A New Zellner's *g*-Prior for Bayesian Model Averaging in Regression Analysis

Rotimi Ogundeji,[1]* Ismaila Adeleke[2] and Ray Okafor[1]
[1]Department of Mathematics, University of Lagos, Akoka-Yaba, Lagos, Nigeria
[2]Department of Actuarial Science and Insurance, University of Lagos, Akoka-Yaba, Lagos, Nigeria
**rogundeji@unilag.edu.ng*

Abstract
In regression analysis, one of the main challenges is selecting a single model among competing models when making inferences. Likewise, the issue of the choice of prior distribution has been delicate in data analysis. Informative prior distributions related to a natural conjugate prior specification are investigated under a limited choice of a single scalar hyperparameter called *g*-prior, which corresponds to the degree of prior uncertainty on regression coefficients. This research identified a set of 11 candidate default priors (Zellner's *g*-priors) prominent in the Literature and applicable in Bayesian model averaging. Some new sets of *g*-prior structures were investigated with a view to proposing an improved *g*-prior specification for regression coefficients in Bayesian Model Averaging (BMA) and the predictive performance of these *g*-priors were compared. Results obtained include the respective prior distributions, posterior distributions and sampling properties of the regression parameters, based on the new set of *g*-prior structures investigated. Also, empirical findings revealed that the proposed *g*-prior structure exhibited equally competitive and consistent predictive ability when compared with identified *g*-prior structures from the Literature.

Keywords: bayesian regression, model uncertainty, normal linear model, predictive performance, prior elicitation

Introduction

Research on Bayesian methodology and applications has progressed remarkably in the past few decades and issues of the choice of prior distribution have been quite delicate in data analysis. Procedures for assessing informative prior distributions for the parameters in Bayesian regression models have been put forward by Zellner (1983, 1986), Agliari *et al.* (1988), Fernandez *et al.* (2001a), Eicher *et al.* (2007) and Raftery *et al.* (1997).

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; Rossi *et al*., 2005). 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 (Davison, 2008).

In linear regression model analysis in which *g*-prior is used, it has been noted that the choice of a scalar hyperparameter *g* is crucial for the behaviour of Bayesian Model Averaging (BMA) procedures. The use of BMA provides a natural solution to model uncertainties that lead to better predictions than simply selecting and using one model (Clyde and George, 2004).

The Zellner's *g*-prior structure has proven universally popular in BMA since it leads to simple closed form expressions of posterior quantities and because it reduces prior elicitation to the choice of a single hyperparameter *g*. The elicitation of *g* is subject to intense debate (e.g., Liang *et al.*, 2008; Hoeting *et al.*, 1999; Fernandez *et al.*, 2001a; Eicher *et al.*, 2007) and constitutes the focus of this research. The approach to prior specification in multiple regression models presented here draws inspiration from the work of Feldkircher *et al.* (2012), Fouskakis and Ntzowfras (2013), Hanson *et al.* (2014), and Li and Clyde (2015).

The aim of this study is to investigate new sets of *g*-priors and propose an improved *g*-prior for averaging in large competing model spaces in the context of Bayesian model averaging. The specific objectives of the study are to:

(i) investigate the different informative prior distributions referred to as scalar hyper parameter *g*-prior identified in literature
(ii) investigate some new set of *g*-priors and to propose *g*-priors specification on parameters estimates and also compare the predictive performance of these *g*-priors and

(iii) derive both prior distributions and posterior distributions of the regression parameters using the respective new *g*-prior investigated and obtain posterior quantities for inferences

Materials and Method

This study gives an overview of the set up of Bayesian linear regression model, the techniques of Bayesian model averaging methodology as a Bayesian solution to the problem of model selection and the concepts of Zellner's *g*-prior in improving predictive ability of Bayesian models.

Bayesian linear regression is an approach to linear regression in which the statistical analysis is undertaken within the context of Bayesian inference. Given a prior distribution, explicit results may be obtained for the posterior probability distributions of the model's parameters. A Bayesian linear model is set up as follows (Lee, 2004):

(i) A linear regression model with a vector *Y* regressed on a number of explanatory variables chosen from a set of *k* variables in a matrix *X* of dimension *n* x (*k*+1), are considered:

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

where $Y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} \varepsilon = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix} \varepsilon = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_k \end{bmatrix}$

$$X = \begin{bmatrix} 1 & X_{12} & . & . & X_{1k} \\ 1 & X_{22} & . & . & X_{2k} \\ . & . & . & . & . \\ . & . & . & . & . \\ 1 & X_{n2} & . & . & X_{nk} \end{bmatrix}$$

with β, a vector of unknown regression coefficients and ε is the error term having constant variance σ^2.

(ii) The likelihood function of β and σ^2 for the model (1) based on the sample is the joint probability density function for all the data conditional on the unknown parameters (β, σ^2):

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

(iii) A Bayesian model builds upon the linear regression of *Y* using conjugate priors by specifying

$$P(\beta, \sigma^2) = P(\beta|\sigma^2)P(\sigma^2) = N(\mu_{\beta,}\sigma^2 V_\beta) \quad x$$
$$IG(a, b) = NIG(\mu_\beta, V_\beta, a, b). \quad (3)$$

Thus, conjugate prior for β and σ^2 is the normal-inverse gamma (NIG) probability distribution with parameters: μ_β, V_β, a, b.

(iv) Inference proceed from the posterior distribution:

$$P(\beta, \sigma^2|Y) = \frac{P(\beta,\sigma^2)\ P(Y|\beta,\sigma^2)}{P(Y)} \quad (4)$$

where $P(Y) = \int P(\beta, \sigma^2)\ P(Y|\beta, \sigma^2) d\beta d\sigma^2$ is marginal likelihood of the data *Y* and from (4), all required posterior quantities for BMA inference can be computed analytically.

Bayesian Model Averaging (BMA) is a technique designed to help account for the uncertainty inherent in the model selection process; BMA focuses on which regressors to include in the analysis. 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 (Hoeting *et al.*, 1999). Given a linear regression model with constant term and *k* potential explanatory variables *x1, x2,...,* $\beta_0 xk$ of the form:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ \ldots + \beta_k x_k + \varepsilon \quad (5)$$

This gives rise to *2k* possible sampling models (indexed *Mj, j = 1,2,...,2k*), depending on whether we include or exclude each of the regressors. Once the model space has been determined, the posterior distribution of any coefficient of interest (say β_h), given the data *D* is:

$$P(\beta_h|D) = \sum_{j=1}^{2^k} P(\beta_h|M_j)\ P(M_j|D) \quad (6)$$

BMA uses each model's posterior probability, $P(M_j|D)$ as weights. Each model (a set of variables) 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 features for estimation via BMA and depend upon a number of key features of the averaging exercise including the choice of prior specified (Montgomery *et al*., 2010). The posterior model probability of *Mj* is given by Raftery *et al*. (2010):

$$P(M_j|D) = P(D|M_j)\frac{P(M_j)}{P(D)} = P(D|M_j)\frac{P(M_j)}{\sum_{i=1}^{2^k} P(D|M_i)P(M_i)} \quad (7)$$

where $P(D|M_j) = \int P(D|\beta^j, M_j)P(\beta^j|M_j)d\beta^j \quad (8)$

and β^j is the vector of parameters from model *Mj*, $P(\beta^j \mid M_j)$ is a prior probability distribution assigned to the parameters of model *Mj* and $P(M_j)$ is the prior probability that *Mj* is the true model.

The estimated posterior means and standard deviations of $\hat{\beta} = (\hat{\beta}_0, \hat{\beta}_1, \ldots, \hat{\beta}_k)$ for model *Mj* are then constructed, by García-donato *et al*. (2013), as:

$$E[\hat{\beta} \mid D] = \sum_{j=1}^{2^k} \hat{\beta} P(M_j|D) \quad (9)$$

$$V[\hat{\beta} \mid D] = \sum_{j=1}^{2^k}\left(Var[\beta \mid D, M_j] + \hat{\beta}^2\right)P(M_j|D) - E[\beta \mid D]^2 \quad (10)$$

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. Thus, Zellner's g reduces the elicitation of the covariance structure by simply choosing the scalar g (Agliari *et al*., 1988).

