OVERCOMING THE COMPLEXITY OF 2^k IN BAYESIAN MODEL AVERAGING

O. O. John; D. F. Adiele and J. C. Nwabueze

Abstract

A new method SWR-BMA which involves the sampling without replacement from the predictor variables has been proposed in Bayesian Model Averaging analysis to confront the problem posed by the summation over all model when the number of covariates is very large. In this study, an analysis was done with Gross Domestic Product (GDP) as the response variable and 19 predictor variables using the proposed method. A newly proposed g-prior which stems from the benchmark prior and Hannan-Quinn criterion was used for this analysis. The results of the study show that the new method compares favourably with the MCMC method and in some cases dominates.

Keywords: Bayesian model averaging, Markov Chain Monte Carlo, Sampling Without Replacement-Bayesian Model Averaging, averaging-g-prior

A model can fit a data reasonably well and also provide sensible parameter estimates (Hoeting, Madigan, Raftery and Volinsky, 1999). An alternative model can also provide a good fit to the same data but leads to substantively different parameter estimates. A procedure for choosing the best model of all the competing models according to some criterion is model selection. However, basing inference on a single selected model may be risky as conditioning on a single selected model ignores model uncertainty and consequently leads to the underestimation of uncertainty when making inferences about the quantities of interest.

Model uncertainty has been a central problem in regression analysis and has in particular played a major role in economic growth and research since the 1990s, during which time a large number of economics literatures were written in attempt to evaluate the new growth determinants, (Durlauf, Johnson and Temple, 2005). Sparks, Khare and Ghosh (2015) considered posterior consistency for parameter estimation, rather than model selection. Bayesian Model Averaging (BMA) as developed by Leamer (1978), Raftery (1998), Madigan and Raftery (1994) overcomes this problem of uncertainty in model selection. The Bayesian Model Averaging (BMA) has become widely accepted, as its strength is in accounting for uncertainty involved in model selection. Thus, from a Bayesian perspective, the current approach to addressing the problem of model uncertainty lies in the method of Bayesian Model Averaging, Kaplan and Lee (2015)

Several works have been done in this area. Annest, Bumgarner, Raftery and Yeung (2009) used the iterative Bayesian Model Averaging (BMA) method in applying survival analysis to microarray data to determine a highly predictive model of patient's time to event (such as death, relapse or metastasis) using a small number of selected genes. Bayesian model averaging (BMA) has also been applied in estimating the association between air pollutants and fatal health outcomes (Fang, Li, Kan, Bottai, Fang and Cao, 2016). Kaplan and Chen (2014) investigated the use of Bayesian Model Averaging in propensity score analysis for quasi-experimental or observational studies. Eicher, Papageorgiou and Raftery (2011) used predictive performance in Bayesian Model Averaging to assess the prior distributions.

Bayesian model averaging is a method that combines the predictive densities generated by individual members of an ensemble, Kleiber, Raftery, Baars, Gneiting, Mass and Grimit (2010), and has been used in postprocessing of ensembles in forecasting weather quantities such as 2-m temperature, sea level pressure, precipitation, Sloughter, Raftery, Gneiting and Fraley (2007), wind speed, Sloughter, Gneiting and Raftery (2010), and also hydrologic streamflow (Duan, Ajami, Gao and Sorooshian, 2007). Sloughter, Gneiting and Raftery (2013) also extended Bayesian Model Averaging (BMA) methodology to use bivariate distributions to provide probabilistic forecast of wind vector.

According to Piironen and Vehtari (2017), Bayesian Model Averaging is generally better on the basis of prediction than a single model. Lyocsa, Molnar and Todorova (2017) demonstrated that Bayesian Model Averaging improves results in forecasting.

However, the implementation of Bayesian Model Averaging is difficult as summing over all competing models, 2^k models is almost impractical when k (number of independent variables) is large (Hoeting, Madigan, Raftery and Volinsky, 1999), Liang, Truong and Wong (2001), Steel (2019). Another implementation challenge with Bayesian model averaging (BMA) is the specification of prior probability over all parameters in all the models and the specification of prior probability of each model, Eicher, Pagageorgiou and Raftery (2011).

