Bima Journal of Science and Technology, Vol. 6 (2) Aug, 2022 ISSN: 2536-6041

ZELLNER'S G-PRIORS IN BAYESIAN MODEL AVERAGING OF LARGE MODEL SPACE USING MARKOV CHAIN MONTE CARLO MODEL COMPOSITION APPLICABLE UNDER BAYESIAN MODEL SAMPLING

[1*]OGUNDEJI R. K.[1*] and [2]ADELEKE I. A.

[1]Department of Statistics, University of Lagos, Nigeria.
[2]Department of Actuarial Science and Insurance, University of Lagos, Nigeria

Corresponding Author: rogundeji@unilag.edu.ng

ABSTRACT

Applied researchers are frequently faced with the issue of model uncertainty in situations where many possible models exist. For large model space in regression analysis, the challenge has always been how to select a single model among competing large model space when making inferences. Bayesian Model Averaging (BMA) is a technique designed to help account for the uncertainty inherent in the model selection process. Informative prior distributions related to a natural conjugate prior specification are investigated under a limited choice of a single scalar hyper parameter called g-prior which corresponds to the degree of prior uncertainty on regression coefficients. This study focuses on situations with extremely large model space made up of large set of regressors generated by a small number of observations, when estimating model parameters. A set of g-prior structures in literature are considered with a view to identify an improved g-prior specification for regression coefficients in Bayesian Model Averaging. The study demonstrates the sensitivity of posterior results to the choice of g-prior on simulated data and real-life data. Markov Chain Monte Carlo (MCMC) are used to generate a process which moves through large model space to adequately identify the high posterior probability models using the Markov Chain Monte Carlo Model Composition (MC^3), a method applicable under Bayesian Model Sampling (BMS). To assess the sensitivity and predictive ability of the g-priors, predictive criteria like Log Predictive Score (LPS) and Log Marginal Likelihood (LML) are employed. The results reveal a g-prior structure that exhibited equally competitive and consistent predictive ability among considered g-prior structures in literature.

Keywords: Bayesian Model Averaging, Bayesian Model Sampling, Posterior Model Probability, Predictive Performance, Zellner's g-Priors.

INTRODUCTION

In regression analysis, picking a single model among competing models tends to ignore the uncertainty associated with the specification of a selected model as a result of overstatement of the strength of evidence via p-values that are too small (Clyde and George, 2004). Thus, Box (1976) states that "all models are wrong, but some are useful", while Einstein also said that "models should be made as small as possible but not simpler", (Nielsen *et al.*, 2014). Hence, applied researchers are frequently faced with the issue of model uncertainty in situations where many possible models exist as a result of regressors or predictors variables motivated by theory that can be large, some over 40 regressors which can cummulate into trillions of possible models, (George, 2000). Thus, data analysts are unsure of which of these regressors are useful. To complicate the situation, the numbers of observations available for analysis are small or relatively limited (in some cases often less than 100 observations or even less than number of regressors).

There are three major difficulties that arise when putting the Bayesian approach into practice in situation of large model spaces: the choice of the prior distributions; the computation of the integrated likelihood and the estimation of τ (posterior distribution over the model space), (García-donato and Martínez-beneito, 2013). An alternative approach to model selection is to compute a weighted average of the estimates of all competing models. This method which has become attractive to many practitioners is called model averaging and is able to incorporate model uncertainty into the analysis. Also, there has been increasing interest in forecasting methods that utilise large data sets and Bayesian Model Averaging (BMA) methods have been widely employed in this area (Feldkircher *et al.*, 2013); (Clyde and George, 2004). BMA is an appropriate framework employed to control model uncertainty by considering useful information provided by all competing models in the set. In BMA, posterior model probability is applied as a measure to determine the performance of each model in the set in comparison with one another (Davison, 2008).

Prior distributions play very crucial roles in Bayesian probability theory as it is attractive to have conditional distributions that have a closed form under sampling (Okafor, 1999); (Lee, 2004). Agliari and Parisetti (1988), Zellner (1983, 1986) proposed a procedure for evaluating a conjugate prior distribution referred to as Zellner's informative g-prior, or simply g-prior. The g-prior has been vastly used in Bayesian analysis in multiple regression models, due to the verity that analytical results are more readily available, better computational efficiency and its simple interpretation (Rossi *et al.*, 2005), (Liang *et al.*, 2008).