Assumed model: $Y = X\beta + \varepsilon$ (11)
with $\varepsilon \sim N(0, \sigma^2 I_n)$, I_n is an identity matrix of order *n*. The likelihood:

$$P(Y \mid X, \beta, \sigma^2) = (2\pi)^{\frac{-n}{2}}(\sigma^2)^{\frac{-n}{2}} \exp\left(-\frac{1}{2\sigma^2}(Y - X\beta)'(Y - X\beta)\right) \quad (12)$$

The Prior: $\beta \mid \sigma^2 \sim N(\beta_0, \ g\Omega)$ (13)

The Posterior: $\beta \mid \sigma^2, \ X \sim N(\beta_0, \sigma^2 \ g(X^TX)^{-1})$ (14)

$$\beta \mid Y, \sigma^2, \ X \sim N\left(\frac{1}{g+1}(\beta_0 + g\hat{\beta}), \ \frac{\sigma^2 g}{g+1}(X^TX)^{-1}\right) \quad (15)$$

$$E[\beta \mid Y, \sigma^2] = \left(\frac{1}{g\sigma^2}X^TX + \frac{1}{\sigma^2}X^TX\right)\left(\frac{1}{g\sigma^2}X^TX\beta_0 + \frac{1}{\sigma^2}X^TY\right) \quad (16)$$

$$E[\beta \mid Y, \sigma^2] = \frac{1}{1+g}\beta_0 + \frac{g}{1+g}(X^TX)^{-1}X^TY \quad (17)$$

$$= \frac{1}{1+g}\beta_0 + \frac{g}{1+g}\hat{\beta} \quad (18)$$

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 hyperparameter *g* embodies how certain a researcher is that the coefficients are indeed zero. The value of *g* corresponds to the degree of prior uncertainty (Hanson *et al*., 2014). 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).

2 major considerations for Zellner's *g*-prior include:

(i) Consistency: the choice of *g* such that posterior model probabilities asymptotically uncover the "true model" *Mj*. That is,
$$P(M_{j,} \mid Y) \ \rightarrow \ 1 \qquad as \ n \rightarrow \infty.$$

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

Given $g > 0$, 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 *Mj*.

Different values of *g* have been assigned in the context of estimation of the regression coefficients of regressors and model sampling for selection. This research identified a set of 11 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 the Literature (Eicher *et al*., 2007), see Table 1.

Table 1: Summary of Identified *g*-Prior Structures Examined

S/N	Structure of *g*-Prior	Comments/Sources
1	$g = n$	Unit Information Prior (UIP) based on number of observations. (Kass and Wasserman, 1996).
2	$g = max(n, \ k^2)$	Corresponds to the benchmark prior suggested by Fernandez *et al.* (2001b).
3	$g = k^2$	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
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.
7	$g = \sqrt{\frac{k}{n}}$	The prior information increases with the number of regressors in the model. (Fernandez *et al*., 2001a)
8	$g = ln(n^3)$	Asymptotically mimics the Hannan–Quinn criterion with CHQ = 3 (Fernandez *et al*., 2001b, p.395).
9	$g = \frac{1}{ln(n^3)}$	The Hannan–Quinn criterion. CHQ = 3 as *n* becomes large. (Hannan and Quinn, 1979).
10	$g = \frac{ln(k+1)}{ln(n)}$	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).

Outlined and adopted are methods and framework used by Zellner (1986) to obtain both the prior distributions and posterior distributions for multiple regression models. Also, the sampling properties in terms of the expected mean and variance of posterior distributions and consistency properties of *g*-priors based on the results of Zhang *et al.* (2008) are outlined.

Prior distributions and posterior distributions using Zellner's g-priors framework for multiple regression models parameters:

Given a regression model:

$$Y = X\beta + \varepsilon \quad (19)$$

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

The likelihood function for the model is given by

$$l(\beta,\sigma \mid Y, X) \quad \alpha \quad \sigma^{-n}\, exp\left(-\frac{1}{2\sigma^2}(Y - X\beta)'(Y - X\beta)\right) \quad (20)$$

$$\alpha \quad \sigma^{-n}\, exp\left\{-\left[vs^2 + (\beta - \hat{\beta})' X'X(\beta - \hat{\beta})\right]/2\sigma^2\right\} \quad (21)$$

where $\hat{\beta} = (X'X)^{-1}X'Y,\ vs^2 = (Y - X\hat{\beta})'(Y - X\hat{\beta})$ and $v = n\text{-}k$.

Given anticipated values of β and σ^2 denoted by β_a and σ_a^2, respectively, from a conceptual or imaginary sample: $Y_0 = X\beta + \varepsilon_0$ (22)

The joint informative g-prior distribution is:

$$P(\beta, \sigma^2 \mid \eta_0)\ \alpha\ \sigma^{-(v+1)} exp\left\{-\frac{v\bar{\sigma}_a^2}{2\sigma^2}\right\} \quad x$$

$$\sigma^{-k}\, exp\{-\mathrm{g}(\beta - \beta_a)'X'X(\beta - \beta_a)/2\sigma^2\} \quad (23)$$

where $\eta_0' = (\beta_a', \bar{\sigma}_a, \mathrm{g}, v)$.

The marginal prior distributions for β and σ are respectively: $P(\beta \mid \beta_a,\ \mathrm{g}, v)\ \alpha\ \{v\bar{\sigma}_a^2 + \mathrm{g}(\beta - \beta_a)'X'X(\beta - \beta_a)\}^{-(v+k)/2}$ (24)

and $P(\sigma \mid \bar{\sigma}_a, V)\ \alpha\ \sigma^{-(v+1)} exp\{-v\bar{\sigma}_a^2/2\sigma^2\}$ (25)

Similarly, based on the results of Zellner (1986) on posterior distribution for β and σ given a g-prior distribution, posterior distributions for β and σ:

$$P(\beta, \sigma \mid D)\ \alpha\ P(\beta, \sigma)\ l(\beta, \sigma \mid Y) \quad (26)$$

$$\alpha\ \sigma^{-(n+k+1)} exp\{-[(Y - X\beta)'(Y - X\beta) + \mathrm{g}(\beta - \bar{\beta})'X'X(\beta - \bar{\beta})]/2\sigma^2\} \quad (27)$$

D denotes the data and $\bar{\beta}$ is the prior mean vector for regression coefficient vector β.

Let $w' = \left(Y' : \mathrm{g}^{\frac{1}{2}}\, \bar{\beta}'\, X'\right)$ and $Z' = \left(X' : \mathrm{g}^{\frac{1}{2}}\, X'\right)$

Then the terms in square brackets in the exponential can be expressed as:

$$(w - Z\beta)'(w - Z\beta) = (w - Z\bar{\bar{\beta}})'(w - Z\bar{\bar{\beta}}) + (\beta - \bar{\bar{\beta}})'Z'Z(\beta - \bar{\bar{\beta}}) \quad (28)$$

where $\bar{\bar{\beta}} = (Z'Z)^{-1}Z'w$.

Thus (27) can be expressed as:

$$P(\beta, \sigma \mid D)\ \alpha\ \sigma^{-(n+k+1)} exp\left\{-\left[(w - Z\bar{\bar{\beta}})'(w - Z\bar{\bar{\beta}}) + (\beta - \bar{\bar{\beta}})'Z'Z(\beta - \bar{\bar{\beta}})\right]/2\sigma^2\right\} \quad (29)$$

where $\bar{\bar{\beta}} = (Z'Z)^{-1}Z'w = (\hat{\beta} + \mathrm{g}\bar{\beta})/(1 + \mathrm{g})$.

This is the mean of the posterior distribution with $\hat{\beta} = (X'X)^{-1}X'Y$.

