

BAYESIAN MODEL AVERAGING (BMA)

IHS EVIEWS

March 2012

Overview

The Bayesian model averaging (BMA) technique addresses the fact that inference using a regression model is typically conducted without accounting for variable selection carried out prior to estimating the model. Ignoring the uncertainty associated with variable selection may yield misleading inference.

BMA approaches the variable selection problem by averaging quantities of interest over all possible models. BMA estimates the unknown quantities of interest under each competing model (with different sets of explanatory variables), and then computes the weighted average of the quantities. The weights are based on a Bayesian estimate of how likely it is that the candidate model is the “correct” model.

This EViews package provides a front end to the R packages for conducting BMA.¹ The EViews package produces the posterior means and standard deviations of coefficients of interest, along with the posterior probabilities (in %) of the variables being in the model.

The package has two add-ins: the **bma** add-in can be applied to linear or generalized linear models, and the **bmamlogit** add-in is designed for multinomial logit models. These add-ins require R and the BMA and/or mlogitBMA packages.

¹The R version of BMA (written by Adrian Raftery, Jennifer Hoeting, Chris Volinsky, Ian Painter, and Ka Yee Yeung) and mlogitBMA (written by Hana Sevcikova and Adrian Raftery) are available at <http://cran.r-project.org/web/packages/BMA/index.html> and <http://cran.r-project.org/web/packages/mlogitBMA/index.html>, respectively.

Bayesian Model Averaging

Suppose δ is a quantity of interest. Then the BMA posterior distribution of δ given the data D is given by

$$\Pr(\delta|D) = \sum_{k=1}^K \Pr(\delta|M_k, D) \Pr(M_k|D). \quad (1)$$

which is a weighted average of the posterior distributions $\Pr(\delta|M_k, D)$ under each model M_k ($k = 1, \dots, K$), where the weights are the posterior model probabilities $\Pr(M_k|D)$ that M_k is the “correct” model. It is worth noting that BMA requires posterior model probabilities for use as weights.

We are interested in the posterior means and variances, $E[\delta|M_k, D]$ and $\text{Var}[\delta|M_k, D]$. Using Equation (1), we may derive expressions for the posterior mean and variance:

$$\begin{aligned} E[\delta|D] &= \int \delta \Pr(\delta|D) d\delta \\ &= \sum_{k=1}^K \left(\int \delta \Pr(\delta|M_k, D) d\delta \right) \Pr(M_k|D) \\ &= \sum_{k=1}^K E[\delta|M_k, D] \Pr(M_k|D) \\ \text{Var}[\delta|D] &= \sum_{k=1}^K (\text{Var}[\delta|M_k, D] + E[\delta|D]^2) - E[\delta|D]^2. \end{aligned}$$

where $E[\delta|M_k, D]$ and $\text{Var}[\delta|M_k, D]$ are the posterior mean and variance which are derived from the model M_k by the Bayes rule.

The posterior model probability for model M_k is

$$\Pr(M_k|D) = \frac{\Pr(D|M_k) \Pr(M_k)}{\sum_{l=1}^K \Pr(D|M_l) \Pr(M_l)} \quad (2)$$

where

$$\Pr(D|M_k) = \int L(D|\theta_k, M_k) \Pr(\theta_k|M_k) d\theta_k,$$

is the marginal likelihood under model M_k , θ_k is the vector of parameters under model M_k , $\Pr(\theta_k|M_k)$ is the prior density of θ_k under model M_k , and $L(D|\theta_k, M_k)$ is

the standard likelihood, and $\Pr(M_k)$ is the prior probability that M_k is the correct model.

It is worth noting that the **BMA** package produces posterior model probabilities and associated Bayesian information criteria (BIC) values. Both posterior model probabilities and BICs may be used to find a “preferred” model out of the competing models; given a set of competing models for the data, the preferred model is the one with the lowest BIC and/or the highest posterior model probability.