This study focuses on situations with extremely large model space made up of large set of regressors generated by a small number of observations, when estimating model parameters. The study demonstrates the sensitivity of posterior results to the choice of g-prior on simulated data and real-life data. Markov Chain Monte Carlo (MCMC) are used to generate a process which moves through large model space to adequately identify the high posterior probability models using the "Markov Chain Monte Carlo Model Composition" (MC^3), a method applicable under Bayesian Model Sampling (BMS). To assess the predictive ability of the g-priors, predictive criteria like Log Predictive Score (LPS) and Log Marginal Likelihood (LML) are employed.

The rest of the research is organized into three sections. Section 2 examines the Zellner's g-prior properties, concepts of Bayesian Model Averaging and predictive criteria employed; Section 3 comprises the implementation of discussed methods, results and discussion of findings based on the simulated and real-life data. The conclusion is contained in Section 4.

MATERIALS AND METHODS

Zellner's g-prior Properties

Zellner's g-priors applied in BMA analysis fixes a constant $g > 0$ and specifies the Gaussian Prior for the regression coefficients β, conditional on σ^2. Zellner's g reduces the elicitation of the covariance structure by simply choosing the scalar g (Agliari and Parisetti, 1988).

Assumed model:

$$Y = X\beta + \varepsilon \qquad (1)$$

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

The likelihood:

$$P(Y \mid X, \beta, \sigma^2) = (2\pi)^{\frac{-n}{2}} (\sigma^2)^{\frac{-n}{2}} \exp\left(-\frac{1}{2\sigma^2} (Y - X\beta)'(Y - X\beta) \right) \quad (2)$$

The joint likelihood density function for all the data conditional on the unknown parameters (β, σ^2):

$$L = \prod_{i=1}^{n} f(Y \mid X, \beta, \sigma^2) = (2\pi)^{\frac{-n}{2}} (\sigma^2)^{\frac{-n}{2}} \exp\left(-\frac{1}{2\sigma^2} (Y - X\beta)'(Y - X\beta) \right) \quad (3)$$

The Prior:

$$\beta \mid \sigma^2 \sim N(\beta_0, g\Omega) \quad (4)$$

The model (1) yields the maximum likelihood estimate $\hat{\beta}$, which has variance $\sigma^2 (X^T X)^{-1} = \Omega$.

The Posterior:

$$\beta \mid \sigma^2, X \sim \mathrm{N}\left(\beta_0, \sigma^2 g (X^T X)^{-1}\right) \quad (5)$$

$$\beta \mid Y, \sigma^2, X \sim N\left(\frac{1}{g+1}(\beta_0 + g\hat{\beta}), \frac{g\sigma^2}{g+1}(X^T X)^{-1} \right) \quad (6)$$

Inference proceeds from the posterior distribution:

$$P(\beta, \sigma^2 \mid Y) = \frac{P(\beta, \sigma^2)\, P(Y \mid \beta, \sigma^2)}{P(Y)} \quad (7)$$

Where P(Y) = $\int P(\beta, \sigma^2)\, P(Y \mid \beta, \sigma^2)$ is marginal likelihood of the data Y.

$$E\left[\beta \mid Y, \sigma^2\right] = \left(\frac{1}{g\sigma^2} X^T X + \frac{1}{\sigma^2} X^T X \right) \left(\frac{1}{g\sigma^2} X^T X \beta_0 + \frac{1}{\sigma^2} X^T Y \right) \quad (8)$$

$$= \frac{1}{1+g} \beta_0 + \frac{g}{1+g} (X^T X)^{-1} X^T Y \quad (9)$$

$$= \frac{1}{1+g} \beta_0 + \frac{g}{1+g} \hat{\beta} \quad (10)$$

Thus the parameter g allows for direct weighting of the prior, β_0, and data, $\hat{\beta}$. This prior is known as Zellner's informative g-prior, or often referred to simply as" g-prior". The hyper parameter g embodies how certain a researcher is that the coefficients are indeed zero. The value of g corresponds to the degree of prior uncertainty. The g-prior is not only intuitive to use in the model and prior definition, but also leads to familiar posterior results (Zhang *et al.*, 2008).

