Bayesian model averaging

Meghan K. Cain, Ph.D.
StataCorp, LLC

STATA 18

Outline

- Motivating example
- Bayesian model selection and inference
- Bayesian model averaging
  - Theoretical introduction
  - Using `bmaregress`
  - Interpretation
  - Model fit
  - Prior selection and sensitivity
    - On the model space
    - On the regression coefficients
  - Prediction

Motivating example

```
. use houseprice
(Ames house data)

. codebook, compact

Variable        Obs Unique      Mean    Min     Max  Label
--------------------------------------------------------------------------------------------------------
saleprice       968    514  186101.8  40000  625000  Sales price, $
yrsold          968      4  2007.492   2006    2009  Year sold
lotfrontage     968    101   70.1405     21     313  Linear feet of street connected to property
lotarea         968    726  10024.68   1300  215245  Lot size in square feet
lowqualfinsf    968     14  4.887397      0     572  Low-quality finished square feet (all floors)
grlivarea       968    650  1522.224    438    3627  Above grade (ground) living area square feet
masvnrarea      962    251   110.026      0    1600  Masonry veneer area in square feet
stflrsf         968    588  1164.247    438    2524  First floor square feet
ndflrsf         968    316  353.0899      0    1818  Second floor square feet
garagearea      968    369  504.1756    160    1390  Size of garage in square feet
wooddecksf      968    201  90.82335      0     736  Wood deck area in square feet
openporchsf     968    166  47.52996      0     547  Open porch area in square feet
screenporch     968     56  16.63533      0     480  Screen porch area in square feet
enclosedpo~h    968     84  20.28719      0     552  Enclosed porch area in square feet
ssnporch        968     15   3.39876      0     508  Three-season porch area in square feet
poolarea        968      5  2.329545      0     648  Pool area in square feet
```

Bayesian linear regression

```
. bayes, rseed(92823) burnin(50000): regress saleprice soldage

Burn-in ...
Simulation ...

Model summary
------------------------------------------------------------------------------
Likelihood:
  saleprice ~ regress(xb_saleprice,{sigma2})

Priors:
  {saleprice:soldage _cons} ~ normal(0,10000)                              (1)
                   {sigma2} ~ igamma(.01,.01)
------------------------------------------------------------------------------
(1) Parameters are elements of the linear form xb_saleprice.

Bayesian linear regression                       MCMC iterations  =     60,000
Random-walk Metropolis–Hastings sampling         Burn-in          =     50,000
                                                 MCMC sample size =     10,000
                                                 Number of obs    =        968
                                                 Acceptance rate  =      .3335
                                                 Efficiency:  min =     .08277
```

```
Likelihood:
  saleprice ~ regress(xb_saleprice,{sigma2})

Priors:
  {saleprice:soldage _cons} ~ normal(0,10000)                          (1)
                   {sigma2} ~ igamma(.01,.01)
------------------------------------------------------------------------------
(1) Parameters are elements of the linear form xb_saleprice.

Bayesian linear regression                       MCMC iterations  =     60,000
Random-walk Metropolis-Hastings sampling         Burn-in          =     50,000
                                                 MCMC sample size =     10,000
                                                 Number of obs    =        968
                                                 Acceptance rate  =      .3335
                                                 Efficiency:  min =     .08277
                                                              avg =       .119
Log marginal-likelihood = -13148.484                          max =      .1516

------------------------------------------------------------------------------
             |                                                Equal-tailed
             |      Mean   Std. dev.     MCSE     Median  [95% cred. interval]
-------------+----------------------------------------------------------------
saleprice    |
     soldage |  913.4528   85.76423   2.98108   912.9563   741.7923   1085.019
       _cons |  49.67633   102.3313   2.92254   54.27965  -152.7373   257.0327
-------------+----------------------------------------------------------------
      sigma2 |  3.33e+10   1.67e+09   4.3e+07   3.33e+10   3.03e+10   3.68e+10
------------------------------------------------------------------------------
Note: Default priors are used for model parameters.
```

```
Likelihood:
  saleprice ~ regress(xb_saleprice,{sigma2})

Priors:
  {saleprice:soldage _cons} ~ normal(0,10000)                              (1)
                   {sigma2} ~ igamma(.01,.01)
------------------------------------------------------------------------------
(1) Parameters are elements of the linear form xb_saleprice.

Bayesian linear regression                       MCMC iterations  =     60,000
Random-walk Metropolis–Hastings sampling         Burn-in          =     50,000
                                                 MCMC sample size =     10,000
                                                 Number of obs    =        968
                                                 Acceptance rate  =      .3335
                                                 Efficiency:  min =     .08277
                                                              avg =       .119
Log marginal-likelihood = -13148.484                          max =      .1516
```

	Mean	Std. dev.	MCSE	Median	Equal-tailed [95% cred.	interval]
saleprice						
soldage	913.4528	85.76423	2.98108	912.9563	741.7923	1085.019
_cons	49.67633	102.3313	2.92254	54.27965	-152.7373	257.0327
sigma2	3.33e+10	1.67e+09	4.3e+07	3.33e+10	3.03e+10	3.68e+10

Note: Default priors are used for model parameters.

- Fitting model,

$$\texttt{saleprice} = \alpha + \beta\texttt{soldage} + \varepsilon$$

where $\varepsilon \sim N(0, \sigma^2)$

```
Likelihood:
  saleprice ~ regress(xb_saleprice,{sigma2})

Priors:
  {saleprice:soldage _cons} ~ normal(0,10000)                              (1)
                   {sigma2} ~ igamma(.01,.01)
------------------------------------------------------------------------------
(1) Parameters are elements of the linear form xb_saleprice.

Bayesian linear regression                       MCMC iterations  =     60,000
Random-walk Metropolis-Hastings sampling         Burn-in          =     50,000
                                                 MCMC sample size =     10,000
                                                 Number of obs    =        968
                                                 Acceptance rate  =      .3335
                                                 Efficiency:  min =     .08277
                                                              avg =       .119
Log marginal-likelihood = -13148.484                          max =      .1516
```

	Mean	Std. dev.	MCSE	Median	Equal-tailed [95% cred.	interval]
saleprice						
soldage	913.4528	85.76423	2.98108	912.9563	741.7923	1085.019
_cons	49.67633	102.3313	2.92254	54.27965	-152.7373	257.0327
sigma2	3.33e+10	1.67e+09	4.3e+07	3.33e+10	3.03e+10	3.68e+10

Note: Default priors are used for model parameters.