However, it can be shown that averaging over all of the models, as measured by a logarithmic score rule, can provide better predictive ability than using a single model. Given that averaging over candidate models may be thought of as a “soft” form of model selection, BMA is particularly attractive when we wish avoid choosing a particular model.

One of the stumbling blocks in implementing BMA is to incorporate every possible uncertainty into the model because the number of models can be enormous. For instance, if we have 10 explanatory variables in a linear model setting, BMA needs to combine $2^{10} = 1024$ different models $\{M_1, \dots, M_{1024}\}$. A solution to this problem involves applying the Occam’s window algorithm which averages only over a set of likely models. Denote a set of likely models and unlikely models to be R and Q , then the Occam’s algorithm:

1. Identifies a largest marginal score: $\Pr(M_k|D) = \operatorname{argmax}_j \Pr(M_k|D)$.
2. Given the value C , excludes a set of models $Q_1 = \{M_j; \Pr(M_k|D)/\Pr(M_j|D)\} \geq C$ and get the likely models $R_1 = \{M_j; M_j \notin Q_1\}$.
3. For $M_j \in R_1$, excludes a set of models $Q_2 = \{M_j; M_l \subset M_j, M_l, M_j \in R_1\}$ and derive a set of likely models $R_2 = \{M_j; M_j \in R_1, M_j \in Q_2\}$.

Thus Equation (1) is replaced by

$$\Pr(\delta|D) = \frac{\sum_{M_k \in R_2} \Pr(\delta|M_k, D) \Pr(M_k|D) \Pr(M_k)}{\sum_{M_k \in R_2} \Pr(M_k|D) \Pr(M_k)}$$

and therefore the number of models in the sum can be greatly reduced.

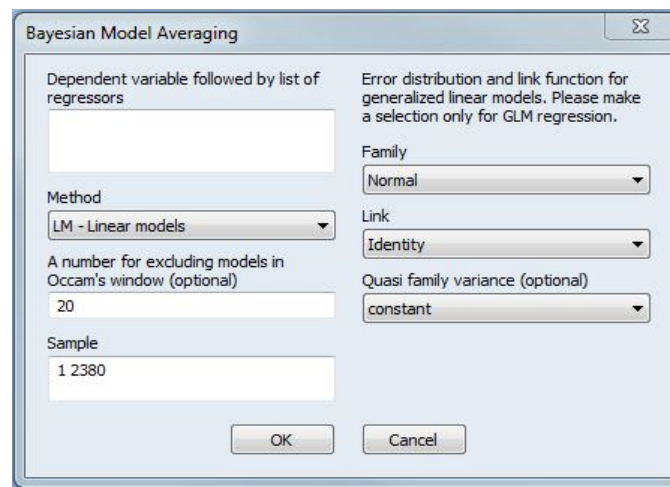
BMA for (Generalized) Linear models

Package Name: bma

Default Proc Name: bma

Default Menu Text: : BMA for GLM/LM

Dialog: Upon running the add-in from the menus, a dialog will appear:



The dialog box is titled "Bayesian Model Averaging" and contains several input fields and dropdown menus. The first field is for the "Dependent variable followed by list of regressors". Below it is a "Method" dropdown menu set to "LM - Linear models". To the right, there is a section for "Error distribution and link function for generalized linear models" with a "Family" dropdown set to "Normal", a "Link" dropdown set to "Identity", and a "Quasi family variance (optional)" dropdown set to "constant". Below the "Method" dropdown is a field for "A number for excluding models in Occam's window (optional)" set to "20". At the bottom left is a "Sample" field with the value "1 2380". "OK" and "Cancel" buttons are at the bottom right.

In the first box, you should specify the dependent variable followed by list of regressors. Please note that an explicit equation specification (e.g. $y=c(1)+c(2)*x$) is not allowed in this box.

The “Method” combo box lets you select the regression model of interest - particularly, linear model (LM) or generalized linear model (GLM).