The covariance matrix of the conditional normal posterior distribution for β given σ, denoted by $V(\beta \mid \sigma,\ D)$, is

$$V(\beta \mid \sigma,\ D) = (Z'Z)^{-1}\sigma^2 \quad (30)$$

$$= (X'X)^{-1}\sigma^2/(1 + \mathrm{g}) \quad (31)$$

$$P(\beta \mid D)\ \alpha\ \left\{\begin{matrix}(w - Z\bar{\bar{\beta}})'(w - Z\bar{\bar{\beta}}) \\ +(\beta - \bar{\bar{\beta}})'Z'Z(\beta - \bar{\bar{\beta}})\end{matrix}\right\}^{-(n+k)/2} \quad (32)$$

with covariance matrix

$$V(\beta \mid D) = (Z'Z)^{-1}\sigma^2 = (X'X)^{-1}\sigma^2/(1 + \mathrm{g}) \quad (33)$$

where $(n - 2)\sigma^2 \equiv (w - Z\bar{\bar{\beta}})'^{(w - Z\bar{\bar{\beta}})} =$

$$(Y - X\bar{\beta})'(Y - X\bar{\beta}) + \mathrm{g}(\bar{\bar{\beta}} - \beta)'X'X(\bar{\bar{\beta}} - \beta) \quad (34)$$

Also, the marginal posterior distribution for σ obtained from (29) by integrating with respect to β, is

$$P(\sigma \mid D)\ \alpha\ \sigma^{-(n+1)} exp\left\{-(w - Z\bar{\bar{\beta}})'(w - Z\bar{\bar{\beta}})/2\sigma^2\right\} \quad (35)$$

Sampling properties of the mean of the posterior distribution using Zellner's g-priors framework in multiple regression models:

The expected means of $\bar{\bar{\beta}}$ is given by:

$$E\bar{\bar{\beta}} = (\beta_0 + \mathrm{g}\bar{\beta})/(1 + \mathrm{g}) \quad (36)$$

where β_0 is the true unknown value of the regression coefficient vector. The bias of $\bar{\bar{\beta}}$:

$$B(\bar{\bar{\beta}}) = E\bar{\bar{\beta}} - \beta_0 = \mathrm{g}(\bar{\beta} - \beta_0)/(1 + \mathrm{g}) \quad (37)$$

The second moment matrix of $\bar{\bar{\beta}} - \beta_0$ is:

$$E(\bar{\bar{\beta}} - \beta_0)(\bar{\bar{\beta}} - \beta_0)' = V(\hat{\beta})/(1 + \mathrm{g})^2 + B(\bar{\bar{\beta}})B(\bar{\bar{\beta}})' \quad (38)$$

with $V(\hat{\beta}) = (X'X)^{-1}\sigma^2$, the covariance matrix of the least squares estimator.

The variance-covariance matrix of $\bar{\bar{\beta}}$ is:

$$E(\bar{\bar{\beta}} - E\bar{\bar{\beta}})(\bar{\bar{\beta}} - E\bar{\bar{\beta}})' = V(\hat{\beta})/(1 + \mathrm{g})^2 \quad (39)$$

Results and Discussion

Based on the methods above and relying on the results of Zellner (1986) on prior and posterior distributions for multiple regression models, the respective prior and posterior distributions for β and σ were obtained using each of the proposed g-prior structures investigated. Also, the sampling properties in terms of the expected mean and variance of the posterior distributions were obtained using each proposed g-prior structures investigated.

The prior and posterior distributions for β and σ using the proposed g-prior, $\mathrm{g} = n/\sqrt{k}$:

Before observing Y, a conceptual or imaginary sample is considered such that:

$$Y_1 = X\beta + \varepsilon_1 \quad (40)$$

Let β_a and σ_a^2 denote anticipated values of β and σ^2, respectively, from the conceptual/imaginary sample.

The joint informative g-prior distribution using $\mathrm{g} = \frac{\boldsymbol{n}}{\sqrt{\boldsymbol{k}}}$ is: $P(\beta, \sigma^2 \mid \eta_0)\ \alpha\ \sigma^{-(v+1)} exp\left\{-\frac{v\bar{\sigma}_a^2}{2\sigma^2}\right\}$ x

$$\sigma^{-k}\, exp\left\{-\left(\frac{\mathrm{n}}{\sqrt{\mathrm{k}}}\right)(\beta - \beta_a)'X'X(\beta - \beta_a)/2\sigma^2\right\} \quad (41)$$

where $\eta_0' = \left(\beta_a', \bar{\sigma}_a, \left(\frac{\mathrm{n}}{\sqrt{\mathrm{k}}}\right), v\right)$.

The marginal prior distributions for β and σ are respectively:

$$P\left(\beta \mid \beta_a, \left(\frac{n}{\sqrt{k}}\right), v\right) \ \alpha \ \left\{v\bar{\sigma}_a^2 + \left(\frac{n}{\sqrt{k}}\right)(\beta - \beta_a)'X'X(\beta - \beta_a)\right\}^{-(v+k)/2} \quad (42)$$

and $P(\sigma \mid \bar{\sigma}_a, V) \ \alpha \ \sigma^{-(v+1)} exp\{-v\bar{\sigma}_a^2/2\sigma^2\}$ (43)

The posterior distribution for β and σ**:**

$$P(\beta, \sigma \mid D) \ \alpha \ P(\beta, \sigma)\, l(\beta, \sigma \mid Y_1) \quad (44)$$

$$\alpha \ \sigma^{-(n+k+1)} exp\left\{-\left[(Y_1 - X\beta)'(Y_1 - X\beta) + \left(\frac{n}{\sqrt{k}}\right)(\beta - \bar{\beta})'X'X(\beta - \bar{\beta})\right]/2\sigma^2\right\} \quad (45)$$

D denotes the data and $\bar{\beta}$ is the prior mean vector for regression coefficient vector β.

Let $w' = \left(Y_1' : \left(\frac{n}{\sqrt{k}}\right)^{\frac{1}{2}} \bar{\beta}' X'\right)$ and $Z' = \left(X' : \left(\frac{n}{\sqrt{k}}\right)^{\frac{1}{2}} X'\right)$

Then the terms in square brackets in the exponential can be expressed as

$$(w - Z\beta)'(w - Z\beta) = (w - Z\bar{\bar{\beta}})'(w - Z\bar{\bar{\beta}}) + (\beta - \bar{\bar{\beta}})'Z'Z(\beta - \bar{\bar{\beta}}) \quad (46)$$

where $\bar{\bar{\beta}} = (Z'Z)^{-1}Z'w$.

Thus (45) becomes:

$$P(\beta, \sigma \mid D) \ \alpha \ \sigma^{-(n+k+1)} exp\left\{-\left[(w_1 - Z_1\bar{\bar{\beta}})'(w_1 - Z_1\bar{\bar{\beta}}) + (\beta - \bar{\bar{\beta}})'Z_1'Z_1(\beta - \bar{\bar{\beta}})\right]/2\sigma^2\right\} \quad (47)$$

where $\bar{\bar{\beta}} = (Z_1'Z_1)^{-1}Z_1'w_1 = \left(\hat{\beta} + \left(\frac{n}{\sqrt{k}}\right)\bar{\beta}\right)/\left(1 + \left(\frac{n}{\sqrt{k}}\right)\right)$.

This is the mean of the posterior distribution with $\hat{\beta} = (X'X)^{-1}X'Y_1$.

The covariance matrix of the conditional normal posterior distribution for β given σ, denoted by $V(\beta \mid \sigma, D)$, is $V(\beta \mid \sigma, D) = (Z_1'Z_1)^{-1}\sigma^2$ (48)

$$= (X'X)^{-1}\sigma^2/\left(1 + \left(\frac{n}{\sqrt{k}}\right)\right) \quad (49)$$

The marginal posterior distribution for β obtained from (47) by integrating with respect to σ, is

$$P(\beta \mid D) \ \alpha \ \left\{(w_1 - Z_1\bar{\bar{\beta}})'(w_1 - Z_1\bar{\bar{\beta}}) + (\beta - \bar{\bar{\beta}})'Z_1'Z_1(\beta - \bar{\bar{\beta}})\right\}^{-(n+k)/2} \quad (50)$$

with covariance matrix:

$$V(\beta \mid D) = (Z_1'Z_1)^{-1}\sigma^2 = (X'X)^{-1}\sigma^2/\left(1 + \left(\frac{n}{\sqrt{k}}\right)\right) \quad (51)$$

Also, the marginal posterior distribution for σ obtained from (47) by integrating with respect to β is

$$P(\sigma \mid D) \ \alpha \ \sigma^{-(n+1)} exp\left\{-(w_1 - Z_1\bar{\bar{\beta}})'(w_1 - Z_1\bar{\bar{\beta}})/2\sigma^2\right\} \quad (52)$$

Sampling properties of the mean of the posterior distribution:

The expected mean of $\bar{\bar{\beta}}$ using $g = \frac{\boldsymbol{n}}{\sqrt{\boldsymbol{k}}}$ is given by:

$$E\bar{\bar{\beta}} = \left(\beta_0 + \left(\frac{n}{\sqrt{k}}\right)\bar{\beta}\right)/\left(1 + \left(\frac{n}{\sqrt{k}}\right)\right) \quad (53)$$

where β_0 is the true unknown value of the regression coefficient vector.