- Fitting model,

$$\text{saleprice} = \alpha + \beta\,\text{soldage} + \varepsilon$$

where $\varepsilon \sim N(0, \sigma^2)$

- With priors,

$$\alpha \sim N(0{,}10000)$$
$$\beta \sim N(0{,}10000)$$

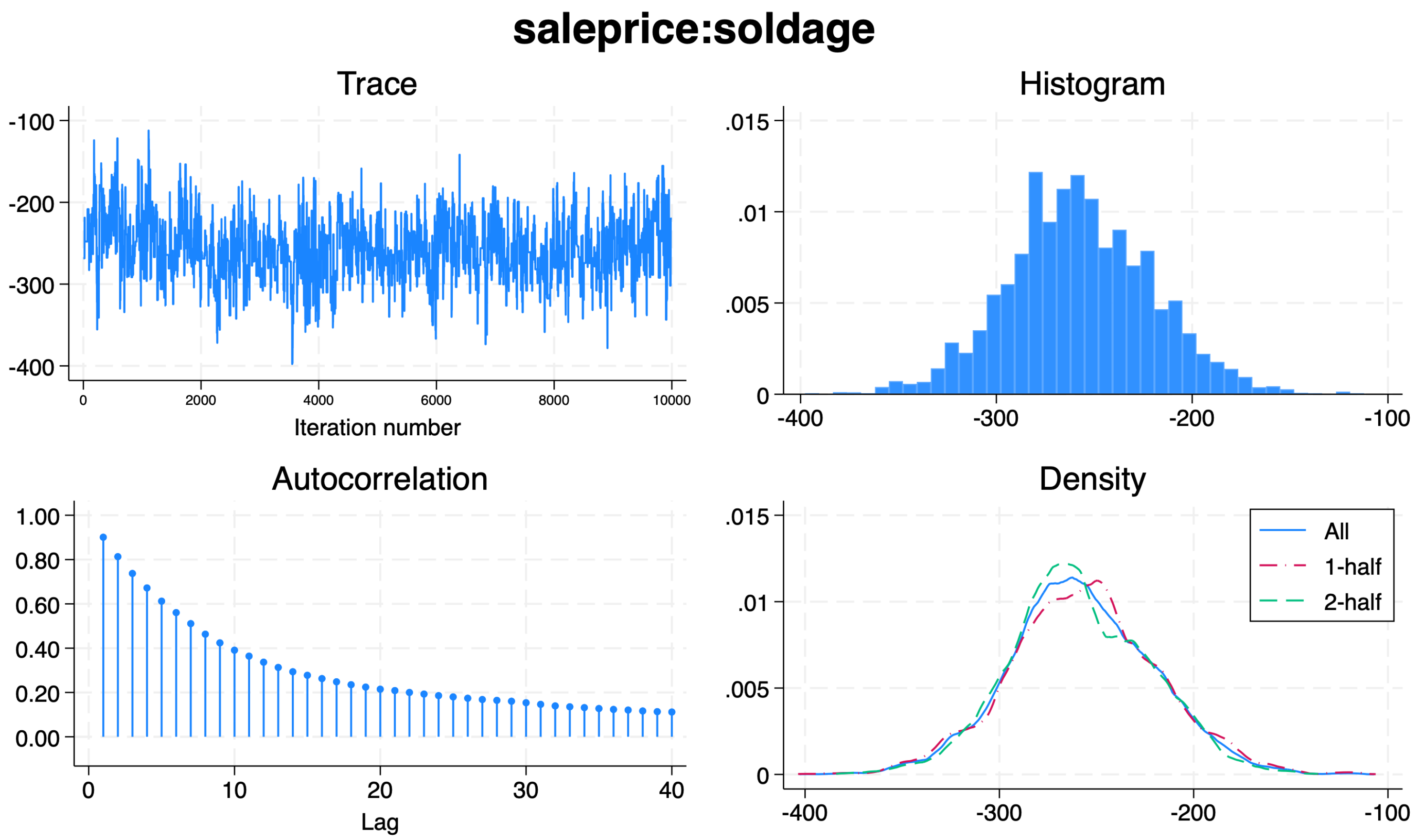
$$\sigma^2 \sim IG(0.01{,}0.01)$$

Bayesian model selection

```
. bayes, saving(m0, replace) rseed(92823) burnin(80000): regress saleprice soldage
. estimates store m0

. bayes, saving(m1, replace) rseed(92823) burnin(80000): regress saleprice soldage lotarea
  grlivarea garagearea wooddecksf totalbsmtsf
. estimates store m1

. bayes, saving(m2, replace) rseed(92823) burnin(80000): regress saleprice soldage lotarea
  grlivarea garagearea wooddecksf totalbsmtsf fullbath halfbath bedroomabvgr
. estimates store m2

. bayes, saving(m3, replace) rseed(92823) burnin(80000): regress saleprice soldage  lotarea
  grlivarea garagearea wooddecksf totalbsmtsf fullbath halfbath bedroomabvgr overallqual overallcond
  exterqual extercond kitchenqual
. estimates store m3

. bayes, saving(m4, replace) rseed(92823) burnin(80000): regress saleprice soldage  lotarea
  grlivarea garagearea wooddecksf totalbsmtsf fullbath halfbath bedroomabvgr overallqual overallcond
  exterqual extercond kitchenqual i.foundation remodage i.fireplaces lotfrontage
. estimates store m4
```

Convergence check

```
. bayesgraph diagnostic {soldage}
```



Bayesian model selection

```
. bayesstats ic m*

Bayesian information criteria
```

	DIC	log(ML)	log(BF)
m0	26201.03	-13148.54	.
m1	23133.59	-11607.01	1541.529
m2	27698.29	-15055.83	-1907.29
m3	23241.77	-13995.57	-847.0305
m4	23866.49	-12339.18	809.3622

```
Note: Marginal likelihood (ML) is computed
      using Laplace-Metropolis approximation.
```

Bayesian model selection

```
. bayestest model m1 m4

Bayesian model tests
```

	log(ML)	P(M)	P(M|y)
m1	-1.16e+04	0.5000	1.0000
m4	-1.23e+04	0.5000	0.0000

```
Note: Marginal likelihood (ML) is computed using
      Laplace-Metropolis approximation.
```

Bayesian model selection

```
. bayestest model m1 m4

Bayesian model tests
```

	log(ML)	P(M)	P(M|y)
m1	**-1.16e+04**	**0.5000**	**1.0000**
m4	**-1.23e+04**	**0.5000**	**0.0000**

```
Note: Marginal likelihood (ML) is computed using
      Laplace-Metropolis approximation.
```

$$p(M|y) = \frac{p(y|M) \times p(M)}{p(y)}$$

Model inference

```
. estimates restore m1
(results m1 are active now)

. bayesstats summary, hpd

Posterior summary statistics                      MCMC sample size =     10,000
```

	Mean	Std. dev.	MCSE	Median	HPD [95% cred.	interval]
saleprice						
soldage	-600.9391	33.75897	2.52158	-602.266	-667.0756	-535.4315
lotarea	.7926465	.1520565	.007882	.7986418	.4832124	1.068803
grlivarea	76.00627	2.722637	.154828	76.01736	70.45908	81.16987
garagearea	54.58883	7.203619	.511928	54.33465	41.2468	68.52216
wooddecksf	35.33023	10.85242	.904212	35.33743	13.68734	56.48988
totalbsmtsf	48.96077	3.410513	.307285	49.17325	41.66393	54.8153
_cons	4.630684	96.00815	5.27784	5.997862	-172.7183	199.9711
sigma2	1.39e+09	6.57e+07	1.4e+06	1.39e+09	1.26e+09	1.51e+09

Bayesian model averaging

- BMA uses Bayes theorem to account for model uncertainty by considering model, M, as a random variable.

$$p(M|D) = \frac{p(D|M) \times p(M)}{p(D)}$$

- Considers all the models in the user-defined model space.
- Rather than selecting a single model, BMA takes model uncertainty into account when making inferences and predictions.
  - Posterior model probabilities (PMPs) are used as weights to average parameter estimates across candidate models
  - Posterior inclusion probabilities (PIPs) are used to assess a variable's importance

Linear regression BMA

- We combine the likelihoods from j linear regression models

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

with the following prior distributions for models M_j and regression parameters β_j, α, and σ.

$$\begin{array}{rcl} M_j & \sim & P(M_j) \\ \boldsymbol{\beta}_j | \alpha, \sigma, M_j & \sim & N_{p_j}\{\mathbf{0}, \sigma^2 g(\mathbf{X}_j{}'\mathbf{X}_j)^{-1}\} \\ \alpha | \sigma, M_j & \propto & 1 \\ \sigma | M_j & \propto & \sigma^{-1} \end{array}$$

BMA with `bmaregress`

```
. bmaregress saleprice soldage lotarea grlivarea garagearea wooddecksf
  totalbsmtsf overallqual overallcond exterqual extercond kitchenqual fullbath halfbath
  bedroomabvgr lotfrontage i.foundation remodage i.fireplaces, rseed(92823) saving(bma, replace)

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

Bayesian model averaging                         No. of obs         =        968
Linear regression                                No. of predictors  =         20
MC3 sampling                                                 Groups =         20
                                                             Always =          0
                                                 No. of models      =        210
                                                     For CPMP >= .9 =         65
Priors:                                          Mean model size    =     13.743
  Models: Beta-binomial(1, 1)                    Burn-in            =      2,500
   Cons.: Noninformative                         MCMC sample size   =     10,000
   Coef.: Zellner's g                            Acceptance rate    =     0.1647
       g: Benchmark, g = 968                     Shrinkage, g/(1+g) =     0.9990
  sigma2: Noninformative                         Mean sigma2        =  8.381e+08

Sampling correlation = 0.9940
```

saleprice	Mean	Std. dev.	Group	PIP
soldage	-472.6591	53.18211	1	1
lotarea	.925373	.1333132	2	1
grlivarea	66.68548	4.275396	3	1
totalbsmtsf	30.70645	3.290573	6	1
overallqual	11624.31	1323.767	7	1
overallcond	8658.505	1045.051	8	1
exterqual	-12514.29	2126.038	9	1
kitchenqual	-9330.598	1540.806	11	1
bedroomabvgr	-8740.395	1733.571	14	1
garagearea	33.20419	6.724143	4	.99996
lotfrontage	158.0653	57.75354	15	.95897
wooddecksf	19.27603	12.34527	5	.79655
fullbath	-4850.452	4008.749	12	.6813
fireplaces				
2	6779.054	6368.133	20	.62012
1	1150.107	2255.75	19	.26919
extercond	117.4145	573.8413	10	.091207
halfbath	171.1511	986.8878	13	.088435
remodage	4.777644	26.4113	18	.086625
foundation				
PConc	166.0346	1047.56	17	.080274
CBlock	-81.34824	754.5213	16	.070731
Always				
_cons	13575.17	15032.19	0	1

saleprice	Mean	Std. dev.	Group	PIP
soldage	-472.6591	53.18211	1	1
lotarea	.925373	.1333132	2	1
grlivarea	66.68548	4.275396	3	1
totalbsmtsf	30.70645	3.290573	6	1
overallqual	11624.31	1323.767	7	1
overallcond	8658.505	1045.051	8	1
exterqual	-12514.29	2126.038	9	1
kitchenqual	-9330.598	1540.806	11	1
bedroomabvgr	-8740.395	1733.571	14	1
garagearea	33.20419	6.724143	4	.99996
lotfrontage	158.0653	57.75354	15	.95897
wooddecksf	19.27603	12.34527	5	.79655
fullbath	-4850.452	4008.749	12	.6813
fireplaces				
2	6779.054	6368.133	20	.62012
1	1150.107	2255.75	19	.26919
extercond	117.4145	573.8413	10	.091207
halfbath	171.1511	986.8878	13	.088435
remodage	4.777644	26.4113	18	.086625
foundation				
PConc	166.0346	1047.56	17	.080274
CBlock	-81.34824	754.5213	16	.070731
Always				
_cons	13575.17	15032.19	0	1

$$\mathrm{PIP}(X_k) = \sum_{j \in J_F} I(X_k \in M_j) P(M_j | \mathbf{y})$$

saleprice	Mean	Std. dev.	Group	PIP
soldage	-472.6591	53.18211	1	1
lotarea	.925373	.1333132	2	1
grlivarea	66.68548	4.275396	3	1
totalbsmtsf	30.70645	3.290573	6	1
overallqual	11624.31	1323.767	7	1
overallcond	8658.505	1045.051	8	1
exterqual	-12514.29	2126.038	9	1
kitchenqual	-9330.598	1540.806	11	1
bedroomabvgr	-8740.395	1733.571	14	1
garagearea	33.20419	6.724143	4	.99996
lotfrontage	158.0653	57.75354	15	.95897
wooddecksf	19.27603	12.34527	5	.79655
fullbath	-4850.452	4008.749	12	.6813
fireplaces				
2	6779.054	6368.133	20	.62012
1	1150.107	2255.75	19	.26919
extercond	117.4145	573.8413	10	.091207
halfbath	171.1511	986.8878	13	.088435
remodage	4.777644	26.4113	18	.086625
foundation				
PConc	166.0346	1047.56	17	.080274
CBlock	-81.34824	754.5213	16	.070731
Always				
_cons	13575.17	15032.19	0	1

$$\mathrm{PIP}(X_k) = \sum_{j \in J_F} I(X_k \in M_j) P(M_j | \mathbf{y})$$

$$\widehat{\boldsymbol{\beta}}_{\mathrm{BMA}} = \sum_{j=1}^{2^p} P(M_j | D) \widehat{\boldsymbol{\beta}}_{M_j}$$