The “A number for excluding models in Occam’s window” box lets you specify the value C in the Occam’s algorithm (Step 2). The number of models in the Occam’s algorithm increases as the value C decreases (the default value for C is 20).

When you choose the GLM regression, the “Family” and “Link” combos let you specify the family distribution of errors and the link function. Table 1 shows available options to users. Since the variance function is determined by the family, all families other than the *quasi* family do not require the variance specification. When the *quasi* family is chosen, the “Variance” box lets you specify the variance function.

Table 1: Family and Link functions

Family	Link	Variance
Normal	identity	Normal
Binomial	logit, probit and cloglog	Binomial
Poisson	log, identity and sqrt	Poisson
Gamma	inverse, identity and log	Gamma
Inverse-Gaussian	$1/\mu^2$	Inverse-Gaussian
Quasi	User-specified	Constant, $\mu(1 - \mu)$, μ , μ^2 , μ^3

Note: The quasi family accepts the logit, probit, cloglog, identity, inverse, log, $1/\mu^2$ and sqrt links.

Command line:

Syntax: bma(options) a list of variables

Options	Explanation
method=arg	Linear models (LM, default), Generalized linear models (GLM)
family=arg	Distribution family: Normal (NORMAL), Binomial (BINOMIAL), Poisson (POISSON), Gamma (GAMMA), Inverse Gaussian (IGAUSS), Quasi (QUASI)
link=arg	Link function: Identity (IDENTITY), Log (LOG), Logit (LOGIT), Probit (PROBIT), Complementary Log-log (CLOGLOG), Inverse (INVERSE), Squared-root (SQRT), $1/\mu^2$ (BOX)
var=arg	Variance for the quasi family: Constant (CON), $\mu(1 - \mu)$ (BIN), μ (MU), μ^2 (MU2), μ^3 (MU3)
smpl=arg	Specify the sample

Example: To illustrate how the add-in can be used for Bayesian model averaging, we use the “401kjae” dataset (see User Guide II, “Generalized Linear Models” on page 313); this is available in your EViews installation directory (Usually, C:\Program Files\EViews 7\Example Files\EV7 Manual Data\Chapter 27 - Generalized Linear Models).

```
bma(method=GLM, family=BINOMIAL, link=LOGIT) mrate log(totemp)
    log(totemp)2 age age2 sole
```

will employ a GLM specification using a binomial proportion family and logit link and proceed the BMA regression. It will create an EViews spool (see Figure 1) called *bma*, with two tables, *bma* and *models*; the first table contains the averaged posterior means and standard deviations and posterior probabilities that the variable is in the model (in %). The second table contains the posterior values under each model, along with BICs and posterior model probabilities.

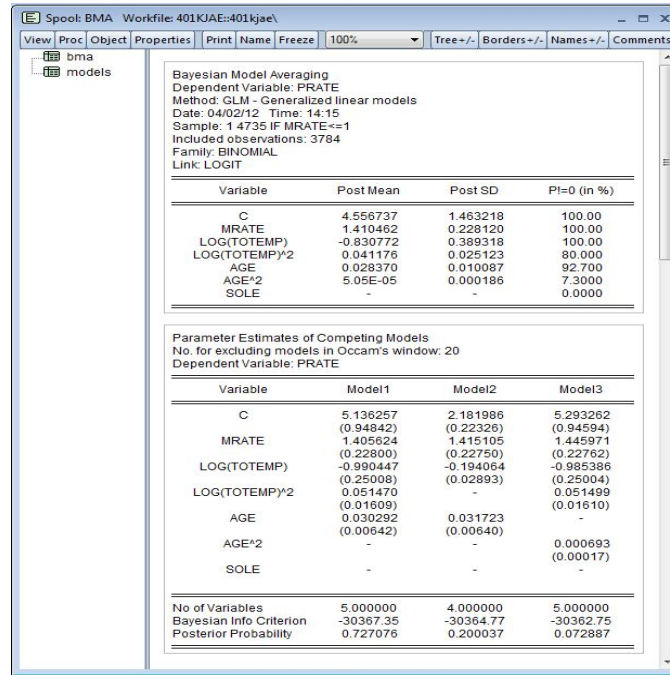


Figure 1: 401K tax advantaged savings plan participation rates (PRATE)