The bias of $\bar{\bar{\beta}}$:

$$B(\bar{\bar{\beta}}) = E\bar{\bar{\beta}} - \beta_0 = \left(\frac{n}{\sqrt{k}}\right)(\bar{\beta} - \beta_0)/\left(1 + \left(\frac{n}{\sqrt{k}}\right)\right) \quad (54)$$

The second moment matrix of $\bar{\bar{\beta}} - \beta_0$ is:

$$E(\bar{\bar{\beta}} - \beta_0)(\bar{\bar{\beta}} - \beta_0)' = \frac{V(\hat{\beta})}{\left(1 + \left(\frac{n}{\sqrt{k}}\right)\right)^2} + B(\bar{\bar{\beta}})B(\bar{\bar{\beta}})' \quad (55)$$

with $V(\hat{\beta}) = (X'X)^{-1}\sigma^2$, the covariance matrix of the least squares estimator.

The variance-covariance matrix of $\bar{\bar{\beta}}$ is:

$$E(\bar{\bar{\beta}} - E\bar{\bar{\beta}})(\bar{\bar{\beta}} - E\bar{\bar{\beta}})' = V(\hat{\beta})/\left(1 + \left(\frac{g}{\sqrt{k}}\right)\right)^2 \quad (56)$$

The prior and posterior distributions for β and σ using the proposed g-prior, $= \sqrt{(1/k)}$:

The joint informative g-prior distribution using $g = \sqrt{\frac{1}{k}}$ is: $P(\beta, \sigma^2 \mid \eta_0) \ \alpha \ \sigma^{-(v+1)} exp\left\{-\frac{v\bar{\sigma}_a^2}{2\sigma^2}\right\}$ x

$$\sigma^{-k}\, exp\left\{-\left(\frac{1}{k}\right)^{\frac{1}{2}}(\beta - \beta_a)'X'X(\beta - \beta_a)/2\sigma^2\right\} \quad (57)$$

where $\eta_0' = \left(\beta_a', \bar{\sigma}_a, \left(\frac{1}{k}\right)^{\frac{1}{2}}, v\right)$.

The marginal prior distributions for β and σ are respectively:

$$P\left(\beta \mid \beta_a, \left(\frac{1}{k}\right)^{\frac{1}{2}}, v\right) \ \alpha \ \left\{v\bar{\sigma}_a^2 + \left(\frac{1}{k}\right)^{\frac{1}{2}}(\beta - \beta_a)'X'X(\beta - \beta_a)\right\}^{-(v+k)/2} \quad (58)$$

and $P(\beta, \sigma \mid D) \ \alpha \ P(\beta, \sigma)\, l(\beta, \sigma \mid Y_2)$ (59)

Similarly, the posterior distribution for β and σ**:**

$$P(\beta, \sigma \mid D) \ \alpha \ P(\beta, \sigma)\, l(\beta, \sigma \mid Y) \quad (60)$$

$$\alpha \ \sigma^{-(n+k+1)} exp\left\{-\left[(Y_2 - X\beta)'(Y_2 - X\beta) + \left(\frac{1}{k}\right)^{\frac{1}{2}}(\beta - \bar{\beta})'X'X(\beta - \bar{\beta})\right]/2\sigma^2\right\} \quad (61)$$

D denotes the data and $\bar{\beta}$ is the prior mean vector for regression coefficient vector β.

Let $w_2' = \left(Y_2' : \left(\frac{1}{k}\right)^{\frac{1}{4}} \bar{\beta}' X'\right)$ and $Z_2' = \left(X' : \left(\frac{1}{k}\right)^{\frac{1}{4}} X'\right)$.

Then the terms in square brackets in the exponential can be expressed as:

$$(w_2 - Z_2\beta)'(w_2 - Z_2\beta) = (w_2 - Z_2\bar{\bar{\beta}})'(w_2 - Z_2\bar{\bar{\beta}}) + (\beta - \bar{\bar{\beta}})'Z_2'Z_2(\beta - \bar{\bar{\beta}}) \quad (62)$$

where $\bar{\bar{\beta}} = (Z_2'Z_2)^{-1}Z_2'w_2$.

Thus (61) becomes:

$$P(\beta, \sigma \mid D) \ \alpha \ \sigma^{-(n+k+1)} exp\left\{-\left[(w_2 - Z_2\bar{\bar{\beta}})'(w_2 - Z_2\bar{\bar{\beta}}) + (\beta - \bar{\bar{\beta}})'Z_2'Z_2(\beta - \bar{\bar{\beta}})\right]/2\sigma^2\right\} \quad (63)$$

where $\bar{\bar{\beta}} = (Z_2'Z_2)^{-1}Z_2'w_2 = \left(\hat{\beta} + \left(\frac{1}{k}\right)^{\frac{1}{2}}\bar{\beta}\right)/\left(1 + \left(\frac{1}{k}\right)^{\frac{1}{2}}\right)$

is the mean of the posterior distribution with $\hat{\beta} = (X'X)^{-1}X'Y_2$.

The covariance matrix of the conditional normal posterior distribution for β given σ, denoted by $V(\beta \mid \sigma,\ D)$, is $V(\beta \mid \sigma,\ D) = (Z_2'Z_2)^{-1}\sigma^2$ (64)

$$= (X'X)^{-1}\sigma^2 / \left(1 + \left(\frac{1}{\mathrm{k}}\right)^{\frac{1}{2}}\right) \quad (65)$$

The marginal posterior distribution for β obtained from (63) by integrating with respect to σ, is

$$P(\beta \mid D)\ \alpha\ \left\{(w_2 - Z_2\bar{\bar{\beta}})'(w_2 - Z_2\bar{\bar{\beta}}) + (\beta - \bar{\bar{\beta}})'Z_2'Z_2(\beta - \bar{\bar{\beta}})\right\}^{-(n+k)/2} \quad (66)$$

with covariance matrix:

$$V(\beta \mid D) = (Z_2'Z_2)^{-1}\sigma^2 = (X'X)^{-1}\sigma^2 / \left(1 + \left(\frac{1}{\mathrm{k}}\right)^{\frac{1}{2}}\right) \quad (67)$$

Also, the marginal posterior distribution for σ obtained from (63) by integrating with respect to β, is

$$P(\sigma \mid D)\ \alpha\ \sigma^{-(n+1)} exp\left\{-(w_2 - Z_2\bar{\bar{\beta}})'(w_2 - Z_2\bar{\bar{\beta}})/2\sigma^2\right\} \quad (68)$$

Sampling properties of mean of the posterior distribution:

The expected mean of $\bar{\bar{\beta}}$ using $\mathrm{g} = \sqrt{(1/\mathrm{k})}$ is given by: $E\bar{\bar{\beta}} = \left(\beta_0 + \left(\frac{1}{k}\right)^{\frac{1}{2}}\bar{\beta}\right) / \left(1 + \left(\frac{1}{k}\right)^{\frac{1}{2}}\right)$ (69)

where β_0 is the true unknown value of the regression coefficient vector.

The bias of $\bar{\bar{\beta}}$:

$$B(\bar{\bar{\beta}}) = E\bar{\bar{\beta}} - \beta_0 = \left(\frac{1}{k}\right)^{\frac{1}{2}}(\bar{\beta} - \beta_0) / \left(1 + \left(\frac{1}{k}\right)^{\frac{1}{2}}\right) \quad (70)$$

The second moment matrix of $\bar{\bar{\beta}} - \beta_0$ is:

$$E(\bar{\bar{\beta}} - \beta_0)(\bar{\bar{\beta}} - \beta_0)' = V(\hat{\beta}) / \left(1 + \left(\frac{1}{\mathrm{k}}\right)^{\frac{1}{2}}\right)^2 + B(\bar{\bar{\beta}})B(\bar{\bar{\beta}})' \quad (71)$$

with $V(\hat{\beta}) = (X'X)^{-1}\sigma^2$, the covariance matrix of the least squares estimator.