Posterior samples of regression coefficients

```
. bmacoefsample, rseed(92823) saving(bmacoefs1, replace)

Simulation (10000): ....5000....10000 done

file bmacoefs1.dta saved.
```

Posterior samples of regression coefficients

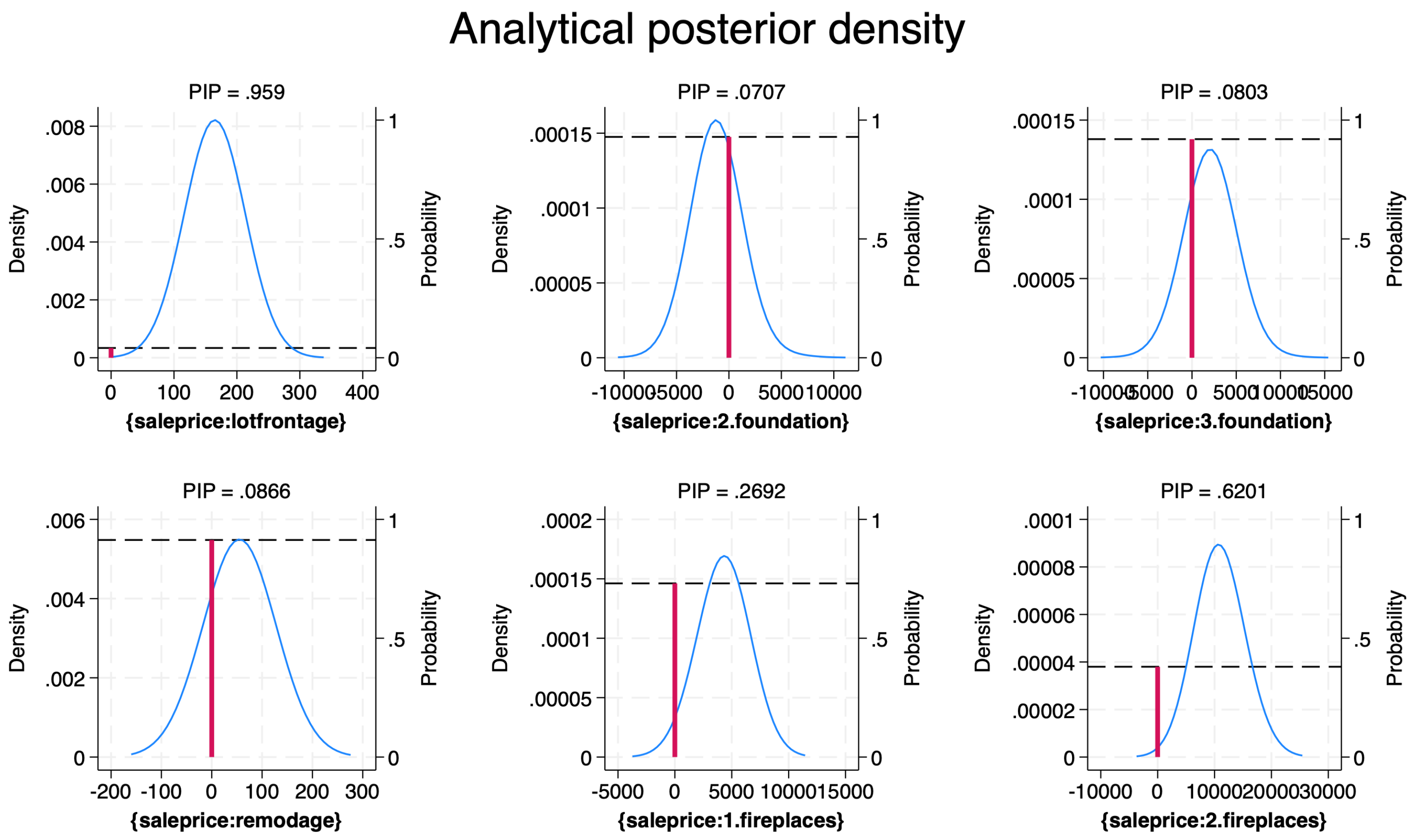
```
. bmacoefsample, rseed(92823) saving(bmacoefs1, replace)

. bayesstats summary, hpd
```

Posterior summary statistics MCMC sample size = **10,000**

	Mean	Std. dev.	MCSE	Median	HPD [95% cred.	interval]
saleprice						
soldage	-471.6422	53.33605	.518248	-472.1445	-576.7097	-367.0692
lotarea	.9250615	.1337143	.001337	.9246568	.6694403	1.194342
grlivarea	66.64104	4.257269	.042573	66.69693	58.01271	74.72243
garagearea	33.20198	6.741999	.068297	33.12693	20.03734	46.48407
wooddecksf	19.63358	12.26022	.122602	21.86233	0	37.33628
totalbsmtsf	30.71164	3.291804	.033379	30.71839	24.47351	37.43925
overallqual	11619.61	1320.846	13.2085	11619.76	9084.395	14256.42
overallcond	8645.001	1035.157	10.3516	8640.961	6575.55	10619.13
exterqual	-12542.98	2120.991	21.4282	-12559.68	-16778.84	-8490.214
extercond	123.4763	584.7643	5.84764	0	-36.88394	1790.93
kitchenqual	-9334.386	1522.694	15.2269	-9308.012	-12379.54	-6428.487
fullbath	-4875.536	4015.625	40.1562	-5413.793	-11202.11	0

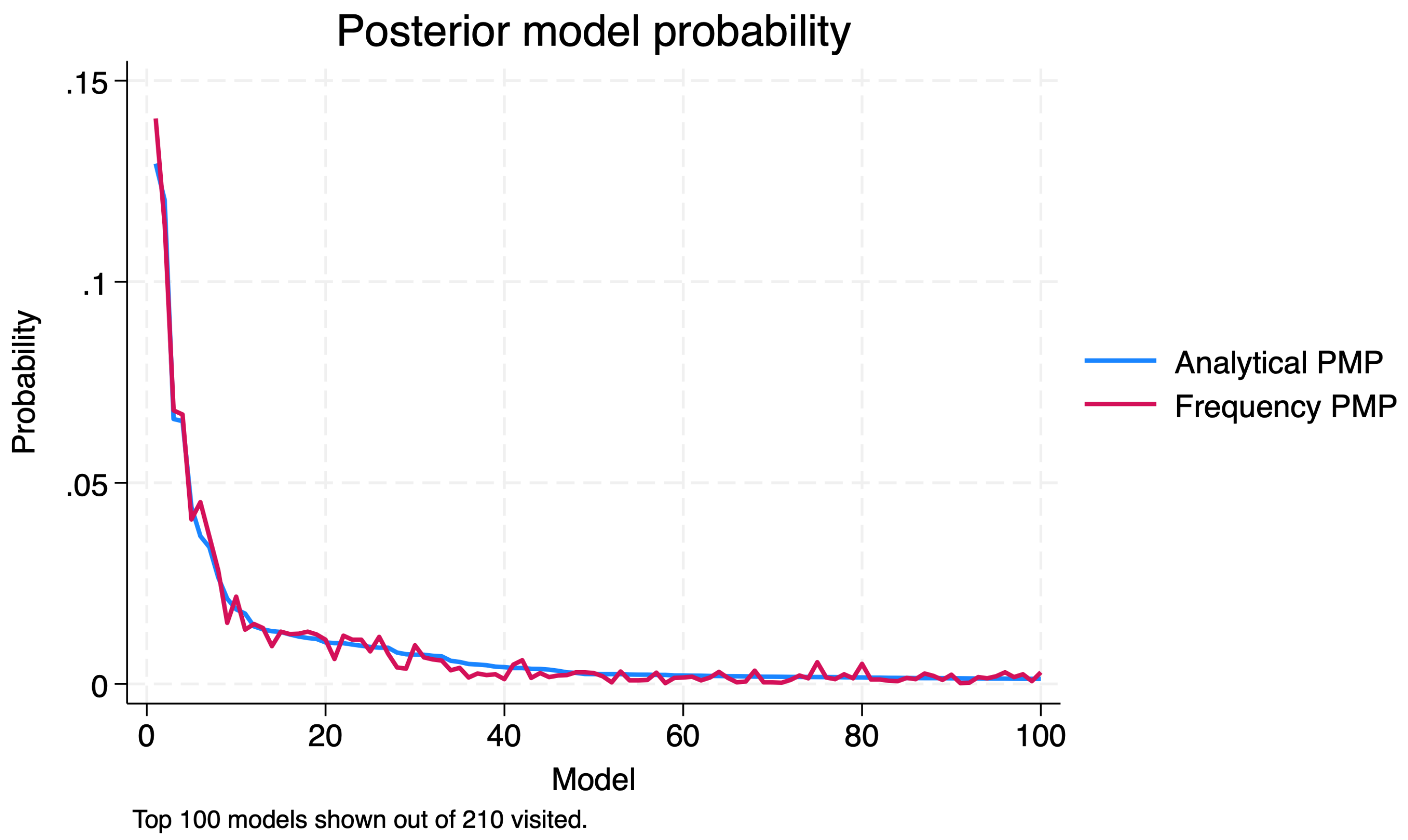
Posterior densities of regression coefficients

```
. bmagraph coefdensity lotfrontage i.foundation remodage i.fireplaces, combine
```



Convergence check

```
. bmagraph pmp
```



Model and variable-inclusion summaries

```
. bmastats models

Computing model probabilities ...

Model summary                              Number of models:
                                               Visited = 210
                                              Reported =   5
```

	Analytical PMP	Frequency PMP	Model size
Rank			
1	.1294	.1406	13
2	.1203	.1143	14
3	.06587	.068	13
4	.06534	.067	15
5	.04403	.0409	12

Note: Using analytical PMP for model ranking.