Madigan and Raftery (1994) proposed a method of averaging over a set of models that are supported by the data using the Occam's window method. They argued that if a model predicts the data far less well than the model which provides the best predictions, then it should be excluded. Again complex models which receive less support from the data than their simpler counterparts were excluded. The Occam's window is among the methods implemented in the BMA R package of (Raftery, Hoeting, Volinsky, Painter and Yeung, 2010). Madigan and York (1995) applied a Markov Chain Monte Carlo (MC^3) method to directly approximate the posterior distribution of the quantity of interest given the data.

The main objective of this research is to reduce the encumbrance in the use of BMA by proffering a method that will address the problem of summing over all 2^k models for large k. We propose a method that involves sampling without replacement from the independent variables and applying BMA to the generated models. This method is hereinafter referred to as SWR-BMA.

On the specification of parameter priors, Fernandez et al (2001) have suggested prior should be large enough as to minimize the effect it will have on the result so as to keep the result close to ordinary least squares coefficients. However, Ciccone and Jarocinski (2010) demonstrated that under noisy data such a large prior may not accommodate the noise and this consequently may lead to over-fitting. Eicher et al (2009) suggested a fixed value be given for priors. The uniform information prior has been criticized as being too conservative, Raftery (1999). In this study, we propose a prior which stems from the benchmark prior of Fernandez et al (2001) and Hannan-Quinn criterion of Hannan and Quinn (1979), and use it in the BMA analysis

Methodology

Bayesian Model Averaging

Bayesian Model Averaging accounts for the uncertainty involved in model selection. This approach is to make inference from a posterior distribution defined on the model space. If Δ is the quantity of interest such as a future observation or the utility of a course of action, and $M = \left(M_1, ..., M_l\right)$ denotes the set of all models considered, then the posterior distribution of Δ given the data D is

$$P\left(\Delta \mid D\right) = \sum_{i=1}^{l} P\left(\Delta \mid M_i, D\right) P\left(M_i \mid D\right) \qquad 1$$

This is an average of the posterior distributions under each of the models considered weighted by their corresponding posterior model probabilities. The posterior probability of the model M_i is given by

$$P\left(M_i \mid D\right) = \frac{P\left(D \mid M_i\right) P\left(M_i\right)}{\sum_{j=1}^{l} P\left(D \mid M_j\right) P\left(M_j\right)} \qquad 2$$

Where $P\left(M_i\right)$ is the prior probability that M_i is the true model, and $P\left(D \mid M_i\right) = \int P\left(D \mid \theta_i, M_i\right) P\left(\theta_i \mid M_i\right) d\theta_i$ is the marginal likelihood of M_i , θ_i is the vector of parameters of model M_i , $P\left(\theta_i \mid M_i\right)$ is the prior density of θ_i under model M_i, $P\left(D \mid \theta_i, M_i\right)$ is the likelihood.

The posterior mean and variance are as follows:

$$E\left(\Delta \mid D\right) = \sum_{i=0}^{l} \hat{\Delta}_i P\left(M_i \mid D\right) \qquad 3$$

$$\operatorname{var}\left(\Delta \mid D\right) = \sum_{i=0}^{l} \left(\operatorname{var}\left[\Delta \mid D, M_i\right] + \hat{\Delta}_i^2\right) P\left(M_i \mid D\right) - E\left(\Delta \mid D\right)^2 \qquad 4$$

where $\hat{\Delta}_i = E\left(\Delta \mid D, M_i\right)$ Raftery (1993) and Draper (1995).

This averaging over all models enables better predictive ability than using any single model if performance is measured by a logarithmic scoring rule (Madigan and Raftery, 1994), Ley and Steel (2009).

Zellner's g prior

Each model to be considered is of the form

$$y = \alpha + \beta_1 x_1 + ... + \beta_k x_k + \varepsilon \qquad 5$$

Suppose for each individual model M_i, the error term is independently and identically normally distributed random variable , $\varepsilon \sim N\left(0, \sigma^2 I\right)$. To obtain the posterior distribution on the model parameters, priors are specified on the constant and error variance

$$P(\alpha_i) \propto 1 \text{ and } P(\sigma^2) \propto (\sigma^2)^{-1} \tag{6}$$