The variance-covariance matrix of $\bar{\bar{\beta}}$ is:

$$E(\bar{\bar{\beta}} - E\bar{\bar{\beta}})(\bar{\bar{\beta}} - E\bar{\bar{\beta}})' = V(\hat{\beta}) / \left(1 + \left(\frac{1}{\mathrm{k}}\right)^{\frac{1}{2}}\right)^2 \quad (72)$$

The prior and posterior distributions for β and σ using the proposed g-prior, $= 1/\mathrm{k}$.

The joint informative g-prior distribution using $\mathrm{g} = \frac{1}{\mathrm{k}}$ is: $P(\beta, \sigma^2 \mid \eta_0)\ \alpha\ \sigma^{-(v+1)} exp\left\{-\frac{v\bar{\sigma}_a^2}{2\sigma^2}\right\}$ x

$$\sigma^{-k}\ exp\left\{-\left(\frac{1}{\mathrm{k}}\right)(\beta - \beta_a)'X'X(\beta - \beta_a)/2\sigma^2\right\} \quad (73)$$

where $\eta_0' = \left(\beta_a', \bar{\sigma}_a, \left(\frac{1}{\mathrm{k}}\right), v\right)$.

The marginal prior distributions for β and σ are respectively:

$$P\left(\beta \mid \beta_a, \left(\frac{1}{\mathrm{k}}\right), v\right)\ \alpha\ \left\{v\bar{\sigma}_a^2 + \left(\frac{1}{\mathrm{k}}\right)(\beta - \beta_a)'X'X(\beta - \beta_a)\right\}^{-(v+k)/2} \quad (74)$$

and $P(\sigma \mid \bar{\sigma}_a, V)\ \alpha\ \sigma^{-(v+1)} exp\{-v\bar{\sigma}_a^2/2\sigma^2\}$ (75)

The posterior distribution for β and σ**:**

$$P(\beta, \sigma \mid D)\ \alpha\ P(\beta, \sigma)\ l(\beta, \sigma \mid Y_3) \quad (76)$$

$$\alpha\ \sigma^{-(n+k+1)} exp\left\{-\left[(Y_3 - X\beta)'(Y_3 - X\beta) + \left(\frac{1}{k}\right)(\beta - \bar{\beta})'X'X(\beta - \bar{\beta})\right]/2\sigma^2\right\} \quad (77)$$

D denotes the data and $\bar{\beta}$ is the prior mean vector for regression coefficient vector β.

Let $w_3' = \left(Y_3' : \left(\frac{1}{\mathrm{k}}\right)^{\frac{1}{2}} \bar{\beta}'\ \mathrm{X}'\right)$ and $Z_3' = \left(X' : \left(\frac{1}{\mathrm{k}}\right)^{\frac{1}{2}}\ \mathrm{X}'\right)$.

Then the terms in square brackets in the exponential can be expressed as

$$(w_3 - Z_3\beta)'(w_3 - Z_3\beta) = (w_3 - Z_3\bar{\bar{\beta}})'(w_3 - Z_3\bar{\bar{\beta}}) + (\beta - \bar{\bar{\beta}})'Z_3'Z_3(\beta - \bar{\bar{\beta}}) \quad (78)$$

where $\bar{\bar{\beta}} = (Z_3'Z_3)^{-1}Z_3'w_3$.

Thus (77) becomes:

$$P(\beta, \sigma \mid D)\ \alpha\ \sigma^{-(n+k+1)} exp\left\{-\left[(w_3 - Z_3\bar{\bar{\beta}})'(w_3 - Z_3\bar{\bar{\beta}}) + (\beta - \bar{\bar{\beta}})'Z_3'Z_3(\beta - \bar{\bar{\beta}})\right]/2\sigma^2\right\} \quad (79)$$

where $\bar{\bar{\beta}} = (Z_3'Z_3)^{-1}Z_3'w_3 = \left(\hat{\beta} + \left(\frac{1}{\mathrm{k}}\right)\bar{\beta}\right) / \left(1 + \left(\frac{1}{\mathrm{k}}\right)\right)$.

.

This is the mean of the posterior distribution with $\hat{\beta} = (X'X)^{-1}X'Y_3$.

The covariance matrix of the conditional normal posterior distribution for β given σ, denoted by $V(\beta \mid \sigma,\ D)$, is $V(\beta \mid \sigma,\ D) = (Z_3'Z_3)^{-1}\sigma^2$ (80)

$$= (X'X)^{-1}\sigma^2 / \left(1 + \left(\frac{1}{\mathrm{k}}\right)\right) \quad (81)$$

The marginal posterior distribution for β obtained from (79) by integrating with respect to σ, is:

$$P(\beta \mid D)\ \alpha\ \left\{(w_3 - Z_3\bar{\bar{\beta}})'(w_3 - Z_3\bar{\bar{\beta}}) + (\beta - \bar{\bar{\beta}})'Z_3'Z_3(\beta - \bar{\bar{\beta}})\right\}^{-(n+k)/2} \quad (82)$$

with covariance matrix:

$$V(\beta \mid D) = (Z_3'Z_3)^{-1}\sigma^2 = (X'X)^{-1}\sigma^2 / \left(1 + \left(\frac{1}{\mathrm{k}}\right)\right) \quad (83)$$

Also, the marginal posterior distribution for σ obtained from (79) by integrating with respect to β, is

$$P(\sigma \mid D)\ \alpha\ \sigma^{-(n+1)} exp\left\{-(w_3 - Z_3\bar{\bar{\beta}})'(w_3 - Z_3\bar{\bar{\beta}})/2\sigma^2\right\} \quad (84)$$

Sampling properties of the mean of the posterior distribution:

The expected mean of $\bar{\bar{\beta}}$ using $\mathrm{g} = \frac{1}{\mathrm{k}}$ is given by:

$$E\bar{\bar{\beta}} = \left(\beta_0 + \left(\frac{1}{k}\right)\bar{\beta}\right) / \left(1 + \left(\frac{1}{k}\right)\right) \quad (85)$$

where β_0 is the true unknown value of the regression coefficient vector.

The bias of $\bar{\bar{\beta}}$:

$$B(\bar{\bar{\beta}}) = E\bar{\bar{\beta}} - \beta_0 = \left(\frac{1}{k}\right)(\bar{\beta} - \beta_0) / \left(1 + \left(\frac{1}{k}\right)\right) \quad (86)$$

The second moment matrix of $\bar{\bar{\beta}} - \beta_0$ is:

$$E(\bar{\bar{\beta}} - \beta_0)(\bar{\bar{\beta}} - \beta_0)' = V(\hat{\beta}) / \left(1 + \left(\frac{1}{\mathrm{k}}\right)\right)^2 + B(\bar{\bar{\beta}})B(\bar{\bar{\beta}})' \quad (87)$$

with $V(\hat{\beta}) = (X'X)^{-1}\sigma^2$, the covariance matrix of the least squares estimator.
The variance-covariance matrix of $\bar{\bar{\beta}}$ is:

$$E(\bar{\bar{\beta}} - E\bar{\bar{\beta}})(\bar{\bar{\beta}} - E\bar{\bar{\beta}})' = V(\hat{\beta})/\left(1+\left(\frac{1}{k}\right)\right)^2 \qquad (88)$$

The prior and posterior distributions for β and σ using the proposed g-prior, $g = n/k$:
The joint informative g-prior distribution using $g = \frac{n}{k}$ is: $P(\beta, \sigma^2 \mid \eta_0)\ \alpha\ \sigma^{-(v+1)} exp\left\{-\frac{v\bar{\sigma}_a^2}{2\sigma^2}\right\}$ x

$$\sigma^{-k}\, exp\left\{-\left(\frac{n}{k}\right)(\beta-\beta_a)'X'X(\beta-\beta_a)/2\sigma^2\right\} \qquad (89)$$

where $\eta_0' = \left(\beta_a', \bar{\sigma}_a, \left(\frac{n}{k}\right), v\right)$.