Variable-inclusion summary

1	.1294	.1406	13
2	.1203	.1143	14
3	.06587	.068	13
4	.06534	.067	15
5	.04403	.0409	12

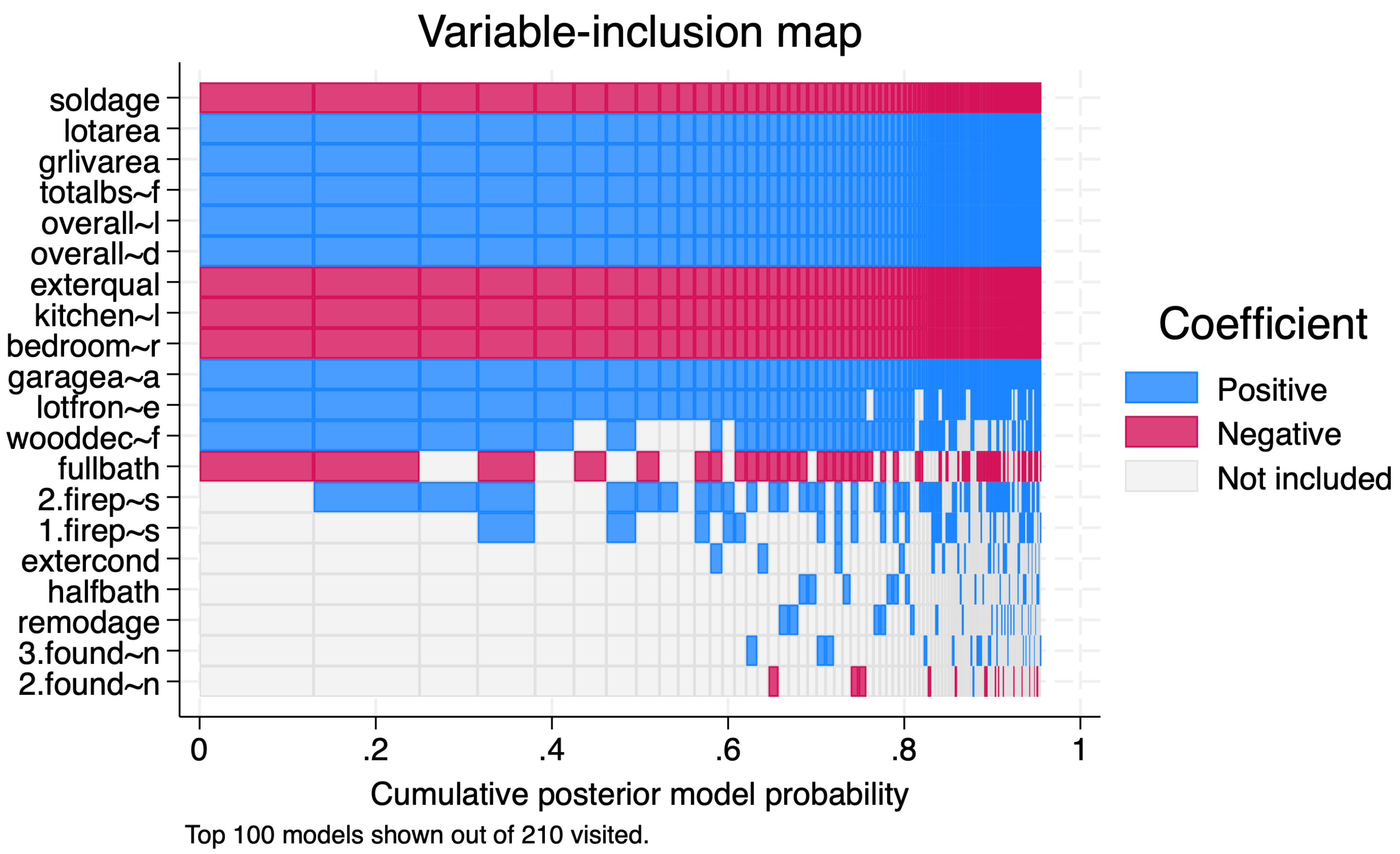
Note: Using analytical PMP for model ranking.

Variable-inclusion summary

	Rank 1	Rank 2	Rank 3	Rank 4	Rank 5
soldage	x	x	x	x	x
lotarea	x	x	x	x	x
grlivarea	x	x	x	x	x
garagearea	x	x	x	x	x
wooddecksf	x	x	x	x	x
totalbsmtsf	x	x	x	x	x
overallqual	x	x	x	x	x
overallcond	x	x	x	x	x
exterqual	x	x	x	x	x
kitchenqual	x	x	x	x	x
fullbath	x	x		x	
bedroomabvgr	x	x	x	x	x
lotfrontage	x	x	x	x	x
fireplaces					
2		x	x	x	
1				x	

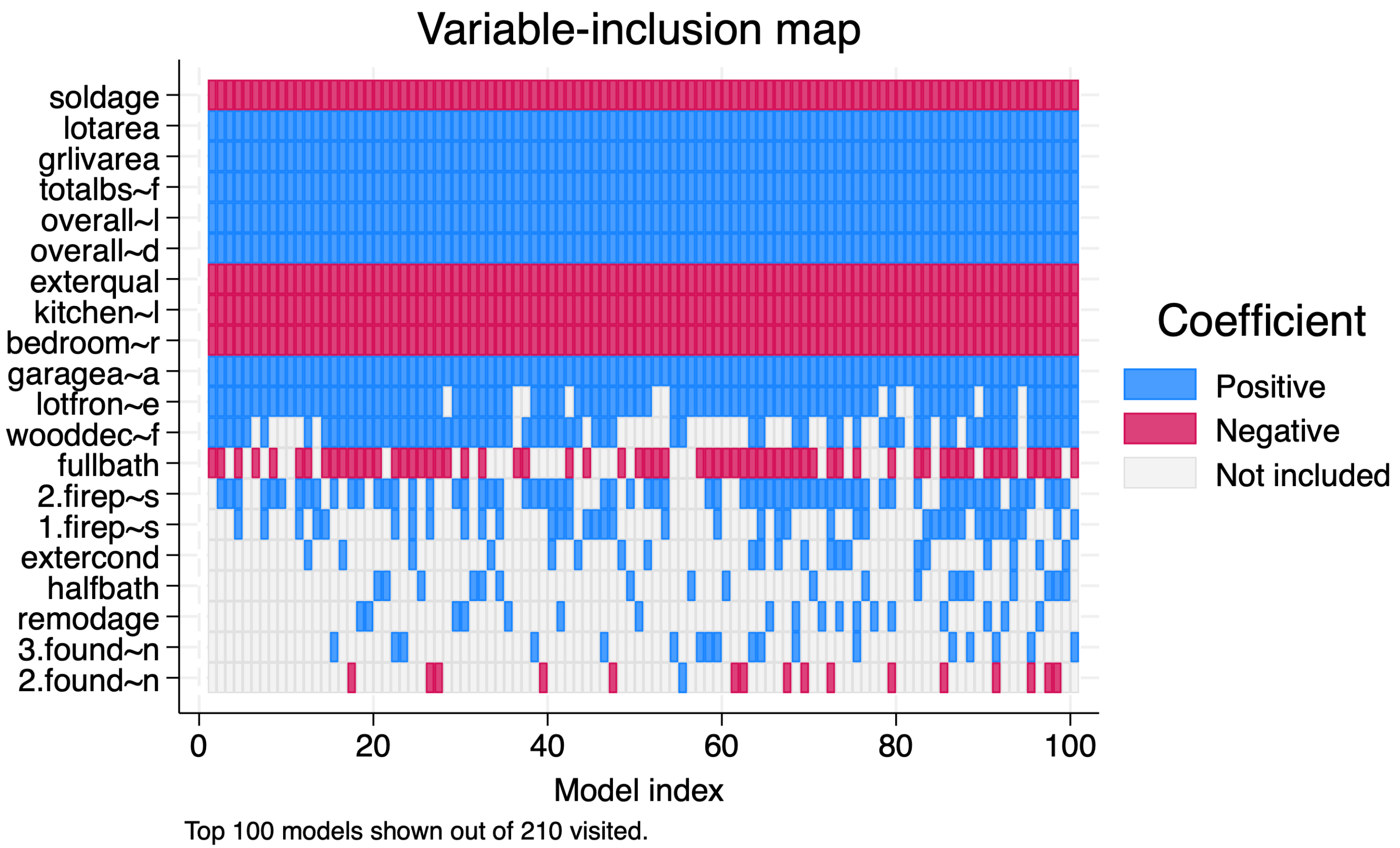
Variable-inclusion map

```
. bmagraph varmap
```



Variable-inclusion map

```
. bmagraph varmap, equalwidth
```



Jointness

```
. bmastats jointness lotfrontage extercond

Computing model probabilities ...

Variables: lotfrontage extercond
```

	Jointness
Doppelhofer–Weeks	1.320401
Ley–Steel type 1	.0938487
Ley–Steel type 2	.1035684
Yule's Q	.5784967

Notes: Using analytical PMPs. See thresholds.

