Network for Research and Development in Africa
International Journal of Pure and Applied Science Research
ISSN: 2384-5918. Volume 11, Number 5
Pages 39-51 (September, 2019)
www.arcnjournals.org

Bayessian Regression Technique For Modeling Multicolinear Data

Danyaro, M.L.; Mohammed, A. Zakar; Mohammed Audu; and Bagare, S.A.
Department of Basic Science, College of Agriculture Gujba, Yobe State Nigeria

Abstract: *This paper examined the behaviours of two frequentist regression methods (the ridge regression and ordinary least squares (OLS)) and Bayesian linear regression method on data with inherent collinear structure. Data sets with reasonable degrees of multicollinearity at some selected sample sizes were simulated. The three regression types were fitted to various data and the performances of both the ridge and OLS estimators were compared with that of the Bayesian linear regression estimators using Normal-Gamma conjugate prior. The goal is to examine the relative efficiency of the Bayesian estimator, which integrates some prior information with the information available in the data in its regression estimation, over the two frequentist regression techniques. Results from Monte Carlo studies established the supremacy of the Bayesian estimators over both the OLS and the ridge estimators. Although, the ridge regression estimators expectedly performs better than the OLS estimators given the degree of multicollinearity in the simulated data, the results generally showed that Bayesian linear regression estimator is relatively more efficient (with smaller mean square errors) than the two frequentist regression techniques given the same data structure.*

Keywords*: Bayessian, Modeling, Multicolinear, Regression, and Simulation*

1.0 Introduction

In statistical inference, there are two broad categories of interpretations of probability: Bayesian inference (Byron Hall, STATISTICAT, LLC, Bayesian Inference Article p. 1-2) and frequentist inference (Byron Hall, STATISTICAT, LLC, and Bayesian Inference Article). These views often differ with each other on the fundamental nature of probability. Frequentist inference loosely defines probability as the limit of an event's relative frequency in a large number of trials, and only in the context of experiments that are random and well-defined. Bayesian inference, on the other hand, is able to assign probabilities to any statement, even when a random process is not involved. In Bayesian inference, probability is a way to represent an individual's degree of belief in a statement, or given evidence. Within Bayesian inference, there are also different interpretations of probability, and different approaches evolved based on those interpretations. The most popular interpretations and approaches are objective Bayesian inference (Berger 2006) and subjective Bayesian inference (Anscombe and Aumann

1963; *Bernardo* 2008). Objective Bayesian inference is often associated with Bayes and Price (1763), Laplace (1814), and *Roberts* (2007). Subjective Bayesian inference is often associated with Ramsey (1926), *Simon*, (2009), and *Bernardo and Smith, (2000)*.

1.1 Bayes' Theorem

Bayes' theorem shows the relation between two conditional probabilities that are the reverse of each other. This theorem is named after Reverend Thomas Bayes (1702-1761), and is also referred to as Bayes' law or Bayes' rule (Bayes and Price 1763). Bayes' theorem expresses the conditional probability, or `posterior probability', of an event A after B is observed in terms of the 'prior probability' of A, prior probability of B, and the conditional probability of B given A. Bayes' theorem is valid in all common interpretations of probability.

When no data are available, a *prior distribution* is used to quantify our knowledge about the parameter. When data are available, we can update our prior knowledge using the conditional distribution of parameters, given the data. The transition from the prior to the posterior is possible via the Bayes theorem. Suppose that before the experiment our prior distribution describing Aisπ (*A*)*:* The data are coming from the assumed model (likelihood) which depends on the parameter and is denoted by *f(x/A).*Bayes theorem updates the prior,$\pi\ (A)$ to the posterior by accounting for the data *x* through the relationship.

$$\pi(A/x) = \frac{f(x/A)\pi(A)}{m(x)} \qquad (1.1)$$

Where *m*(*x*) is a normalizing constant, $m(x) = \int f(x/A)\pi(A)\, dA$ (1.2)

Once the data *x* are available, *A* is the only unknown quantity and the posterior distribution π(*A/x*) completely describes the uncertainty. There are two key advantages of Bayesian paradigm: (i) once the uncertainty is expressed via the probability distribution and the statistical inference can be automated, it follows a conceptually simple recipe, and (ii) available prior information is coherently incorporated into the statistical model.