The marginal prior distributions for β and σ are respectively:

$$P\left(\beta \mid \beta_a, \left(\frac{n}{k}\right), v\right)\ \alpha\ \left\{v\bar{\sigma}_a^2 + \left(\frac{n}{k}\right)(\beta-\beta_a)'X'X(\beta-\beta_a)\right\}^{-(v+k)/2} \qquad (90)$$

and $P(\sigma \mid \bar{\sigma}_a, V)\ \alpha\ \sigma^{-(v+1)} exp\{-v\bar{\sigma}_a^2/2\sigma^2\}$ (91)
The posterior distribution for β and σ**:**

$$P(\beta,\sigma \mid D)\ \alpha\ P(\beta,\sigma)\, l(\beta,\sigma \mid Y_4) \qquad (92)$$

$$\alpha\ \sigma^{-(n+k+1)} exp\left\{-\left[(Y_4 - X\beta)'(Y_4 - X\beta) + \left(\frac{n}{k}\right)(\beta-\bar{\beta})'X'X(\beta-\bar{\beta})\right]/2\sigma^2\right\}. \qquad (93)$$

D denotes the data and $\bar{\beta}$ is the prior mean vector for regression coefficient vector β.
Let $w_4' = \left(Y_4' : \left(\frac{n}{k}\right)^{\frac{1}{2}} \bar{\beta}' X'\right)$ and $Z_4' = \left(X' : \left(\frac{n}{k}\right)^{\frac{1}{2}} X'\right)$.
Then the terms in square brackets in the exponential can be expressed as

$$(w_4 - Z_4\beta)'(w_4 - Z_4\beta) = (w_4 - Z_4\bar{\bar{\beta}})'(w_4 - Z_4\bar{\bar{\beta}}) + (\beta - \bar{\bar{\beta}})'Z_4'Z_4(\beta - \bar{\bar{\beta}}) \qquad (94)$$

where $\bar{\bar{\beta}} = (Z_4'Z_4)^{-1}Z_4'w_4$.
Thus (93) becomes:

$$P(\beta,\sigma \mid D)\ \alpha\ \sigma^{-(n+k+1)} exp\left\{-\left[(w_4 - Z_4\bar{\bar{\beta}})'(w_4 - Z_4\bar{\bar{\beta}}) + (\beta - \bar{\bar{\beta}})'Z_4'Z_4(\beta - \bar{\bar{\beta}})\right]/2\sigma^2\right\} \qquad (95)$$

where $\bar{\bar{\beta}} = (Z_4'Z_4)^{-1}Z_4'w_4 = \left(\hat{\beta} + \left(\frac{n}{k}\right)\bar{\beta}\right)/\left(1+\left(\frac{n}{k}\right)\right)$.
This is the mean of the posterior distribution with $\hat{\beta} = (X'X)^{-1}X'Y_4$.

The covariance matrix of the conditional normal posterior distribution for β given σ, denoted by $V(\beta \mid \sigma,\ D)$, is $(Z'V(\beta \mid \sigma,\ D) = (Z_4'Z_4)^{-1}\sigma^2)^{-1}\sigma^2$ (96)

$$= (X'X)^{-1}\sigma^2/\left(1+\left(\frac{n}{k}\right)\right) \qquad (97)$$

The marginal posterior distribution for β obtained from (95) by integrating with respect to σ, is:

$$P(\beta \mid D)\ \alpha\ \left\{(w_4 - Z_4\bar{\bar{\beta}})'(w_4 - Z_4\bar{\bar{\beta}}) + (\beta - \bar{\bar{\beta}})'Z_4'Z_4(\beta - \bar{\bar{\beta}})\right\}^{-(n+k)/2} \qquad (98)$$

with covariance matrix:

$$V(\beta \mid D) = (Z_4'Z_4)^{-1}\sigma^2 = (X'X)^{-1}\sigma^2/\left(1+\left(\frac{n}{k}\right)\right) \qquad (99)$$

Also, the marginal posterior distribution for σ obtained from (95) by integrating with respect to β, is:

$$P(\sigma \mid D)\ \alpha\ \sigma^{-(n+1)} exp\left\{-(w_4 - Z_4\bar{\bar{\beta}})'(w_4 - Z_4\bar{\bar{\beta}})/2\sigma^2\right\} \qquad (100)$$

Sampling properties of the mean of the posterior distribution:
The expected mean of $\bar{\bar{\beta}}$ using $g = \frac{n}{k}$ is given by:

$$E\bar{\bar{\beta}} = \left(\beta_0 + \left(\frac{n}{k}\right)\bar{\beta}\right)/\left(1+\left(\frac{n}{k}\right)\right) \qquad (101)$$

where β_0 is the true unknown value of the regression coefficient vector.
The bias of $\bar{\bar{\beta}}$:

$$B(\bar{\bar{\beta}}) = E\bar{\bar{\beta}} - \beta_0 = \left(\frac{n}{k}\right)(\bar{\beta} - \beta_0)/\left(1+\left(\frac{n}{k}\right)\right) \qquad (102)$$

The second moment matrix of $\bar{\bar{\beta}} - \beta_0$ is:

$$E(\bar{\bar{\beta}} - \beta_0)(\bar{\bar{\beta}} - \beta_0)' = V(\hat{\beta})/\left(1+\left(\frac{n}{k}\right)\right)^2 + B(\bar{\bar{\beta}})B(\bar{\bar{\beta}})' \qquad (103)$$

with $V(\hat{\beta}) = (X'X)^{-1}\sigma^2$, the covariance matrix of the least squares estimator.
The variance-covariance matrix of $\bar{\bar{\beta}}$ is:

$$E(\bar{\bar{\beta}} - E\bar{\bar{\beta}})(\bar{\bar{\beta}} - E\bar{\bar{\beta}})' = V(\hat{\beta})/\left(1+\left(\frac{n}{k}\right)\right)^2 \qquad (104)$$

The prior and posterior distributions for β and σ using the proposed g-prior, $g = \sqrt{(n/k)}$:
The joint informative g-prior distribution using $g = \sqrt{\frac{n}{k}}$ is: $P(\beta, \sigma^2 \mid \eta_0)\ \alpha\ \sigma^{-(v+1)} exp\left\{-\frac{v\bar{\sigma}_a^2}{2\sigma^2}\right\}$ x

$$\sigma^{-k}\, exp\left\{-\left(\frac{n}{k}\right)^{\frac{1}{2}}(\beta-\beta_a)'X'X(\beta-\beta_a)/2\sigma^2\right\} \qquad (105)$$

where $\eta_0' = \left(\beta_a', \bar{\sigma}_a, \left(\frac{n}{k}\right)^{\frac{1}{2}}, v\right)$.

The marginal prior distributions for β and σ are respectively:

$$P\left(\beta \mid \beta_a, \left(\frac{n}{k}\right)^{\frac{1}{2}}, v\right)\ \alpha\ \left\{v\bar{\sigma}_a^2 + \left(\frac{n}{k}\right)^{\frac{1}{2}}(\beta-\beta_a)'X'X(\beta-\beta_a)\right\}^{-(v+k)/2} \qquad (106)$$

and $P(\sigma \mid \bar{\sigma}_a, V)\ \alpha\ \sigma^{-(v+1)} exp\{-v\bar{\sigma}_a^2/2\sigma^2\}$ (107)
The posterior distribution for β and σ**:**

$$P(\beta,\sigma \mid D)\ \alpha\ P(\beta,\sigma)\, l(\beta,\sigma \mid Y_5) \qquad (108)$$

$$\alpha\ \sigma^{-(n+k+1)} exp\left\{-\left[(Y_5 - X\beta)'(Y_5 - X\beta) + \left(\frac{n}{k}\right)^{\frac{1}{2}}(\beta-\bar{\beta})'X'X(\beta-\bar{\beta})\right]/2\sigma^2\right\}. \qquad (109)$$

D denotes the data and $\bar{\beta}$ is the prior mean vector for regression coefficient vector β.
Let $w_5' = \left(Y_5' : \left(\frac{n}{k}\right)^{\frac{1}{4}} \bar{\beta}' X'\right)$ and $Z_5' = \left(X' : \left(\frac{n}{k}\right)^{\frac{1}{4}} X'\right)$.