Two major considerations for Zellner's g-prior include:

Consistency: the choice of g such that posterior model probabilities

asymptotically uncover the "true model", M_j. That is, $P(M_j \mid Y) \to 1 \quad as\ n \to \infty$.

The importance of "g" as a penalty term enforcing parameter parsimony factor $(1+g)^{\frac{k_j + k_s}{2}}$.

Given g, it follows a t-distribution with expected value $E(\beta_j \mid y, X, g, M_j) = \frac{g}{1+g} \hat{\beta}_j$, where $\hat{\beta}_j$ is the standard OLS estimator for the model j.

Different values of g ($0 < g < \infty$) have been assigned in the context of estimation of the regression coefficients of regressors and model sampling from selection. This study identified and considered some set of candidate default

priors (Zellner's informative g-prior that is based on a sample of n observations and k regression coefficients of independent variables) advocated in literature (Eicher *et al.*, 2007), see Table 1.

Table 1: Summary of Identified g-Prior

S/N	Structure of g-prior	Comments and Sources
1	g = n	Unit Information Prior (UIP) based on number of observations. (Kass and Wasserman, 1996).
2	g = max(n, K2)	Corresponds to the benchmark prior suggested by Fernandez, Ley and Steel (2001b).
3	g = K2	Conforms to the risk inflation criterion by Foster and George (1994).
4	$g = \frac{1}{n}$	It is in the spirit of the" unit information priors" of Kass and Wasserman (1996).
5	$g = \frac{k}{n}$	Here, we assign more information to the prior as we have more regressors in the model (Hanson, 2014)
6	$g = \sqrt{\frac{1}{n}}$	This is an intermediate case, where we choose a smaller asymptotic penalty term for large models than in the Schwarz criterion. (Hanson, 2014)
7	$g = \sqrt{\frac{k}{n}}$	Suggested by FLS, where prior information increases with the number of regressors in the model (Fernandez et al., 2001a).
8	$g = In(n^3)$	Asymptotically mimics the Hannan-Quinn criterion with CHQ=3 (Fernandez et al., 2001, p.395)
9	$g = \frac{1}{In(n^3)}$	The Hannan-Quinn criterion. CHQ=3 as n becomes large (Hannan and Quinn, 1979).
10	$g = \frac{In(k+1)}{In(n)}$	Prior information decreases even slower with sample size and there is asymptotic convergence to the Hannan-Quinn criterion with CHQ = 1.
11	$g = \frac{1}{k^2}$	This prior is suggested by the risk inflation criterion (RIC). (Foster and George, 1994).
12	$g = \frac{n}{\sqrt{k}}$	This prior is suggested for increased regressors generated with small observations (Ogundeji et al., 2018)

Structures Examined

Bayesian Model Averaging under Zellner's g-Priors

One way to account for model uncertainty is to allow all models to contribute to inference by averaging. By averaging across a large set of models one can determine those variables which are relevant to the data generating process for a given set of priors used in the analysis. BMA requires a prior probability of each model and a prior probability distribution over the parameters of each model (Montgomery *et al.*, 2010). Once the model space has been determined, the posterior distribution of any coefficient of interest (say β_h), given the data D is:

$$P(\beta_h \mid D) = \sum_{j=1}^{2^k} P(\beta_h \mid M_j) P(M_j \mid D) \quad . \quad (11)$$

BMA uses each model's posterior probability, $P(M_j \mid D)$ as weights. Each model (a set of regressors) receives a weight and the final estimates are constructed as a weighted average of the parameter estimates from each of the models. BMA includes all of the

variables within the analysis, but shrinks the impact of certain variables towards zero through the model weights. These weights are the key feature for estimation via BMA and will depend upon a number of key features of the averaging exercise including the choice of prior specified (Raftery *et al.*, 1993, 1997).

