                           Mean model size =    3.578
                                                           Burn-in        =    2,500
                                                           MCMC sample size = 10,000
                                                           Acceptance rate =    0.6488

Priors:
  Models: Beta-binomial, mean = 3
  Cons.: Noninformative
  Coef.: Zellner's g
         g: Hyper-g(3)
  sigma2: Noninformative
                                                           Mean sigma2    =    0.069

Sampling correlation = 0.9772
file bma_mprior.dta saved.

```

Outline

Model  
averagingThe Bayesian  
approach

BMA

bma  
commands

Example

BMA regression

PMP

PIP

Jointness

Compare OLS

CRI

Sensitivity

Predictions

References

## Sensitivity analysis: Model prior (Estimation)

Outline

Model  
averaging

The Bayesian  
approach

BMA

bma  
commands

Example

BMA regression

PMP

PIP

Jointness

Compare OLS

CRI

Sensitivity

Predictions

References

```
. bmaregress dlife_exp dmrte_m dmrte_f dpopulation dgdp_cap ///
> dforest_area delectr_r durban_pop dwater (dfood_prd,always), ///
> mprior(betabinomial 3) gprior(hyperg 3) ///
> rseed(123) saving(bma_mprior, replace) noheader
```

Burn-in ...

Simulation ...

Computing model probabilities ...

dlife_exp	Mean	Std. dev.	Group	PIP
dmrte_m	-.0469339	.0241444	1	.8982
dmrte_f	-.0403306	.0387409	2	.6403
durban_pop	.1098543	.3370131	7	.227
delectr_r	.0010525	.0057874	6	.1828
dwater	-.0219699	.2423952	8	.1711
dforest_area	.061596	.5022718	5	.1676
dpopulation	-3.64e-08	4.38e-07	3	.1507
dgdp_cap	2.88e-06	.0000717	4	.1407
<b>Always</b>				
dfood_prd	.0271962	.0175803	0	1
_cons	.015783	.149966	0	1

Note: Coefficient posterior means and std. dev. estimated from 197 models.

	Mean	Std. dev.	MCSE	Median	Equal-tailed [95% cred. interval]	
g	17.92258	75.00153	1.30593	8.979797	1.417264	83.33791
Shrinkage	.8717893	.1073472	.002952	.8997976	.5863091	.9881429

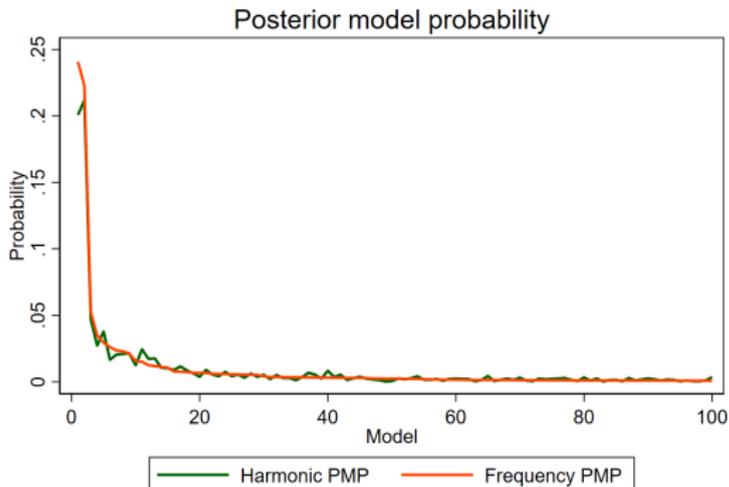
file bma\_mprior.dta saved.

. 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
```



Top 100 models shown out of 208 visited.

## 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 estimation with the model prior (`bma_mprior`) would be the best alternative.

```
. bmastats lps bmareg_enum2 bma_robust bma_mprior, compact
Log predictive-score (LPS)
Number of observations = 63
```

LPS	Mean	Minimum	Maximum
<code>bmareg_enum2</code>	.7212922	-.4892696	4.65655
<code>bma_robust</code>	.8365542	-.4436279	4.336097
<code>bma_mprior</code>	.7040244	-.4125325	3.603595

Notes: Results using analytical and frequency PMPs.  
Result `bma_mprior` 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_enum2
(results bmareg_enum2 are active now)
. bmapredict pmean,mean
note: computing analytical posterior predictive means.
```

- Use `bmacoefsample` to produce the simulated mcmc data with the mprior model.

```
. estimates restore bma_mprior
(results bma_mprior are active now)
. bmacoefsample, saving(bma_coef, replace)
Simulation (10000): ....5000....10000 done
file bma_coef.dta saved.
```

Outline

Model  
averaging

The Bayesian  
approach

BMA

bma  
commands

Example

BMA regression

PMP

PIP

Jointness

Compare OLS

CRI

Sensitivity

Predictions

References

## 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* if e(sample)
```

Variable	Obs	Mean	Std. dev.	Min	Max
dlife_exp	28	.1872143	.3772941	-.949	.852
pmean	28	.1872143	.2834054	-.6076933	.6304601
pmean_simul	28	.1875378	.2513545	-.5202287	.5831841
cri_l	28	-.3873159	.2567555	-1.148985	.0038576
cri_u	28	.7619181	.2299818	.1484607	1.137557

Outline

Model  
averagingThe Bayesian  
approach

BMA

bma  
commands

Example

BMA regression

PMP

PIP

Jointness

Compare OLS

CRI

Sensitivity

Predictions

References

## 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.

Outline

Model  
averaging

The Bayesian  
approach

BMA

bma  
commands

Example

BMA regression

PMP

PIP

Jointness

Compare OLS

CRI

Sensitivity

Predictions

References

## 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
- Sanchez G. "Introduction to Bayesian model averaging in Stata" Mexican Stata Conference, October 2023.  
[https://www.stata.com/meeting/mexico23/slides/Mexico23\\_Sanchez.pdf](https://www.stata.com/meeting/mexico23/slides/Mexico23_Sanchez.pdf)

Outline

Model  
averaging

The Bayesian  
approach

BMA

bma  
commands

Example

BMA regression

PMP

PIP

Jointness

Compare OLS

CRI

Sensitivity

Predictions

References