From the “bma” table, it is clear that the employer matching contribution rate (MRATE), the log of total employment (LOG(TOTEMP), LOG(TOTEMP)²), plan age (AGE), and the constant (C) have high posterior probabilities of being in the model, while AGE² and a binary indicator for whether the plan is the only pension plan offered by the plan sponsor (SOLE) have low posterior probabilities. The SOLE variable, especially, has 0 probability to be in the model and so a great deal of uncertainty about whether it should be included remains.

The “models” table shows that the models favored by BMA do not include the AGE² variable.

BMA for Multinomial Logit models

Package Name: bma

Default Proc Name: bnamlogit

Default Menu Text: : BMA for Multinomial logit models

Description: When analyzing a categorical response, multinomial logistic (MNL) regression is one often-used strategy. Given that N individuals make one choice among J alternatives, the probability of individual i choosing alternative j is

$$P_{ij} = \frac{\exp(V_{ij})}{\sum_{j=1}^J \exp(V_{ij})}, \quad i = 1, \dots, N, \quad j = 1, \dots, J$$
$$V_{ij} = c_j + \alpha_j X_i + \beta_j W_{ij}$$

where c_j is the alternative-specific intercepts, $\{X_i\}$ is a set of individual-specific explanatory variables and $\{W_{ij}\}$ is a set of alternative-specific explanatory variables.

The choice for one alternative is modeled based on the differences between alternatives, and so we set a base alternative and treat the remaining alternatives as differences from the base. For example, if the base alternative is 1 (i.e. $V_{i1} = 0$ for all i), then we will be interested in the difference between the satisfaction of two different alternatives j and 1,

$$V_{ij} = c_j + \alpha_j X_i + \beta_j (W_{ij} - W_{i1}) \quad j = 2, \dots, J.$$

Dialog: Upon running the add-in from the menus, a dialog will appear (see Figure 2).

The first box in the dialog lets you specify the categorically distributed response variable which is the choice made by the N individuals. The following two boxes let you specify the individual-specific regressors X_i and the alternative-specific regressors W_{ij} .

When you would like to set the base alternative (default is 1), you need to specify the index of the alternative of interest in the “Index of base alternative” box. Since the names of the alternatives will be automatically determined by R based on the structure of the response variable, you should be careful to choose the index of the base alternative (e.g. if the names of the alternatives are assigned to be (1 3 2) by R, then the index “2” denotes the alternative 3).

Figure 2: Dialog of BMA for multinomial logit

The “Include alternative-specific intercepts” checkbox lets you decide whether alternative specific constants c_j should be included in the selected models.

The “Estimation method” combo asks you to choose the estimation method in the model averaging procedure. It is worth noting that the Begg and Gray approximation (Begg and Gray, 1984) provides a close approximation to maximum-likelihood estimation of the full MNL, and is particularly attractive due to its computational efficiency when the response variable has multiclass classification (Yeung *et al.*, 2005).

Command line:

Syntax: `bmamlogit(options) response individual_specific_var @ alternative_specific_var`

Options:

	Explanation
<code>base=int</code>	The index of the base alternative (The default choice is 1)
<code>mnl</code>	Perform maximum likelihood estimation (The default method is the Begg and Gray approximation)
<code>exc</code>	Exclude alternative-specific constants
<code>smpl=arg</code>	Specify the sample

Example: By using the Boston HMDA dataset², we will illustrate how the add-in may be used for Bayesian model averaging for multinomial logit models. In the following command,

bmamlogit s7 s11 s13 s15 s17 @ s42

the variable *s7* is a dependent variable, *s11-s17* are individual specific variables and *s42* is an alternative specific variable. It will create an EViews spool (“bmamlogit”) which contains three tables, *multinomial_res*, *bma*, and *models*.