Prior Specification for Model Selection in BMA

A key aspect of the problem is the uncertainty about the choice of regressors *i.e.*, model uncertainty. This means that we need to specify a prior distribution over the model space of all 2^k possible models:

$P(M_j) = p_j$, $j = 1, 2, \ldots 2^k$ with $p_j > 0$ and

$$\sum_{j=1}^{2^k} p_j = 1 \qquad (12)$$

The most common model prior in the literature is the uniform distribution that assigns equal prior probability to all models, so that $P(M_k) = 1/k$ for each k (Raftery, 1993, 1995) and (Yuan *et al.*, 2005). This implies that the prior probability of including a regressor is $1/2^k$ independently of the other regressors included in the model.

Assessment of g-Priors using Predictive Performance

To assess the predictive ability of the g-priors, predictive criteria like Log Predictive Score (LPS) and Log Marginal Likelihood (LML) were employed.

Log Predictive Score (LPS)

The predictive ability of any model is measured by the sum of the logarithm of the posterior predictive ordinates for the observations in the hold-out set. The log score for any given model is the observed coordinate of the predictive density given as:

$$-\sum_{\theta' \varepsilon D^H} \log P\left(\theta' | M_k, D^T\right), \qquad (13)$$

where $P\left(\theta' | M_k, D^T\right)$ is the posterior predictive ordinate. The predictive log score for BMA is then

$$LPS(\theta') = -\sum_{\theta' \varepsilon D^H} \log \left\{\sum_{k=1}^{K} P\left(\theta' | M_k, D^T\right) P\left(M | D^T\right)\right\} \qquad (14)$$

The log predictive score is a proper scoring rule for assessing predictive performance and a smaller value of *LPS* makes a Bayes model a prior choice for *g* that is preferable (Fernandez *et al.*, 2001b).

Log Marginal Likelihood (LML)

The marginal likelihood or the model evidence is the probability of observing the data given a specific model and is defined as:

$$P(X | M) = \int P(X | \theta, M) P(\theta | M) d\theta \qquad (15)$$

If we have two models M_1 and M_2, then we can compare the marginal likelihoods of each, i.e., compare $P(X | M_1)$ to $P(X | M_2)$ and ask which is better (i.e. larger), (Kass *et al.*, 1995).

$$\frac{P(M_1 | X)}{P(M_2 | X)} = \frac{P(M_1)}{P(M_2)} \frac{P(X | M_1)}{P(X | M_2)} \qquad (16)$$

In more than two models, the model with the largest marginal likelihood is the best model.

RESULTS

Simulation Study

The effects of the set of g-priors were examined using simulated dataset drawn from multivariate normal distributions. Five groups of dataset (11-variate, 21-variate, 31-variate, 41-variate, 51-variate and 61-variate) normal random variable were simulated using R and other statistical software (Martin and Stefan, 2013). The simulated study was carried out on the basis of $n = 40$ observations with k = 10, 20, 30, 40, 50 and 60 set of regressors or predictor variables (identified as $X_1, X_2, \ldots, X_{60}$ with a common response variable Y). The

simulated dataset generated were organized as shown in Table 2

Table 2: Summary Data frames of simulated dataset with regressors up to 60 (i.e. Quintillions of models!)

Data Group	1	2	3	4	5	6
Dataset	11-variate	21-variate	31-variate	41-variate	51-variate	61-variate
Regressors (k)	10	20	30	40	50	60
Model Space (2^k)	1024	1.05×10^6	1.07×10^9	1.10×10^{12}	1.13×10^{15}	1.15×10^{18}

Based on Posterior Inclusion Probability (PIP) of the regressors under different g-prior distributions applied, the results show consistency in regressors identified as significant in relationship with the dependent variable across the different g-prior structures. An overview of the result using simulated data as reported above in Table 2 shows that the LML for the twelfth g-prior structure at the different model spaces are of the highest.

Real Life Data

The real life data for the implementation of the sensitivity and predictive performance of identified g-prior structures were obtained. The real-life data was sourced from the National Bureau of Statistics (NBS), 2012 annual reports on all official statistics on socio-economic and macro-economic indicators (literacy, unemployment rates, interest rates, inflation etc.), various machinery and tools that have been brought to bear in improving the efficiency and reliability of official statistics.