Jointness

```
. bmastats jointness lotfrontage extercond

Computing model probabilities ...

Variables: lotfrontage extercond
```

	Jointness
Doppelhofer–Weeks	1.320401
Ley–Steel type 1	.0938487
Ley–Steel type 2	.1035684
Yule's Q	.5784967

Notes: Using analytical PMPs. See thresholds.

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

LS2	LS1*	Interpretation
$[0, 0.01)$	$[0, 0.01)$	Decisive disjointness
$[0.01, 0.03)$	$[0.01, 0.03)$	Very strong disjointness
$[0.03, 0.1)$	$[0.03, 0.09)$	Strong disjointness
$[0.1, 0.22)$	$[0.09, 0.18)$	Favorable disjointness
$[0.22, 3)$	$[0.18, 0.75)$	Independent inclusion
$[3, 10)$	$[0.75, 0.91)$	Favorable jointness
$[10, 30)$	$[0.91, 0.97)$	Strong jointness
$[30, 100)$	$[0.97, 0.99)$	Very strong jointness
$[100, \infty)$	$[0.99, 1]$	Decisive jointness

Jointness

```
. bmastats jointness i.fireplaces
note: 0b.fireplaces was omitted from the model.

Computing model probabilities ...

Variables: 1.fireplaces 2.fireplaces
```

	Jointness
Doppelhofer–Weeks	2.001581
Ley–Steel type 1	.368435
Ley–Steel type 2	.5833684
Yule's Q	.7619259

Notes: Using analytical PMPs. See thresholds.

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

LS2	LS1*	Interpretation
$[0, 0.01)$	$[0, 0.01)$	Decisive disjointness
$[0.01, 0.03)$	$[0.01, 0.03)$	Very strong disjointness
$[0.03, 0.1)$	$[0.03, 0.09)$	Strong disjointness
$[0.1, 0.22)$	$[0.09, 0.18)$	Favorable disjointness
$[0.22, 3)$	$[0.18, 0.75)$	Independent inclusion
$[3, 10)$	$[0.75, 0.91)$	Favorable jointness
$[10, 30)$	$[0.91, 0.97)$	Strong jointness
$[30, 100)$	$[0.97, 0.99)$	Very strong jointness
$[100, \infty)$	$[0.99, 1]$	Decisive jointness

Group factor variables

```
. bmaregress saleprice soldage lotarea grlivarea ndflrsf garagearea wooddecksf
  totalbsmtsf overallqual overallcond exterqual extercond kitchenqual fullbath halfbath ,
  bedroomabvgr lotfrontage i.foundation remodage i.fireplaces, groupfv rseed(92823)
```

saleprice	Mean	Std. dev.	Group	PIP
soldage	-472.6868	53.11433	3	1
lotarea	.9462467	.1318879	4	1
grlivarea	67.17343	4.630281	5	1
garagearea	32.60386	6.728266	7	1
totalbsmtsf	31.22253	3.561296	9	1
overallqual	11686.82	1326.367	10	1
overallcond	8652.753	1047.159	11	1
exterqual	-12392.12	2126.573	12	1
kitchenqual	-9398.937	1538.938	14	1
bedroomabvgr	-8888.69	1740.121	17	1
lotfrontage	156.5698	58.18211	18	.95616
wooddecksf	19.02557	12.48653	8	.78505
fullbath	-5220.31	4025.28	15	.71389
fireplaces				
1	1092.677	2237.681	2	.23929

saleprice	Mean	Std. dev.	Group	PIP
soldage	-472.6868	53.11433	3	1
lotarea	.9462467	.1318879	4	1
grlivarea	67.17343	4.630281	5	1
garagearea	32.60386	6.728266	7	1
totalbsmtsf	31.22253	3.561296	9	1
overallqual	11686.82	1326.367	10	1
overallcond	8652.753	1047.159	11	1
exterqual	-12392.12	2126.573	12	1
kitchenqual	-9398.937	1538.938	14	1
bedroomabvgr	-8888.69	1740.121	17	1
lotfrontage	156.5698	58.18211	18	.95616
wooddecksf	19.02557	12.48653	8	.78505
fullbath	-5220.31	4025.28	15	.71389
fireplaces				
1	1092.677	2237.681	2	.23929
2	3096.407	5920.174	2	.23929
extercond	123.9493	589.3916	13	.096236
remodage	5.479112	28.03716	19	.094324
halfbath	182.7362	1030.268	16	.093442
ndflrsf	.3365046	1.985463	6	.085715
Always				
_cons	13092.47	15076.78	0	1

Grouping predictors

```
. bmaregress saleprice soldage (lotarea grlivarea ndflrsf garagearea wooddecksf totalbsmtsf)
  (overallqual overallcond exterqual extercond kitchenqual) (fullbath halfbath bedroomabvgr)
  lotfrontage i.foundation remodage i.fireplaces
```

saleprice	Mean	Std. dev.	Group	PIP
lotarea	.9364923	.1340479	1	1
grlivarea	65.11402	6.10944	1	1
ndflrsf	3.312339	6.092468	1	1
garagearea	33.37016	6.71463	1	1
wooddecksf	24.73969	8.503941	1	1
totalbsmtsf	33.21403	5.114122	1	1
overallqual	11790.78	1308.537	2	1
overallcond	8696.484	1045.992	2	1
exterqual	-12405.65	2123.37	2	1
extercond	1258.075	1449.196	2	1
kitchenqual	-9256.343	1536.278	2	1
soldage	-470.2763	53.95602	4	1
fullbath	-6988.319	2938.175	3	.99943
halfbath	361.1639	2823.499	3	.99943
bedroomabvgr	-8821.761	1756.671	3	.99943
lotfrontage	157.6711	63.74793	5	.93126

Always-included predictors

```
. bmaregress saleprice (lotfrontage, always) soldage lotarea grlivarea ndflrsf garagearea
  wooddecksf totalbsmtsf overallqual overallcond exterqual extercond kitchenqual fullbath
  halfbath bedroomabvgr i.foundation remodage i.fireplaces, rseed(92823)
```

saleprice	Mean	Std. dev.	Group	PIP
soldage	-470.8889	53.01965	1	1
lotarea	.9236817	.1309135	2	1
grlivarea	66.51743	4.495513	3	1
garagearea	32.88755	6.687279	5	1
totalbsmtsf	30.91096	3.475457	7	1
overallqual	11642.4	1319.85	8	1
overallcond	8660.37	1042.949	9	1
exterqual	-12472.75	2126.209	10	1
kitchenqual	-9343.106	1539.618	12	1
bedroomabvgr	-8858.439	1718.42	15	1
wooddecksf	18.39017	12.78852	6	.75621
fullbath	-4496.081	4065.752	13	.62978
fireplaces				
2	6048.679	6316.6	20	.55885
1	900.1932	2049.18	19	.21374

soldage	-470.8889	53.01965	1	1
lotarea	.9236817	.1309135	2	1
grlivarea	66.51743	4.495513	3	1
garagearea	32.88755	6.687279	5	1
totalbsmtsf	30.91096	3.475457	7	1
overallqual	11642.4	1319.85	8	1
overallcond	8660.37	1042.949	9	1
exterqual	-12472.75	2126.209	10	1
kitchenqual	-9343.106	1539.618	12	1
bedroomabvgr	-8858.439	1718.42	15	1
wooddecksf	18.39017	12.78852	6	.75621
fullbath	-4496.081	4065.752	13	.62978
fireplaces				
2	6048.679	6316.6	20	.55885
1	900.1932	2049.18	19	.21374
halfbath	143.2159	898.7959	14	.069316
extercond	88.16331	499.7916	11	.067685
remodage	3.900709	23.83462	18	.067285
ndflrsf	.2505584	1.700551	4	.061588
foundation				
PConc	113.283	862.5292	17	.059177
CBlock	-57.03658	628.6965	16	.053648
Always				
lotfrontage	165.4151	48.6885	0	1
_cons	13100.51	14890.95	0	1

Priors

- Prior for model space

  `mprior()`

`betabinomial`	model size
`uniform`	model space
`binomial`	inclusion probabilities

- Prior for Zellner's g (shrinkage of regression coefficients)

  `gprior()`

`bench` (default)	`sqrtn`	`betashrink` *#1 #2*	`hypergn` *#*
`uip`	`fixed` *#*	`betabench` *#*	`zsiow`
`ric`	`ebl`	`hyperg` *#*	`robust`