The more important prior is placed on the regression coefficients, β_i , which expresses the researcher's belief about the coefficients. A conservative prior of zero mean for the coefficients is common, with variance, $g\sigma^2\left(X_i^T X\right)^{-1}$, where g is referred to as the Zellner's g :

$$\beta_i \mid g \sim N\left(0, g\sigma^2\left(X_i^T X\right)^{-1}\right) \tag{7}$$

A small g implies that the researcher is indeed certain that the coefficients are zero, while a large g indicates researcher's uncertainty about zero coefficients. The posterior distribution follows a t distribution having expected value

$$E(\beta_i \mid y, X, g, M_i) = \frac{g}{1+g}\hat{\beta}_i \tag{8}$$

Where $\hat{\beta}_i$ is the standard ordinary least squares estimator for model M_i . The posterior covariance is given by

$$\operatorname{cov}(\beta_i \mid y, X, g, M_i) = \frac{(y-\bar{y})^T (y-\bar{y})}{N-3}\frac{g}{1+g}\left(1-\frac{g}{1+g}R_i^2\right)\left(X_i^T X\right)^{-1} \tag{9}$$

Where R_i^2 is the usual coefficient of determination for model M_i . The marginal likelihood resulting from this prior framework in Bayesian model averaging is

$$P(y \mid M_i, X, g) \propto (y-\bar{y})^T (y-\bar{y})^{-\frac{N-1}{2}}(1+g)^{-\frac{k_i}{2}}\left(1-\frac{g}{1+g}\right)^{-\frac{N-1}{2}} \tag{10}$$

Model Prior

The uniform model prior was suggested by Raftery (1998), and George and McCulloch (1993). It assigns equal prior probability to all 2^k models, so that $P(M_i) = 2^{-k}$.

Sampling without Replacement-Bayesian Model Averaging (SWR-BMA)

Let k be the number of independent variables. A random sample of size s without replacement from k independent variables with r replications will yield r different models with s independent variables each.

$$y = \alpha + \beta X_s + e \tag{11}$$

where $X_s \subset X$ is the design matrix with s independent variables. Thus, we have

$$\begin{aligned} y_{1i} &= \alpha_1 + \beta_{1i}X_{1s} + e_{1i} \\ y_{2i} &= \alpha_2 + \beta_{2i}X_{2s} + e_{2i} \\ \vdots \quad & \quad \vdots \qquad \vdots \qquad \vdots \\ y_{ri} &= \alpha_2 + \beta_{ri}X_{rs} + e_{ri} \end{aligned} \tag{12}$$

for $i = 1, ..., n$, the number of observations.
A BMA is run on the $r-$ models, and the post mean averaged over all the $r-$ models.
A simple random sample of size 6 was drawn without replacement from the 19 predictor variables, with 24 replications. The generated models were then used for the analysis.

Averaging-g-prior (av-g-prior)
The av-g-prior which sets

$$g = \frac{1}{2}\left[\max\left(n, k^2\right) + \left(\ln n\right)^3\right] \qquad 13$$

was employed in a comparative analysis of different parameter priors, and performed creditably better in some cases than the uniform information prior (UIP), Hannan-Quinn criterion (HQ) and benchmark prior (BRIC), John, Adiele and Nwabueze (2019). In this study, the av-g-prior is used on the two methods: MCMC and SWR-BMA.

Results and Discussion
In this work, we have used data obtained from Central Bank of Nigeria statistical bulletin having gross domestic product (GDP) as the response variable and 19 predictor variables which are described as follows: industrial output (INDQ), money supply (MS), gross fixed capital formation, credit to private sector (CPS), recurrent expenditure (RECEX), balance of payment (BOP), savings (SAV), stock market capitalization (SMC), external reserve (EXTR), external debt (EXTDT), income tax (INCTX), unemployment (UEMP), financial deepening (FD), oil price (OILP), domestic debt (DOMD), inflation (INF), exchange rate (EXCR), capital expenditure (CAPEX), lending rate (LR). Analysis of this data using the Markov Chain Monte Carlo (MCMC) method and the SWR-BMA give the result below. In both methods we have use the newly introduced g-prior.