Data frames comprises of N = 72 observations with k = 41 regressors or predictor variables, cumulating into 2.20×10^{12} models (i.e. Trillions of models!). Normalizing transformations were made on the data sets to make them multivariate normal, achieved by the standardization of the data set and removal of influential and extreme observations.

Posterior Results for g-Priors Structures Investigated

The effects of the set of g-priors using the above datasets and model space were obtained from the posterior results. To analyse these data, uniform model prior was applied as the model prior for the model space across parameter g-prior structures investigated. Given the model space $2^{41} = 2.2\ X\ 10^{12}$ (over two trillions of models!) and with a fairly large amount of drawings (5million), MC^3 sampler is applied to adequately identify the high posterior probability models. The posterior results from a run with 5 million drawings after a burn-in of 1 million discarded drawings were obtained. The posterior quantities include summary results or output of the BMA analysis include Posterior Inclusion Probabilities (PIP) of 41 regressors, corresponding Posterior Means and Posterior Standard deviation of the parameters estimates across the different parameter g-priors examined. Predictive abilities based on these real data under the different choices of g-priors were examined and compared using Log Predicted Scores (LPS) and Log Marginal Likelihood (LML), (see Tables 3 and 4).

Table 3: Predictive ability under different choices of g-priors examined using both LPS and LML for the Normalised Real Life Data.

	g-prior Structures	Normalised Real-life Data	
		LPS	LML
S/N		n = 72, k = 41	n = 72, k = 41
		g = 72	g = 72
1	g = n	-130.7	373.9
	g = max (n, K^2).	g = 1681	g = 1681
2 & 3	$g = k^2$	-135.7	374
	$g = \frac{1}{n}$	g = 0.013889	g = 0.013889
4		-119.7	329.3
	$g = \frac{k}{n}$	g = 0.569444	g = 0.569444
5		-125.1	327.9
	$g = \sqrt{\frac{1}{n}}$	g = 0.117851	g = 0.117851
6		-120.8	328.8
	$g = \sqrt{\frac{k}{n}}$	g = 0.754615	g = 0.754615
7		-126.4	327.8
	$g = In(n^3)$	g = 12.83	g = 12.83
8		-140.2	344.5
	$g = \frac{1}{In(n^3)}$	g = 0.077942	g = 0.077942
9		-120.4	327.8
	$g = \frac{In(k+1)}{In(n)}$	g = 0.873968	g = 0.873968
10		-127.1	329.3
	$g = \frac{1}{k^2}$	g = 0.000595	g = 0.000595
11		-119.5	342.7
	$g = \frac{n}{\sqrt{k}}$	g = 11.24451	g = 11.24451
12		-140.1	383.9

An overview of the result as reported above in Table 3 shows that the LPS for the 12th g-prior structure at the different model spaces using normalized real-life data are of the lowest. Similarly, the results show that the LML for the 12th g-prior structure at the different model spaces using normalized real-life data are of the highest.

Table 4: Predictive Ability under Different Choices of g-priors Examined using Both LPS and LML for the Un-normalised Real Life Data

	g-prior Structures	Un-normalised Real-life Data	
		LPS	LML
S/N		n = 72, k = 41	n = 72, k = 41
		g = 72	g = 72
1	g = n	-137.5	363.9
	g = max (n, K^2)	g = 1681	g = 1681
2 & 3	g = k^2	-115.8	354
	$g = \frac{1}{n}$	g = 0.013889	g = 0.013889
4		-129.9	349.3
	$g = \frac{k}{n}$	g = 0.569444	g = 0.569444
5		-135.2	317.9
	$g = \sqrt{\frac{1}{n}}$	g = 0.117851	g = 0.117851
6		-130.4	328.8
	$g = \sqrt{\frac{k}{n}}$	g = 0.754615	g = 0.754615
7		-136.6	337.8
	$g = In(n^3)$	g = 12.83	g = 12.83
8		-150.7	354.5
	$g = \frac{1}{In(n^3)}$	g = 0.077942	g = 0.077942
9		-125.8	377.8
	$g = \frac{In(k+1)}{In(n)}$	g = 0.873968	g = 0.873968
10		-137.7	309.3
	$g = \frac{1}{k^2}$	g = 0.000595	g = 0.000595
11		-129.7	312.7
	$g = \frac{n}{\sqrt{k}}$	g = 11.24451	g = 11.24451
12		-150.8	393.9

An overview of the result as reported above in table 4 shows that the LPS for the twelfth g-prior structure at the different model spaces using un-normalized real-life data are of the lowest. Similarly, the results show that the LML for the 12th g-prior structure at the different model spaces using normalized real-life data are of the highest.