Figure 3 shows the *multinomial_res* table which displays information on the categorical dependent variable *s7*. Note that the table contains the names (i.e. 1 3 2) and frequencies (i.e. 2025 285 70) of the alternatives, along with information on the base alternative selection.

Dependent Variable: S7			
Base choice name: 1			
Base choice index: 1			
Alternative	1	3	2
Frequency	2025	285	70

Figure 3: *multinomial_res*

Figure 4 displays the main estimation results; the BMA posterior means, standard deviations and posterior probabilities (see Table 4(a)), and the posterior values under each model, along with BICs and posterior model probabilities (see Table 4(b)). It is shown that the individual-specific variable *s13*, the alternative-specific constants *s42*₃ and *s42*₂, and the constant *C* have high posterior probabilities of being in the model, while the remaining individual-specific variables *s11*, *s15*, and *s42* have low posterior probabilities. Especially, the probabilities that *s42*₂ and *s15* to be in the model are 0 so that uncertainty about whether they should be included still remains. The two variables *s11* and *s17* have relatively noticeable probabilities, however, the BMA favored Model 1 does not have these variables.

²Available at <http://wps.aw.com/wps/media/objects/3254/3332253/datasets2e/datasets/hmda.docx>

Bayesian Model Averaging			
Dependent Variable: S7			
Method: Begg and Gray approximation			
Date: 04/02/12 Time: 16:04			
Sample: 1 2380			
Included observations: 2380			
Variable	Post Mean	Post SD	PI=0 (in %)
C	-0.478596	0.421232	100.00
C(2)	-0.887210	0.305317	100.00
S42(3)	0.543111	0.122584	100.00
S42(2)	-	-	0.0000
S11	0.052713	0.131917	16.900
S13	-0.538295	0.071986	100.00
S15	-	-	0.0000
S17	3.92E-07	5.54E-07	37.600

(a) *bma*

Parameter Estimates for Competing Models				
Dependent Variable: S7				
Variable	Model1	Model2	Model3	Model4
C	-0.416077 (0.38673)	-0.407257 (0.38733)	-0.803919 (0.43276)	-0.801898 (0.43285)
C(2)	-0.824066 (0.26548)	-0.812185 (0.26588)	-1.221947 (0.26674)	-1.218104 (0.26722)
S42(3)	0.543649 (0.12235)	0.545029 (0.12258)	0.537668 (0.12306)	0.538315 (0.12334)
S42(2)	-	-	-	-
S11	-	-	0.309673 (0.14806)	0.317351 (0.14809)
S13	-0.547984 (0.06494)	-0.554808 (0.06511)	-0.475820 (0.07390)	-0.481336 (0.07394)
S15	-	-	-	-
S17	-	1.04E-06 (0.00000)	-	1.06E-06 (0.00000)
No of Variables	4.000000	5.000000	5.000000	6.000000
Bayesian Info Criterion	2932.397	2933.442	2935.652	2936.530
Posterior Probability	0.521921	0.309497	0.102493	0.066089

(b) *models*

Figure 4: *bma* and *models*

References

- Begg, C.B., Gray, R. (1984) Calculation of polychotomous logistic regression parameters using individualized regressions. *Biometrika* 71, 11–18.
- Raftery, A.E., Madigan, D., Hoeting, J.A. (1997) Bayesian model averaging for linear regression models. *Journal of the American Statistical Association* 92, 437, 179–191.
- Yeung, K.Y., Bumgarner, R.E., Raftery, A.E. (2005) Bayesian model averaging: development of an improved multiclass, gene selection and classification tool for microarray data. *Bioinformatics* 21, 2394–2402.