Then the terms in square brackets in the exponential can be expressed as

$$(w_5 - Z_5\beta)'(w_5 - Z_5\beta) = \left(w_5 - Z_5\bar{\bar{\beta}}\right)'\left(w_5 - Z_5\bar{\bar{\beta}}\right) + \left(\beta - \bar{\bar{\beta}}\right)' Z_5' Z_5 \left(\beta - \bar{\bar{\beta}}\right) \quad (110)$$

where $\bar{\bar{\beta}} = (Z_5'Z_5)^{-1}Z_5'w_5$.

Thus (109) becomes:

$$P(\beta, \sigma \mid D) \;\alpha\; \sigma^{-(n+k+1)} exp\left\{-\left[\left(w_5 - Z_5\bar{\bar{\beta}}\right)'\left(w_5 - Z_5\bar{\bar{\beta}}\right) + \left(\beta - \bar{\bar{\beta}}\right)' Z_5'Z_5\left(\beta - \bar{\bar{\beta}}\right)\right] / 2\sigma^2\right\} \quad (111)$$

where $\bar{\bar{\beta}} = (Z_5'Z_5)^{-1}Z_5'w_5 = \left(\hat{\beta} + \left(\frac{\mathrm{n}}{\mathrm{k}}\right)^{\frac{1}{2}} \bar{\beta}\right) / \left(1 + \left(\frac{\mathrm{n}}{\mathrm{k}}\right)^{\frac{1}{2}}\right)$.

This is the mean of the posterior distribution with $\hat{\beta} = (X'X)^{-1}X'Y_5$.

The covariance matrix of the conditional normal posterior distribution for β given σ, denoted by $V(\beta \mid \sigma,\ D)$, is $V(\beta \mid \sigma,\ D) = (Z_5'Z_5)^{-1}\sigma^2$ (112)

$$= (X'X)^{-1}\sigma^2 / \left(1 + \left(\frac{\mathrm{n}}{\mathrm{k}}\right)^{\frac{1}{2}}\right) \quad (113)$$

The marginal posterior distribution for β obtained from (109) by integrating with respect to σ, is

$$P(\beta \mid D) \;\alpha\; \left\{\left(w_5 - Z_5\bar{\bar{\beta}}\right)'\left(w_5 - Z_5\bar{\bar{\beta}}\right) + \left(\beta - \bar{\bar{\beta}}\right)' Z_5'Z_5\left(\beta - \bar{\bar{\beta}}\right)\right\}^{-(n+k)/2} \quad (114)$$

with covariance matrix:

$$V(\beta \mid D) = (Z_5'Z_5)^{-1}\sigma^2 = (X'X)^{-1}\sigma^2 / \left(1 + \left(\frac{\mathrm{n}}{\mathrm{k}}\right)^{\frac{1}{2}}\right) \quad (115)$$

Also, the marginal posterior distribution for σ obtained from (109) by integrating with respect to β, is $P(\sigma \mid D) \;\alpha\; \sigma^{-(n+1)} exp\left\{-\left(w_5 - Z_5\bar{\bar{\beta}}\right)'\left(w_5 - Z_5\bar{\bar{\beta}}\right) / 2\sigma^2\right\}$ (116)

Sampling properties of the mean of the posterior distribution:

The expected mean of $\bar{\bar{\beta}}$ using $\mathrm{g} = \sqrt{\frac{\mathrm{n}}{\mathrm{k}}}$ is given by:

$$E\bar{\bar{\beta}} = \left(\beta_0 + \left(\frac{n}{k}\right)^{\frac{1}{2}} \bar{\beta}\right) / \left(1 + \left(\frac{n}{k}\right)^{\frac{1}{2}}\right) \quad (117)$$

where β_0 is the true unknown value of the regression coefficient vector.

The bias of $\bar{\bar{\beta}}$:

$$B(\bar{\bar{\beta}}) = E\bar{\bar{\beta}} - \beta_0 = \left(\frac{n}{k}\right)^{\frac{1}{2}} (\bar{\beta} - \beta_0) / \left(1 + \left(\frac{n}{k}\right)^{\frac{1}{2}}\right) \quad (118)$$

The second moment matrix of $\bar{\bar{\beta}} - \beta_0$ is:

$$E(\bar{\bar{\beta}} - \beta_0)(\bar{\bar{\beta}} - \beta_0)' = \frac{V(\hat{\beta})}{\left(1 + \left(\frac{\mathrm{n}}{\mathrm{k}}\right)^{\frac{1}{2}}\right)^2} + B(\bar{\bar{\beta}})B(\bar{\bar{\beta}})'. \quad (119)$$

with $V(\hat{\beta}) = (X'X)^{-1}\sigma^2$, the covariance matrix of the least squares estimator.

The variance-covariance matrix of $\bar{\bar{\beta}}$ is:

$$E(\bar{\bar{\beta}} - E\bar{\bar{\beta}})(\bar{\bar{\beta}} - E\bar{\bar{\beta}})' = V(\hat{\beta}) / \left(1 + \left(\frac{\mathrm{n}}{\mathrm{k}}\right)^{\frac{1}{2}}\right)^2 \quad (120)$$

The prior and posterior distributions for β and σ using the proposed g-prior, $\mathrm{g} = \frac{n}{k^2}$

The joint informative g-prior distribution using $\mathrm{g} = \frac{n}{k^2}$ is:

$$P(\beta, \sigma^2 \mid \eta_0) \;\alpha\; \sigma^{-(v+1)} exp\left\{-\frac{v\bar{\sigma}_a^2}{2\sigma^2}\right\} \quad x \quad \sigma^{-k}\, exp\left\{-\left(\frac{\mathrm{n}}{\mathrm{k}^2}\right)(\beta - \beta_a)'X'X(\beta - \beta_a)/2\sigma^2\right\} \quad (121)$$

where $\eta_0' = \left(\beta_a', \bar{\sigma}_a, \left(\frac{\mathrm{n}}{\mathrm{k}^2}\right), v\right)$.

The marginal prior distributions for β and σ are respectively:

$$P\left(\beta \mid \beta_a, \left(\frac{n}{\mathrm{k}^2}\right), v\right) \;\alpha\; \left\{v\bar{\sigma}_a^2 + \left(\frac{n}{\mathrm{k}^2}\right)(\beta - \beta_a)'X'X(\beta - \beta_a)\right\}^{-(v+k)/2} \quad (122)$$

and $\quad P(\sigma \mid \bar{\sigma}_a, V) \;\alpha\; \sigma^{-(v+1)} exp\{-v\bar{\sigma}_a^2/2\sigma^2\}$ (123)

The posterior distribution for β and σ**:**

$$P(\beta, \sigma \mid D) \;\alpha\; P(\beta, \sigma)\; l(\beta, \sigma \mid Y_6) \quad (124)$$

$$\alpha\; \sigma^{-(n+k+1)} exp\left\{-\left[(Y_6 - X\beta)'(Y_6 - X\beta) + \left(\frac{n}{k^2}\right)(\beta - \bar{\beta})'X'X(\beta - \bar{\beta})\right] / 2\sigma^2\right\}. \quad (125)$$

D denotes the data and $\bar{\beta}$ is the prior mean vector for regression coefficient vector β.

Let $w_6' = \left(Y_6' : \left(\frac{\mathrm{n}}{\mathrm{k}^2}\right)^{\frac{1}{2}} \bar{\beta}' \mathrm{X}'\right)$ and $Z_6' = \left(X' : \left(\frac{\mathrm{n}}{\mathrm{k}^2}\right)^{\frac{1}{2}} \mathrm{X}'\right)$.

Then the terms in square brackets in the exponential can be expressed as

$$(w_6 - Z_6\beta)'(w_6 - Z_6\beta) = \left(w_6 - Z_6\bar{\bar{\beta}}\right)'\left(w_6 - Z_6\bar{\bar{\beta}}\right) + \left(\beta - \bar{\bar{\beta}}\right)' Z_6'Z_6\left(\beta - \bar{\bar{\beta}}\right) \quad (126)$$

where $\bar{\bar{\beta}} = (Z_6'Z_6)^{-1}Z_6'w_6$.