Priors: Model space

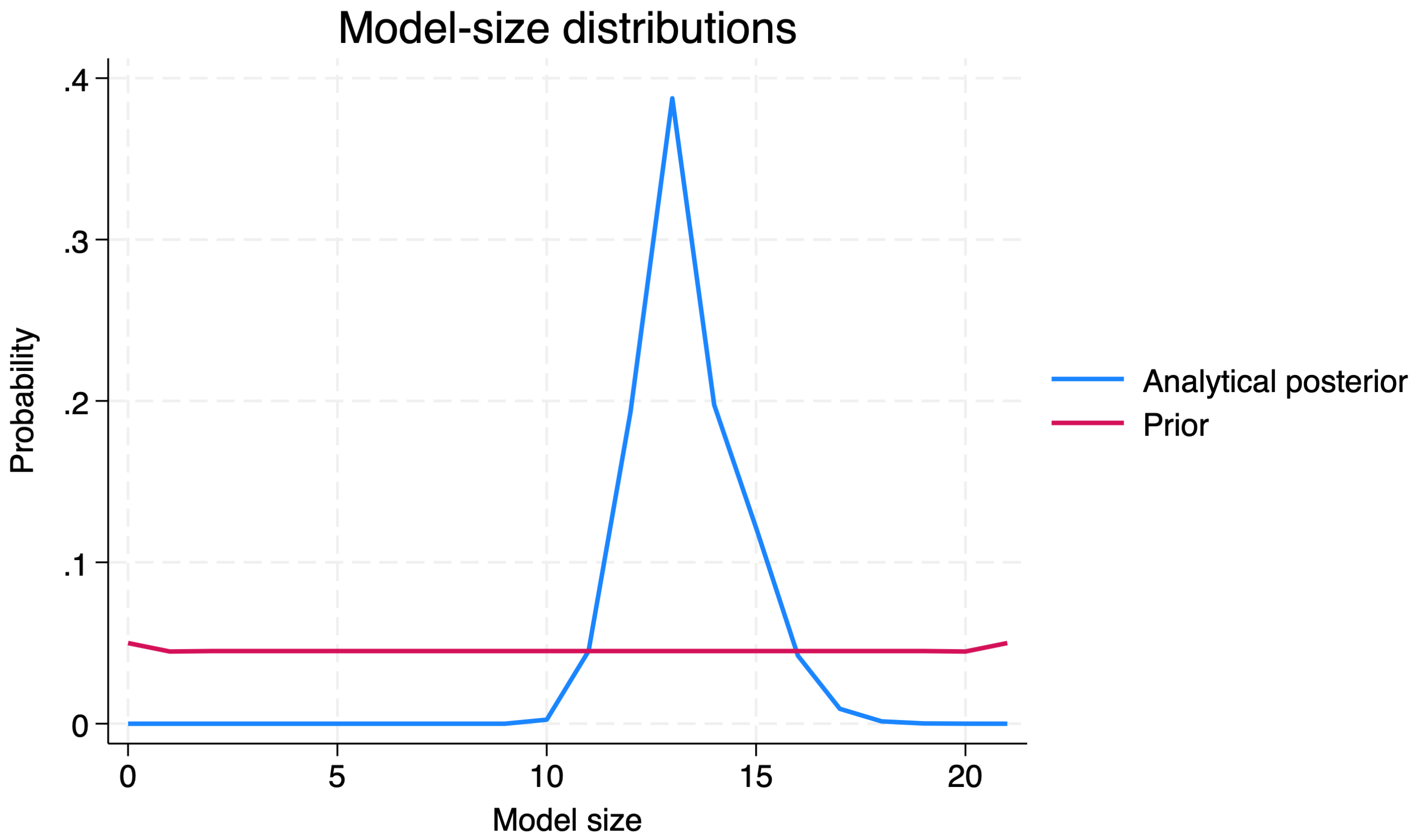
```
. bmaregress saleprice soldage lotarea grlivarea ndflrsf garagearea wooddecksf
  totalbsmtsf overallqual overallcond exterqual extercond kitchenqual fullbath halfbath
  bedroomabvgr lotfrontage i.foundation remodage i.fireplaces, groupfv enumeration

Enumerating models ...
Computing model probabilities ...

Bayesian model averaging                        No. of obs         =        968
Linear regression                               No. of predictors  =         21
Model enumeration                                           Groups =         19
                                                            Always =          0
Priors:                                         No. of models      =    524,288
  Models: Beta-binomial(1, 1)                       For CPMP >= .9 =         34
   Cons.: Noninformative                        Mean model size    =     13.322
   Coef.: Zellner's g
       g: Benchmark, g = 968                    Shrinkage, g/(1+g) =     0.9990
  sigma2: Noninformative                        Mean sigma2        = 8.400e+08
```

Priors: Model space

```
. bmaregress saleprice soldage lotarea grlivarea ndflrsf garagearea wooddecksf
  totalbsmtsf overallqual overallcond exterqual extercond kitchenqual fullbath halfbath
  bedroomabvgr lotfrontage i.foundation remodage i.fireplaces, groupfv enumeration

Enumerating models ...
Computing model probabilities ...

Bayesian model averaging                        No. of obs         =        968
Linear regression                               No. of predictors  =         21
Model enumeration                                           Groups =         19
                                                            Always =          0
Priors:                                         No. of models      =    524,288
  Models: Beta-binomial(1, 1)                       For CPMP >= .9 =         34
   Cons.: Noninformative                        Mean model size    =     13.322
   Coef.: Zellner's g
       g: Benchmark, g = 968                    Shrinkage, g/(1+g) =     0.9990
  sigma2: Noninformative                        Mean sigma2        = 8.400e+08
```

Priors: Model space

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

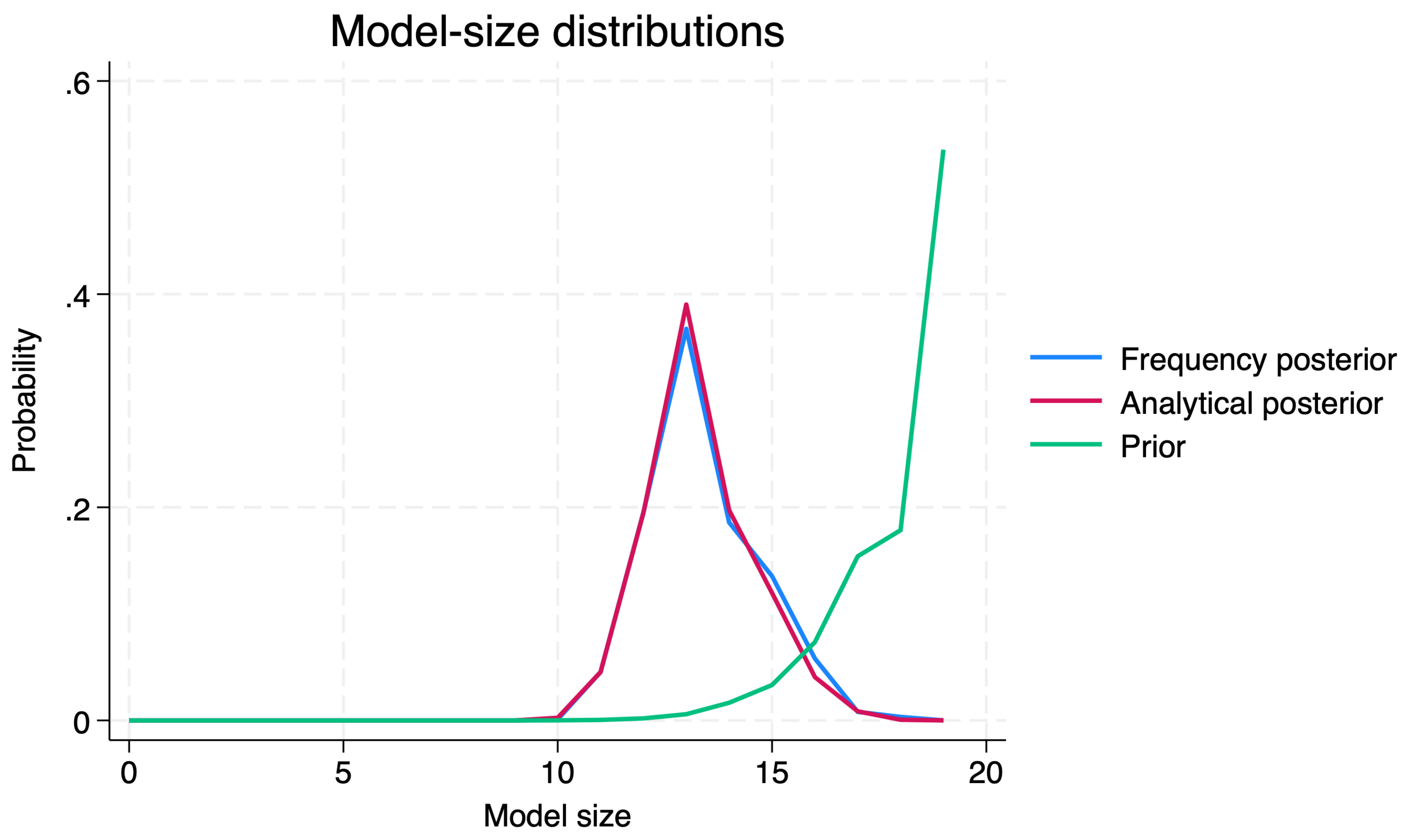


Priors: Model space

```
. bmaregress saleprice soldage lotarea grlivarea ndflrsf garagearea wooddecksf
  totalbsmtsf overallqual overallcond exterqual extercond kitchenqual fullbath halfbath
  bedroomabvgr lotfrontage i.foundation remodage i.fireplaces, groupfv rseed(92823)

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

Bayesian model averaging                         No. of obs         =        968
Linear regression                                No. of predictors  =         21
MC3 sampling                                                 Groups =         19
                                                             Always =          0
                                                 No. of models      =        105
                                                     For CPMP >= .9 =         29
Priors:                                          Mean model size    =     13.303
  Models: Beta-binomial(1, 1)                    Burn-in            =      2,500
   Cons.: Noninformative                         MCMC sample size   =     10,000
   Coef.: Zellner's g                            Acceptance rate    =     0.1235
       g: Benchmark, g = 968                     Shrinkage, g/(1+g) =     0.9990
  sigma2: Noninformative                         Mean sigma2        = 8.399e+08
```