Table 1: Post Mean and Standard Deviation

variable	PIP	AvPostMean	postSD
indQ	0.97762	3.59793	0.83428
ms	0.30312	0.33892	0.76503
gfcf	0.58257	0.00501	0.00401
cps	0.22894	0.11701	0.47288
recEx	0.71154	7.44779	3.23975
bop	0.20455	-0.00012	0.00045
sav	0.26861	0.51695	1.28751
smc	0.32365	0.17701	0.25571
extR	0.46694	0.09737	0.09574
extDt	0.14758	-0.01139	0.11259
incTx	0.55209	0.00251	0.00146
uemp	0.16576	-10.28151	53.89658
fd	0.19270	-10.37443	141.0305
oilp	0.19591	0.54180	2.52806
domD	0.32710	1.28388	1.72679
inf	0.14160	-0.26410	7.92463
excr	0.17378	-1.31057	9.24437
capEx	0.18320	0.34953	2.18442
lr	0.14675	-1.93531	30.52152

Table 1 gives the result of the BMA analysis of the data using the proposed method. Column 2 is the posterior inclusion probability; column 3 is the averaged posterior mean and column 4 is the posterior standard deviation

Table 2: Post Mean and Standard Deviation (MCMC)

variable	PIP	PostMean	postSD
indQ	0.9666	3.34995	0.98338
ms	0.3829	0.82860	1.68692
gfcf	0.4534	0.00202	0.00269
cps	0.1979	-0.36996	1.24501
recEx	0.1643	0.74065	2.30162
bop	0.1702	-0.000096	0.00027
sav	0.1848	0.25800	1.17122
smc	0.1253	0.02990	0.10953
extR	0.1172	0.01079	0.04221
extDt	0.1057	-0.01563	0.06929
incTx	0.1086	0.000088	0.00067
uemp	0.0818	-4.67627	26.2901
fd	0.0737	-1.60157	63.9436
oilp	0.0943	0.17857	0.97112
domD	0.0807	0.00297	0.54302
inf	0.0510	-0.15585	2.65144
excr	0.0645	-0.34547	3.51244
capEx	0.0731	-0.05086	0.68208
lr	0.0539	-0.01767	9.12433

Table 2 gives the result of the BMA analysis using the Markov Chain Monte Carlo method. Column 2 is the posterior inclusion probability; column 3 is the posterior mean and column 4 is the posterior standard deviation

Table 3: Posterior Mean

Variable	MCMC	SWR-BMA
indQ	3.34995	3.59793
ms	0.82860	0.33892
gfcf	0.00202	0.00501
cps	-0.36996	0.11701
recEx	0.74065	7.44779
bop	-0.000096	-0.00012
sav	0.25800	0.51695
smc	0.02990	0.17701
extR	0.01079	0.09737
extDt	-0.01563	-0.01139
incTx	0.000088	0.00251
uemp	-4.67627	-10.28151
fd	-1.60157	-10.37443
oilp	0.17857	0.54180
domD	0.00297	1.28388
inf	-0.15585	-0.26410
excr	-0.34547	-1.31057
capEx	-0.05086	0.34953
lr	-0.01767	-1.93531

Table 3 gives a comparison of the posterior mean of the proposed method and Markov Chain Monte Carlo method. Column 2 shows results of MCMC method while column 3 the result of theproposed method. Here, similarities in posterior mean of some variables namely indQ, gfcf, bop, extDt and inf as obtained from MCMC method and SWR-BMA method can be observed. This can be seen clearly in Table 6.

Table 4: Posterior Standard Deviation

Variable	MCMC	SWR-BMA
indQ	0.98338	0.83428
ms	1.68692	0.76503
gfcf	0.00269	0.00401
cps	1.24501	0.47288
recEx	2.30162	3.23975
bop	0.00027	0.00045
sav	1.17122	1.28751
smc	0.10953	0.25571
extR	0.04221	0.09574
extDt	0.06929	0.11259
incTx	0.00067	0.00146
uemp	26.2901	53.89658
fd	63.9436	141.0305
oilp	0.97112	2.52806
domD	0.54302	1.72679
inf	2.65144	7.92463
excr	3.51244	9.24437
capEx	0.68208	2.18442
lr	9.12433	30.52152

Table 4 gives a comparison of the posterior standard deviation of the proposed method and the Markov Chain Monte Carlo method. The similarities in posterior standard deviation of variables gfcf, bop, sav, incTx and extDt as obtained from the two methods were observed here and can be clearly seen in Table 7. However, SWR-BMA method has smaller standard deviation in some variables namely, indQ, ms, cps, with values 0.8343, 0.7650, 0.4729 respectively as compared with MCMC method which has values 0.9834, 1.6869, 1.2450 respectively