Table 5: Comparing the Predicted Values of the 72nd Observation (Dependent Variable) with its Actual Value using Respective Log Predictive Scores (LPS) Across Parameter g-priors.

S/N	g-prior	Actual Value	Predicted Value	LPS
1	g = n	0.0046	0.0013	-3.716
2	g = max (n, k^2)	0.0046	0.0021	-3.649
3	$g = k^2$	0.0046	0.0021	-3.649
4	$g = \frac{1}{n}$	0.0046	0.021	-2.603
5	$g = \frac{k}{n}$	0.0046	0.016	-2.917
6	$g = \sqrt{\frac{1}{n}}$	0.0046	0.02	-2.683
7	$g = \sqrt{\frac{k}{n}}$	0.0046	0.015	-2.981
8	$g = \ln(n^3)$	0.0046	0.031	-3.633
9	$g = \frac{1}{\ln(n^3)}$	0.0046	0.021	-2.653
10	$g = \frac{\ln(k+1)}{\ln(n)}$	0.0046	0.0145	-3.019
11	$g = \frac{1}{k^2}$	0.0046	0.0214	-2.592
12	$g = \frac{n}{\sqrt{k}}$	0.0046	0.0033	-3.614

Table 5 above compares the sensitivity and predictive abilities of the different g-priors by predicting the 72nd observation (Dependent Variable), using 71 observations against 41 regressors (Explanatory Variables) of the real-life dataset. The results from Table 5 show that the actual value of the dependent variable of the 72nd observation is best predicted by one of the g-prior structure investigated: $g = \frac{n}{\sqrt{k}}$ based on the predicted values (closest to the actual value). Though, preceded by g-prior no. 1, 2 & 3 having one of the lowest LPS.

CONCLUSION

Bayesian methodology provides a formal framework to implement model averaging under large model space with reliable results obtained. This study provides closed-form solutions and reduces the complexity of prior elicitation to one scalar g relying on its virtues for consistency and penalty term for model size. Given huge model space generated by

increased number of regressors with relatively small number of observations, reliable results were obtained using some set of g-prior structures in literature. Under a uniform model prior, the study identifies g-priors no. 1, 2, 3 & 12 (Table 1) to exhibit consistently competitive and reliable out-of-sample predictive performance, using both real life and simulated data set. These findings compliment the results of Fernandez, Ley and Steel (2001a, 2001b) and results compares favorably with the effects and results of other g-prior cited in the paper of (Fernandez, Ley and Steel, 2001a, 2001b).

In conclusion, Zellner's g-priors in Bayesian Model Averaging offer a sound fully Bayesian approach that features the virtues of prior input and predictive gains without incurring the risk of misspecification.

REFERENCES

Agliari A and Parisetti C. (1988). A-g Reference Informative Prior: A Note on Zellner's g-Prior. *Journal of the Royal Statistical Society.* Series D (The Statistician). 37, No. 3, 271-275.

Box G.E.P. (1976). Science and Statistics. *Journal of the American Statistical Association*. 71: 791 – 327.

Clyde, M. and George, E. I. (2004), "Model uncertainty," Statistical Science, 19, 81–94.

Davison A.C. (2008). Statistical Models. Cambridge University Press, New York.

Eicher T., Papageorgiou C., Raftery A. (2007). Determining Growth Determinants: Default Priors and Predictive Performance in Bayesian Model Averaging. Center for Statistics and the Social Sciences University of Washington. Working Paper no. 76.

Fernandez, C., Ley, E. and Steel M. (2001a): Model Uncertainty in Cross-Country Growth Regressions, Journal of Applied Econometrics, 16, 563-576.

Fernandez, C., Ley, E. and Steel M. (2001b). "Benchmark Priors for Bayesian Model Averaging," *Journal of Econometrics* 100, 381-427.