Priors: Model space

```
. bmagraph msize
```



Priors: Sensitivity analysis on model space

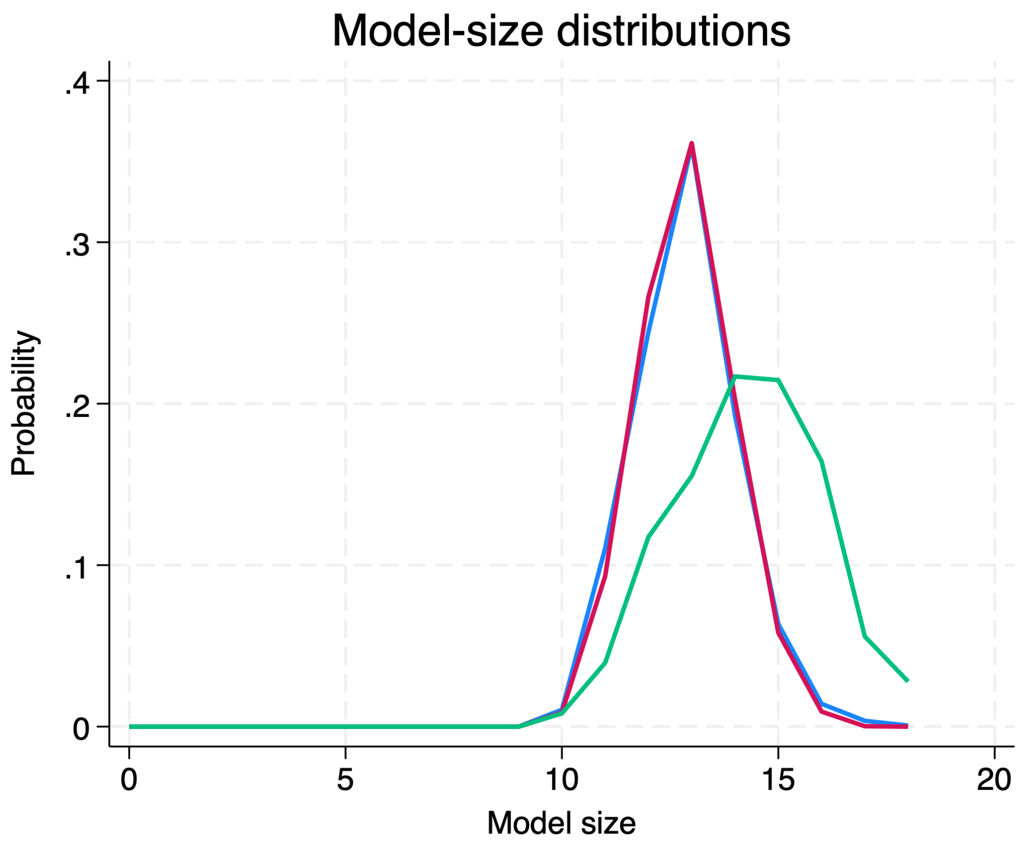
```
. bmaregress saleprice soldage lotarea grlivarea ndflrsf garagearea wooddecksf
  totalbsmtsf overallqual overallcond exterqual extercond kitchenqual fullbath halfbath
  bedroomabvgr lotfrontage i.foundation remodage i.fireplaces, rseed(92823) saving(bma1, replace)
. estimates store default
```

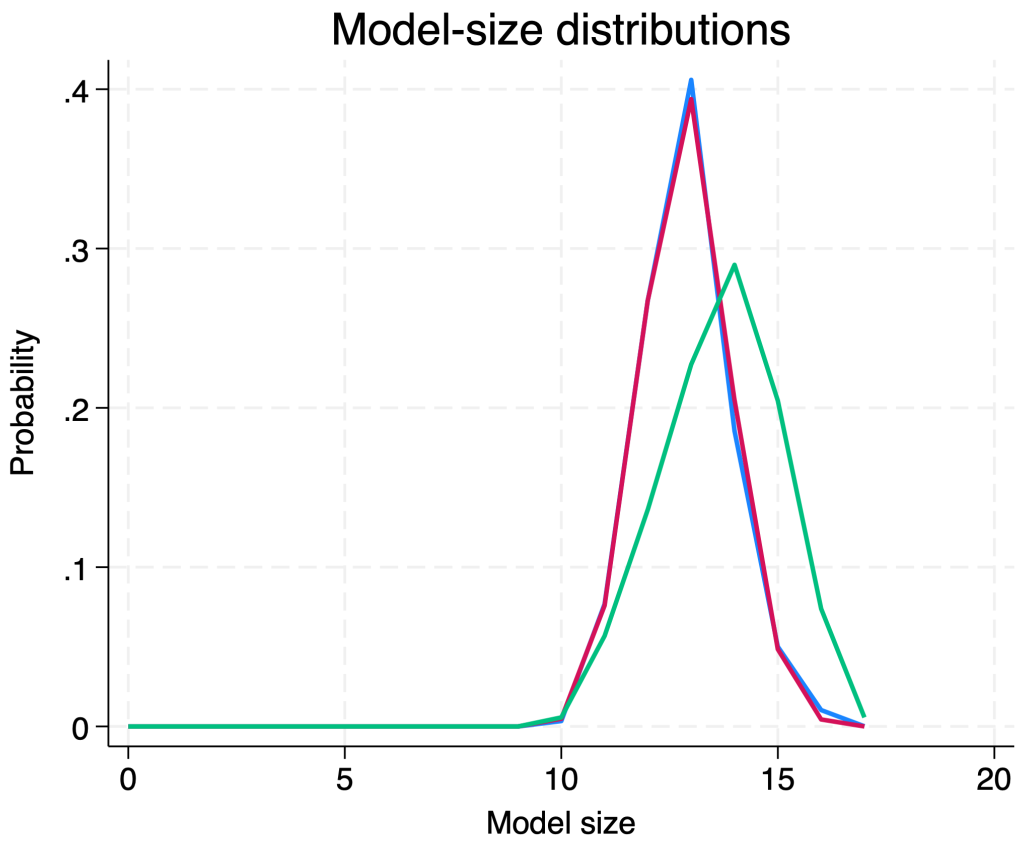
```
. bmaregress saleprice soldage lotarea grlivarea ndflrsf garagearea wooddecksf
  totalbsmtsf overallqual overallcond exterqual extercond kitchenqual fullbath halfbath
  bedroomabvgr lotfrontage i.foundation remodage i.fireplaces, rseed(92823)
  mprior(betabinomial 3) saving(bma2, replace)
. estimates store msize3
```

```
. bmaregress saleprice soldage lotarea grlivarea ndflrsf garagearea wooddecksf
  totalbsmtsf overallqual overallcond exterqual extercond kitchenqual fullbath halfbath
  bedroomabvgr lotfrontage i.foundation remodage i.fireplaces, rseed(92823)
  mprior(uniform) saving(bma3, replace)
. estimates store uniform
```

Priors: Sensitivity analysis on model space

```
. bmagraph msize
```





Priors: Sensitivity analysis on model space

- We can compare model fit of alternative priors using log predictive-scores. With model enumeration,

$$\mathrm{LPS}_{y^*} = -\log\left\{\sum_{j=1}^{2^p} P_a(M_j|\mathbf{y}) f(y^*|M_j, \mathbf{y}, \mathbf{z}_j^*)\right\}$$

```
. bmastats lps default msize3 uniform, compact

Log predictive-score (LPS)

Number of observations = 968
```

LPS	Mean	Minimum	Maximum
default	11.67967	11.1937	30.85662
msize3	11.68142	11.19583	30.52574
uniform	11.68138	11.19556	30.60314

Notes: Using analytical PMPs.
Result **default** has the smallest mean LPS.