Table 5: Posterior Inclusion Probability (PIP)

Variables	MCMC	SWR-BMA
indQ	0.9666	0.97762
ms	0.3829	0.30312
gfcf	0.4534	0.58257
cps	0.1979	0.22894
recEx	0.1643	0.71154
bop	0.1702	0.20455
sav	0.1848	0.26861
smc	0.1253	0.32365
extR	0.1172	0.46694
extDt	0.1057	0.14758
incTx	0.1086	0.55209
uemp	0.0818	0.16576
fd	0.0737	0.19270
oilp	0.0943	0.19591
domD	0.0807	0.32710
inf	0.0510	0.14160
excr	0.0645	0.17378
capEx	0.0731	0.18320
lr	0.0539	0.14675

Table 5 gives a comparison of the posterior inclusion probability of the proposed method and the Markov Chain Monte Carlo method. Table 5 shows that SWR-BMA has higher posterior inclusion probabilities (PIP) in all the variables except ms. The sum of the posterior inclusion probabilities for the new method agrees with the mean number of regressors in the each model, and five variables have PIPs that are approximately greater than or equal to 0.5.

Table 6: Similarities in Posterior Mean

variables	MCMC	SWR-BMA
indQ	3.35	3.59
gfcf	0.002	0.005
bop	-0.0001	-0.0001
extDt	-0.015	-0.011
inf	-0.16	-0.26

Table 7: Similarities in Posterior Standard Deviation

Variables	MCMC	SWR-BMA
gfcf	0.003	0.004
bop	0.0003	0.0004
sav	1.17	1.28
incTx	0.001	0.001
extDt	0.07	0.11

Conclusion

When the number of predictor variables is large, the use of BMA is usually faced with a big challenge as the summation over all the competing 2^k models becomes almost impractical. The results in this study has shown that this challenge can be overcome by using SWR-BMA which reduces the number of models to be summed over and also achieve a good result as compared with MCMC method. The proposed g-prior also shows a reliable non informative prior in this work and suits the proposed method well.

References

Annest, A., Bumgarner, R. E., Raftery, A. E. & Yeung, K. Y. (2009). Iterative Bayesian Model Averaging: A Method for Application of Survival Analysis to High-Dimensional Microarray Data. *Bioinformatics* 10:72

Ciccone, A. & Jarocinski, M. (2010). Determinants of Economic Growth: Will Data Tell? *American Economic Journal: Macroeconomics*, 2 (4): 222 – 246

Duan, Q., Ajami, N. K., Gao, X. & Sorooshian, S. (2007). Multi Model Ensemble Hydrologic Prediction Using Bayesian Model Averaging. *Advances in Water Resources*. 30, 1371 – 1386

Durlauf, S. N., Johnson, P., Temple, J. (2005). Growth Econometrics. In Aghion, P, Durlauf, N (eds) Handbook of Economic Growth, North-Holland: Amsterdam; 555 – 677

Eicher, T., Papageorgiou, C. & Raftery, A. (2009). Determining Growth Determinants: Default Priors and Predictive in Bayesian Model Averaging. *Journal of Applied Economics* 1 – 31

Eicher, T. S., Papageorgiou, C. & Raftery, A. E. (2011). Default Priors and Predictive Performance in Bayesian Model Averaging with Application to Growth Determinants. *Journal of Applied Econometrics* 26, 30 – 35

Fang, X., Li, K., Kan, H., Bottai, M., Fang, F. & Cao, Y. (2016). Bayesian Model Averaging Method for Evaluating Associations Between Air Pollution and Respiatory Mortality: A Time Series Study. *British Medical Journal Open* 6(8): 1- 10

Fernandez, C., Ley, E. & Steel, M. F. J. (2001). Benchmark Prior for Bayesian Model Averaging. *Journal of Econometrics* 100, 381 – 427

George, E. & McCulloch, R. (1993). Variable Selection via Gibbs Sampling. *Journal of the American Statistical Association*. 88, 881 – 889

Hannan, E. J. & Quinn, B. G. (1979). The Determination of the Order of an Autoregression. *Journal of the Royal Statistical Society Series B* 41: 190 – 195

Hoeting, J. A., Madigan, D., Raftery, A. E. & Volinsky, C. T. (1999). Bayesian Model Averaging: A Tutorial. *Statistical Science* 4 (4), 382 – 417

John, O. O., Adiele, D. F. & Nwabueze, J. C. (2019). Overcoming the Complexity of 2^k in Bayesian Models. *Ph.D Dissertation unpublished. Department of Statistics, Michael Okpara University of Agriculture, Umudike*.