Thus (125) becomes:

$$P(\beta, \sigma \mid D) \;\alpha\; \sigma^{-(n+k+1)} exp\left\{-\left[\left(w_6 - Z_6\bar{\bar{\beta}}\right)'\left(w_6 - Z_6\bar{\bar{\beta}}\right) + \left(\beta - \bar{\bar{\beta}}\right)' Z_6'Z_6\left(\beta - \bar{\bar{\beta}}\right)\right] / 2\sigma^2\right\} \quad (127)$$

where $\bar{\bar{\beta}} = (Z_6'Z_6)^{-1}Z_6'w_6 = \left(\hat{\beta} + \left(\frac{\mathrm{n}}{\mathrm{k}^2}\right)\bar{\beta}\right) / \left(1 + \left(\frac{\mathrm{n}}{\mathrm{k}^2}\right)\right)$.

This is the mean of the posterior distribution with $\hat{\beta} = (X'X)^{-1}X'Y_6$.

The covariance matrix of the conditional normal posterior distribution for β given σ, denoted by $V(\beta \mid \sigma,\ D)$, is $V(\beta \mid \sigma,\ D) = (Z_6'Z_6)^{-1}\sigma^2$ (128)

$$= (X'X)^{-1}\sigma^2 / \left(1 + \left(\frac{\mathrm{n}}{\mathrm{k}^2}\right)\right) \quad (129)$$

The marginal posterior distribution for β obtained from (125) by integrating with respect to σ, is

$$P(\beta \mid D) \;\alpha\; \left\{\left(w_6 - Z_6\bar{\bar{\beta}}\right)'\left(w_6 - Z_6\bar{\bar{\beta}}\right) + \left(\beta - \bar{\bar{\beta}}\right)' Z_6'Z_6\left(\beta - \bar{\bar{\beta}}\right)\right\}^{-(n+k)/2} \quad (130)$$

with covariance matrix:

$$V(\beta \mid D) = (Z_6'Z_6)^{-1}\sigma^2 = (X'X)^{-1}\sigma^2 / \left(1 + \left(\frac{\mathrm{n}}{\mathrm{k}^2}\right)\right) \quad (131)$$

Also, the marginal posterior distribution for σ obtained from (125) by integrating with respect to β, is $P(\sigma \mid D) \;\alpha\; \sigma^{-(n+1)} exp\left\{-\left(w_6 - Z_6\bar{\bar{\beta}}\right)'\left(w_6 - Z_6\bar{\bar{\beta}}\right) / 2\sigma^2\right\}$ (132)

Sampling properties of the mean of the posterior distribution:

The expected mean of $\bar{\bar{\beta}}$ using $g = \frac{n}{k^2}$ is given by:

$$E\bar{\bar{\beta}} = \left(\beta_0 + \left(\frac{n}{k^2}\right)\bar{\beta}\right) / \left(1 + \left(\frac{n}{k^2}\right)\right) \qquad (133)$$

where β_0 is the true unknown value of the regression coefficient vector.

The bias of $\bar{\bar{\beta}}$:

$$B(\bar{\bar{\beta}}) = E\bar{\bar{\beta}} - \beta_0 = \left(\frac{n}{k^2}\right)(\bar{\beta} - \beta_0) / \left(1 + \left(\frac{n}{k^2}\right)\right) \qquad (134)$$

The second moment matrix of $\bar{\bar{\beta}} - \beta_0$ is:

$$E(\bar{\bar{\beta}} - \beta_0)(\bar{\bar{\beta}} - \beta_0)' = V(\hat{\beta}) / \left(1 + \left(\frac{n}{k^2}\right)\right)^2 + B(\bar{\bar{\beta}})B(\bar{\bar{\beta}})'. \qquad (103)$$

with $V(\hat{\beta}) = (X'X)^{-1}\sigma^2$, the covariance matrix of the least squares estimator.

The variance-covariance matrix of $\bar{\bar{\beta}}$ is:

$$E(\bar{\bar{\beta}} - E\bar{\bar{\beta}})(\bar{\bar{\beta}} - E\bar{\bar{\beta}})' = V(\hat{\beta}) / \left(1 + \left(\frac{n}{k^2}\right)\right)^2 \qquad (134)$$

The effects of the set of *g*-priors were examined using datasets provided by FLS (Fernandez, Ley and Steel, 2001a). The analysis was based on *n* = 72 observations with *k* = 41 set of regressors or possible variables. To analyse these data, uniform model prior was applied as the model prior for the model space across parameter *g*-prior structures investigated.

Table 2: The Effects of *g*-Prior Structures on Predicted Values compared with its Actual Value

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

Given the model space $2^{41} = 2.2 \; x \; 10^{12}$ and with a fairly large amount of drawings (5 million), Markov Chain Monte Carlo Model Composition (MC^3) sampler was applied to adequately identify the high posterior probability models. The Bayesian analysis was carried out using Bayesian model averaging package "BMS" available in the statistical software R. The study compares the predictive abilities of different *g*-priors from the Literature (see Table 1) and the 6 proposed *g*-prior structures investigated. Table 2 shows the results for the 11 *g*-prior structures (Table 1) and the most reliable and consistent *g*-prior structure among the proposed 6 *g*-prior structures.

The log predictive score (LPS) is a scoring rule for assessing predictive performance and a smaller value of LPS makes a Bayes model a prior choice for *g* that is preferable (Kadane and Lazar, 2004).

The results from Table 2 show that the actual value of the dependent variable of the 72nd observation is best predicted by a new *g*-prior, $g = n/\sqrt{k}$ based on the predicted and actual values and having one of the lowest LPS, though preceded by *g*-prior serial number (S/N: 1, 2 and 3).

Conclusion

The study investigated 11 *g*-priors identified in the Literature and 6 new *g*-priors proposed. The reliability and consistency of the predictive ability of a new *g*-prior structure, $g = n/\sqrt{k}$ *vis-à-vis* the actual value of the dependent variables, predicted values of the dependent variables and corresponding LPS was demonstrated.

Based on the framework and methods enumerated, the prior distributions and posterior distributions of the regression parameters were obtained for the proposed 6 *g*-prior structures investigated. The sampling properties in term of the expected mean and variance of the posterior distribution were obtained for the respective *g*-prior structures investigated. The new proposed *g*-prior, $g = n/\sqrt{k}$, exhibited a reliable, consistent, competitive and predictive performance and offers a sound, fully Bayesian approach that features the virtues of prior input and predictive gains that minimises the risk of misspecification.

References

Adeleke, I. A. and Ogundeji, R. K. (2009) Bayesian Estimation of the Proportion of Subscribers to Nigeria's National Health Insurance Scheme. *Journal of Modern Mathematics and Statistics,* **3**, (3): 56 – 59.

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**(3): 271–275.

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. 578pp.

Eicher, T., Papageorgiou, C. and 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, 48pp.

Eicher T., Papageorgiou, C. and Raftery A. (2011) Default Priors and Predictive Performance in Bayesian Model Averaging, with Application to Growth Determinants. *Journal of Applied Econometrics.* **26**: 30–55.

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

Fouskakis, D. and Ntzowfras I. (2013). Power Conditional-Expected Prior: Using g-Priors with Random Imaginary Data for Variable Selection. Cornell University Library, arXiv:1307.244V1 [Stat. CO].

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.

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.

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

Hoeting, J. A., Madigan, D., Raftery, A. E., and Volinsky, C. T. (1999) "Bayesian Model Averaging: A Tutorial (with discussion)," *Statistical Science*, **14**: 382–401 (corrected version at http://www.stat.washington.edu/research/online/hoeting1999.pdf).

Kadane, J. B. and Lazar, N. A. (2004) Methods and Criteria for Model Selection, *Journal of the American Statistical Association,* **99**(65): 279–290.

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, 268pp.

Li, Y. and Clyde, M. (2015) Mixtures of g-priors in Generalised Linear Models. Cornell University Library, axXiv: 1503.06913V3 [Stat. ME].

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.

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

Olubusoye O. E. and Akanbi O. B. (2015) On g-Prior Elicitation in Bayesian Model Averaging Approach to Normal Linear Regression Model. Conference Proceedings on Perspectives and Developments in Mathematics, p. 147–170.

Raftery, A. E., Karny, M. and Ettler, P. (2010) "Online Prediction Under Model Uncertainty via Dynamic Model Averaging: Application to a Cold Rolling Mill," *Technometrics* **52**: 52–66.

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.

Rossi, P. E, Allenby, G. M, and Mcculloch, R. (2005) Bayesian Statistics and Marketing. Wiley Series in Probability and Statistics, 160pp.

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), **32**(2): 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),* p. 1969–1976.