Feldkircher, M. and S. Zeugner (2013). "Bayesian Model Averaging with BMS," R-package vs. 0.3.0.

Feldkircher, M. (2012). Forecast Combination and Bayesian Model Averaging: A Prior Sensitivity Analysis." Journal of Forecasting, 31(4):361 – 376.

Foster, D. P. and George, E. I. (1994). The Risk Inflation Criterion for Multiple Regression. *Annals of Statistics*, **22**: 1947 – 1975.

García-donato G. & Martínez-beneito M.A. (2013): On Sampling Strategies in Bayesian Variable Selection Problems with Large Model Spaces, *Journal of the American Statistical Association,* 108:501, 340-352.

George, E.I. (2000) The variable selection procedure. *Journal of the American Statistical Association,* 95, 1304-1308.

Hanson, T. E., Branscum, A. J., and Johnson, W. O. (2014). Informative g-Priors for Logistic Regression. Bayesian Analysis, 9(3):597- 612.

Hannan, E. J. and Quinn, B. G. (1979). The Determination of the Order of an Autoregression. *Journal of the Royal Statistical Society.* Series B (Methodological), 41, (2): 190 – 195.

Kass R. and Wasserman L. (1996): The Selection of Prior Distributions by Formal Rules, Journal of the American Statistical Association, 91:435, 1343 - 1370.

Lee P.M. (2004) Bayesian Statistics: An Introduction. Oxford University Press, New York, Third edition, 2004.

Liang, F., Paulo, R., Molina, G., Clyde, M.,

and Berger, J. (2008), "Mixtures of g-priors for Bayesian variable selection," Journal of the American Statistical Association, 103, 410 – 423.

Montgomery J. and Nyhan B. (2010). Bayesian Model Averaging: Theoretical developments and practical applications. Oxford Journal: Political Analysis (2010) 18 (2): 245-270.

Nielsen, J., Christensen, M., Cemgil, A. and Jensen, S. (2014) Bayesian Model Comparison with the g-Prior. I E E E Transactions on Signal Processing.

Ogundeji, R. K., Adeleke, I. A. and Okafor R. (2018). A New Zellner's g-Prior for Bayesian Model Averaging in. Regression Analysis. *Journal of Scientific Research and Development. (JSRD, Unilag),* 18(1): 51-60.

Okafor, Ray. (1999) Using an Empirical Bayes Model to Estimate Currency Exchange Rates. *Journal of Applied Statistics,* 26, (8): 973 – 983.

Raftery, A. E., Madigan, D., and Hoeting, J. A. (1997). "Bayesian model averaging for linear regression models," *Journal of the American Statistical Association,* 92, 179–191.

Raftery, A.; Madigan, D.; Hoeting, J. (1993). *Model Selection and Accounting for Model Uncertainty in Linear Regression Models*. *Technical Report No*. *262*; Department of Statistics, University of Washington: Washington, DC, USA, 1993.

Raftery, A. (1995). Bayesian Model Selection in Social Research. In *Sociological Methodology*; Marsden, P.V., Ed.; Blackwells: Cambridge, UK, 1995; 111-196.

Rossi P.E, Allenby G.M, and Mcculloch R. (2005). Bayesian Statistics and Marketing. Wiley Series in Probability and Statistics.

Yuan Z. and Yang Y. (2005): Combining Linear Regression Models, *Journal of the American Statistical Association,* 100:472, 1202-1214.

Zellner, A. (1983). Applications of Bayesian Analysis in Econometrics. Journal of the Royal Statistical Society. Proceedings of the 1982 I.O.S. Annual Conference on Practical Bayesian Statistics, Series D (The Statistician), Vol. 32, No. 1/2, pp. 23-34.

Zellner, A. (1986): "On Assessing Prior Distributions and Bayesian Regression Analysis with g-Prior Distributions," *The American Statistician*, 49, 327-335.

Zhang Z., Jordan M. and Yeung D. (2008). Posterior Consistency of the Silverman g-prior in Bayesian Model Choice. 22nd Annual Conference Neural Information Processing System (NIPS), 1969-1976.