Kaplan, D. & Chen, J. (2014). Bayesian Model Averaging for Propensity Score Analysis. *Multivariate Behavioural Research*. 49(6): 505 – 517

Kaplan, D. & Lee, C. (2015). Bayesian Model Averaging Over Directed Acyclic Graphs With Implications for the Predictive Performance of Structural Equation Models. *Structural Equation Modelling: A Multidisciplinary Journal*, 00: 1 – 11

Kleiber, W., Raftery, A. E., Baars, J., Gneiting, T., Mass, C. F. & Grimit, E. (2010). Locally Calibrated Probabilistic Temperature Forecasting Using Geostatistical Model Averaging and Local Bayesian Model Averaging. *Monthly Weather Review* 139, 2630 – 2649

Leamer, E. E. (1978). *Specification Searches: Ad Hoc Inference With Non Experimental Data*. New York: Wiley

Ley, E. & Steel, M. F. J. (2009). On the Effect of Prior Assumptions in Bayesian Model Averaging with Application to Growth Regression. *Journal of Applied Econometrics* 24 (4): 651 – 674

Liang, F., Truong, Y. K. & Wong, W. H. (2001). Automatic Bayesian Model Averaging for Linear Regression and Applications in Bayesian Curve Fitting. *Statistica Sinica* 11, 1005 – 1029

Lyocsa, V., Molnar, P. & Todorova, N. (2017). Volatility Forecasting of Non-ferrous Metal Futures: Covariances, Covariates or Combinations? *Journal of International Financial Markets Institutions and Money* 51, 228 – 247

Madigan, D. & Raftery, A. E. (1994). Model Selection and Accounting for Model Uncertainty in Graphical Models Using Occam's Window. *Journal of the American Statistical Association* 89, 1535 – 1546

Piironen, J. & Vehtari, A. (2017). Comparison of Bayesian Predictive Methods for Model Selection. *Statistics and Computing* 27, 711 - 735

Raftery, A. E. (1998). Approximate Bayes Factors for Generalized Linear Models. *Technical Report no. 121*. Departmentof Statistics, University of Washington

Raftery, A. E. (1993). Bayesian Model Selection in Structural Equation Models. In K. Bollen and J. Long (eds.) Testing Structural Equation Models 163 – 180. Newbury Park, CA: Sage

Raftery, A. E. (1999). Bayes Factors and BIC: Comment on "A Critique of Bayesian Information Criterion for Model Selection." *Sociological Methods and Research* 27, 411 – 427

Raftery, A. E., Hoeting, J., Volinsky, C., Painter, I. & Yeung, K. (2010). Bayesian Model Averaging. R package vs. 3.13. http:// CRAN. R – project.org/package = BMA

Sloughter, J. M. L., Raftery, A. E., Gneiting, T. & Fraley, C. (2007). Probabilistic Quantitative Precipitation Forecasting Using Bayesian Model Averaging. *Monthly Weather Review* 135, 3209 – 3220

Sloughter, J. M. L., Gneiting, T.& Raftery, A. E. (2010). Probabilistic Wind Speed Forecasting Using Ensembles and Bayesian Model Averaging. *Journal of the American Association*. 105, 25 – 35

Sloughter, J. M. L., Gneiting, T. & Raftery, A. E. (2013). Probabilistic Wind Vector Forecasting Using Ensembles and Bayesian Model Averaging. *American Meteorological Society*. 2017 – 2119

Sparks, D., Khare, K. & Ghosh, M. (2015). Necessary and Sufficient Conditions for Dimensional Posterior Consistency under g –priors. *Bayesian Analysis* 10, 627 – 664

Steel, M. F. J. (2019). Model Averaging and its Use in Economics. University of Warwick. https:// arxiv.org/pdf/1709.08221.pdf