Priors: Zellner's *g*

- Specified with option `gprior()`

- Zellner's *g* can be fixed

`bench` (default)	`sqrtn`
`uip`	`fixed` $\#$
`ric`	`ebl`

or random

`betashrink` $\#1$ $\#2$	`hypergn` $\#$
`betabench` $\#$	`zsiow`
`hyperg` $\#$	`robust`

Priors: Zellner's *g*

```
. bmaregress saleprice soldage lotarea grlivarea ndflrsf garagearea wooddecksf
  totalbsmtsf overallqual overallcond exterqual extercond kitchenqual fullbath halfbath
  bedroomabvgr lotfrontage i.foundation remodage i.fireplaces, rseed(92823) ///
  gprior(fixed 3) saving(bma4, replace)

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

Bayesian model averaging                        No. of obs         =        968
Linear regression                               No. of predictors  =         21
MC3 sampling                                                Groups =         21
                                                            Always =          0
                                                No. of models      =        455
                                                    For CPMP >= .9 =         89
Priors:                                         Mean model size    =     19.834
  Models: Beta-binomial(1, 1)                   Burn-in            =      2,500
   Cons.: Noninformative                        MCMC sample size   =     10,000
   Coef.: Zellner's g                           Acceptance rate    =     0.1538
       g: g = 3                                 Shrinkage, g/(1+g) =     0.7500
  sigma2: Noninformative                        Mean sigma2        =  2.172e+09

Sampling correlation = 0.9936
```

saleprice	Mean	Std. dev.	Group	PIP
soldage	-353.5303	94.19383	1	1
lotarea	.6786378	.1797678	2	1
grlivarea	47.34431	8.183656	3	1
totalbsmtsf	24.61012	6.916847	7	1
overallqual	8465.82	1875.939	8	1
overallcond	6776.408	1600.507	9	1
kitchenqual	-6904.605	2176.249	12	1
exterqual	-9434.649	2975.65	10	.99991
garagearea	25.84054	9.305165	5	.99898
bedroomabvgr	-6475.413	2436.221	15	.99829
lotfrontage	122.9556	70.59139	16	.9781
wooddecksf	17.25947	12.18964	6	.95564
fireplaces				
2	8935.525	6311.416	21	.95391
fullbath	-4636.856	4152.044	13	.93202
fireplaces				
1	3003.671	3167.673	20	.90785
extercond	783.9046	1907.077	11	.86306
ndflrsf	2.539905	7.84616	4	.85277
foundation				
PConc	1742.234	5957.846	18	.84994
remodage	29.53987	95.46216	19	.84919
halfbath	208.2841	3618.297	14	.84911
foundation				
CBlock	-229.8303	4852.946	17	.84547
Always				
_cons	52580.84	24986.15	0	1

Priors: Zellner's *g*

```
. bmaregress saleprice soldage lotarea grlivarea ndflrsf garagearea wooddecksf
  totalbsmtsf overallqual overallcond exterqual extercond kitchenqual fullbath halfbath
  bedroomabvgr lotfrontage i.foundation remodage i.fireplaces, rseed(92823)
  gprior(betashrink 3 1)

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

Bayesian model averaging                        No. of obs         =        968
Linear regression                               No. of predictors  =         21
MC3 and adaptive MH sampling                                Groups =         21
                                                            Always =          0
                                                No. of models      =        345
                                                   For CPMP >= .9  =        138
Priors:                                         Mean model size    =     14.450
  Models: Beta-binomial(1, 1)                   Burn-in            =      2,500
   Cons.: Noninformative                        MCMC sample size   =     10,000
   Coef.: Zellner's g                           Acceptance rate    =     0.5523
       g: Beta-shrinkage(3, 1)
  sigma2: Noninformative                        Mean sigma2        =  8.456e+08

Sampling correlation = 0.9506
```

Priors: Zellner's *g*

```
. bmaregress saleprice soldage lotarea grlivarea ndflrsf garagearea wooddecksf
  totalbsmtsf overallqual overallcond exterqual extercond kitchenqual fullbath halfbath
  bedroomabvgr lotfrontage i.foundation remodage i.fireplaces, rseed(92823)
  gprior(betashrink 3 1) saving(bma5, replace) mcmcsize(20000)

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

Bayesian model averaging                        No. of obs         =        968
Linear regression                               No. of predictors  =         21
MC3 and adaptive MH sampling                                Groups =         21
                                                            Always =          0
                                                No. of models      =        505
                                                   For CPMP >= .9  =        163
Priors:                                         Mean model size    =     14.490
  Models: Beta-binomial(1, 1)                   Burn-in            =      2,500
   Cons.: Noninformative                        MCMC sample size   =     20,000
   Coef.: Zellner's g                           Acceptance rate    =     0.5474
       g: Beta-shrinkage(3, 1)
  sigma2: Noninformative                        Mean sigma2        =  8.455e+08

Sampling correlation = 0.9831
```

saleprice	Mean	Std. dev.	Group	PIP
soldage	-472.8053	54.45996	1	1
lotarea	.9184792	.1329845	2	1
grlivarea	65.9689	4.824229	3	1
garagearea	33.39393	6.743793	5	1
totalbsmtsf	30.98743	3.75827	7	1
overallqual	11558.55	1337.382	8	1
overallcond	8676.264	1062.889	9	1
exterqual	-12519.7	2135.343	10	1
kitchenqual	-9288.661	1549.962	12	1
bedroomabvgr	-8697.967	1743.483	15	1
lotfrontage	159.4465	56.46735	16	.9669
wooddecksf	19.82252	12.07134	6	.82055
fullbath	-5142.268	3890.278	13	.73455
fireplaces				
2	7781.396	6363.75	21	.69865
1	1570.314	2512.982	20	.3657
extercond	232.0236	793.2074	11	.18295
remodage	7.986031	34.37515	19	.1541
foundation				
PConc	324.9358	1496.115	18	.1452
ndflrsf	.5765441	2.568182	4	.14105
halfbath	216.2857	1183.066	14	.1404
foundation				

fireplaces				
2	7781.396	6363.75	21	.69865
1	1570.314	2512.982	20	.3657
extercond	232.0236	793.2074	11	.18295
remodage	7.986031	34.37515	19	.1541
foundation				
PConc	324.9358	1496.115	18	.1452
ndflrsf	.5765441	2.568182	4	.14105
halfbath	216.2857	1183.066	14	.1404
foundation				
CBlock	-146.0076	1113.551	17	.14
Always				
_cons	13574.75	15521.9	0	1

Note: Coefficient posterior means and std. dev. estimated from 505 models.
Note: Default prior is used for models.

	Mean	Std. dev.	MCSE	Median	Equal-tailed [95% cred.	interval]
g	427.2365	180.3464	3.18699	387.3428	203.5504	887.3675
Shrinkage	.9973139	.0009701	.000018	.997425	.9951112	.9988743

Priors: Sensitivity analysis on shrinkage

```
. bmaregress saleprice soldage lotarea grlivarea ndflrsf garagearea wooddecksf
  totalbsmtsf overallqual overallcond exterqual extercond kitchenqual fullbath halfbath
  bedroomabvgr lotfrontage i.foundation remodage i.fireplaces, rseed(92823)
  gprior(fixed 3) saving(bma4, replace)
. estimates store fixed3

. bmaregress saleprice soldage lotarea grlivarea ndflrsf garagearea wooddecksf
  totalbsmtsf overallqual overallcond exterqual extercond kitchenqual fullbath halfbath
  bedroomabvgr lotfrontage i.foundation remodage i.fireplaces, rseed(92823)
  gprior(betashrink 3 1) saving(bma5, replace) mcmcsize(20000)
. estimates store beta31

. bmaregress saleprice soldage lotarea grlivarea ndflrsf garagearea wooddecksf
  totalbsmtsf overallqual overallcond exterqual extercond kitchenqual fullbath halfbath
  bedroomabvgr lotfrontage i.foundation remodage i.fireplaces, rseed(92823)
  gprior(robust) saving(bma6, replace)
. estimates store robust
```

Priors: Sensitivity analysis on shrinkage

```
. bmastats lps default fixed3 beta31 robust, compact

Log predictive-score (LPS)

Number of observations = 968
```

LPS	Mean	Minimum	Maximum
default	11.67967	11.1937	30.85662
fixed3	11.93608	11.66977	25.30779
beta31	11.67858	11.19884	30.67606
robust	11.6788	11.19851	30.6364

```
Notes: Results using analytical and frequency PMPs.
       Result beta31 has the smallest mean LPS.
```

Predictions: Posterior samples of coefficients

```
. bmacoefsample, rseed(92823) saving(bmacoefs2, replace)

Simulation (10000): ....5000....10000 done

file bmacoefs2.dta saved.
```

Predictions: New dataset

```
. use houseprice2010
(Ames house data in 2010)

. codebook, compact

Variable       Obs Unique      Mean    Min     Max  Label
----------------------------------------------------------------------------------------------------
saleprice      132    118  183990.1  55993  611657  Sales price, $
yrsold         132      1      2010   2010    2010  Year sold
lotfrontage    132     58  72.32576     21     152  Linear feet of street connected to property
lotarea        132    119  9894.197   1680   53504  Lot size in square feet
lowqualfinsf   132      2  2.954545      0     390  Low-quality finished square feet (all floors)
grlivarea      132    126  1508.909    720    3279  Above grade (ground) living area square feet
masvnrarea     132     52  89.39394      0     760  Masonry veneer area in square feet
stflrsf        132    123  1182.886    525    2364  First floor square feet
ndflrsf        132     51  323.0682      0    1589  Second floor square feet
garagearea     132     91  485.2045    200     923  Size of garage in square feet
wooddecksf     132     51  100.4545      0     668  Wood deck area in square feet
openporchsf    132     58  38.63636      0     262  Open porch area in square feet
screenporch    132     12  14.74242      0     385  Screen porch area in square feet
enclosedpo~h   132     26  29.09091      0     242  Enclosed porch area in square feet
ssnporch       132      2  1.090909      0     144  Three-season porch area in square feet
```

Predictions: Coverage

```
. bmapredict CrI_l CrI_u, cri hpd
note: computing credible intervals using simulation.

Computing predictions ...

. generate coverage = saleprice > CrI_l & saleprice < CrI_u

. summarize coverage
```

Variable	Obs	Mean	Std. dev.	Min	Max
coverage	132	.9469697	.2249476	0	1

Predictions: LPS and MSE

```
. bmapredict lps, lps
. bmapredict pmean, mean
note: computing posterior predictive means using simulation; option mcmcsample implied.

Computing predictions ...

. generate mse = (saleprice-pmean)^2
```

STATA® 18