1.2 Bayesian Linear Regression

In statistics, **Bayesian linear regression** is an approach to linear regression in which the statistical analysis is undertaken within the context of Bayesian inference. When the regression model has errors that have a normal distribution, and if a particular form of prior distribution is assumed, explicit results are available for the posterior probability distributions of the model's parameters. Consider a standard linear regression problem, in which for$i = 1, \dots, n$ we specify the conditional distribution of$y_i$given a *1×k* predictor vector$\boldsymbol{x_i}$:

$$y_i = \boldsymbol{x_i}^T\boldsymbol{\beta} + \epsilon_i \qquad (1.3)$$

where $\boldsymbol{\beta}$ is a*k×1* vector of regression parameters to be estimated, and the$\epsilon_i$is independent and identically-distributed and normally distributed random error of the model with $\epsilon_i \sim N(0, \sigma^2)$. From the regression model in (1.2), the following likelihood function is developed within the Bayesian concept:

$$p(\boldsymbol{y}|\boldsymbol{X}, \boldsymbol{\beta}, \sigma^2) \propto (\sigma^2)^{\frac{-n}{2}} exp\left(-\frac{1}{2(\sigma^2)}(\boldsymbol{y} - \boldsymbol{X\beta})^T(\boldsymbol{y} - \boldsymbol{X\beta})\right) \qquad (1.4)$$

The ordinary least squares solution is to estimate the coefficient vector using the Moore-Penrose pseudo-inverse:

$$\hat{\beta} = (X^T X)^{-1}(X^T Y) \quad (1.5)$$

Where **X** is the ***n×k*** design matrix of predictor variables, each row of which is a predictor vector $\boldsymbol{x_i}^T$; and **y** is the column n-vector $(\boldsymbol{y_1}, \ldots, \boldsymbol{y_n})^T$.

This is a frequentist approach, and it assumes that there are enough measurements (samples) to say something meaningful about $\boldsymbol{\beta}$. In the Bayesian approach, the data are supplemented with additional information in the form of a prior probability distribution. The prior belief about the parameters is combined with the data's likelihood function according to Bayes theorem to yield the posterior belief about the parameters $\boldsymbol{\beta}$ and σ. The prior can take different functional forms depending on the domain and the information that is available a priori.

2.1 Linear Regression Model

The linear regression model is use to study the relationship between a dependent variable and one or more independently variables. The generic form of the linear regression model is

$$y = f(x_1, x_2, \ldots, x_k) + \varepsilon \quad (2.1)$$
$$= \beta_o + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_k x_k + \varepsilon \quad (2.2)$$

Where y is the dependent variable or explained variable and $x_1, \ldots, x_k$ are the independent or explanatory variables. The function $f(x_1, \ldots, x_k)$ is commonly called population regression equation of y on $x_1, \ldots, x_k$. In this setting, y is the regress and x_k, k=1 … K, are the regressors or covariates. The term ε is a random disturbance so named because it "disturbs" an otherwise stable relationship. (Greene, 2000)

2.1.2 Normal Linear Regression Model

The above equation can be expressed in matrix form as a normal linear regression model and thus the model is given as:

$$y = X\beta + \epsilon \quad (2.3)$$

Where $y = (y_1, \ldots, y_n)^T, \beta = (\beta_1, \ldots, \beta_j)^T$ and assuming initially that $\epsilon \sim N(0, \sigma^2)$

The addition of the assumption of normality of ϵ leads to normal linear regression model. (Gujarati (2004).

2.1.3 Linear Regression Model in Matrixnotation

Suppose we have data on a dependent variable, y_i, and k explanatory variables $x_{i1}, \ldots, x_{ik}$ for $i = 1, \ldots, n$. The linear regression model is given by:

$$y_i = \beta_1 + \beta_2 x_{i2} + \cdots + \beta_k x_{ik} + \epsilon_i \quad (2.4)$$

The above notation is such that x_{i1} is implicitly set to 1 to allow for an intercept. This model can be written more compactly in matrix notation by defining the $n \times 1$ vectors:

$$y = \begin{bmatrix} y_1 \\ y_2 \\ . \\ . \\ . \\ y_n \end{bmatrix}$$

and

$$\epsilon = \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ . \\ . \\ . \\ \epsilon_n \end{bmatrix}$$

the $k \times 11$

=β_1β_2...β_k
and the $n \times k$ matrix

$$X = \begin{bmatrix} 1 x_{12} \cdots x_{1k} \\ 1 x_{22} \cdots x_{2k} \\ . \quad . \quad \cdots \quad . \\ . \quad . \quad \cdots \quad . \\ . \quad . \quad \cdots \quad . \\ . \quad . \quad \cdots \quad . \\ 1 x_{n2} \cdots x_{nk} \end{bmatrix}$$

And writing

$$y = X\beta + \epsilon \qquad (2.5)$$

Using the definition of matrix multiplication it can be verified that (2) is equivalent to the n equations defined by (1) (Koop. 2003),

3.0 Methodology

3.1 Bayesian Method

The paper presents the posterior distribution of the Bayesian normal linear regression, properties of the parameters of interest with respect to their credible intervals and highest posterior densities.

3.1.1 The Posterior Distribution of Bayesian Normal Linear Regression

The posterior distribution $p(\beta, h|y)$ defined as:

$$p(\beta, h|y) \propto p(\beta, h) \times p(y|\beta, h) \quad (3.1)$$

Where $p(\beta, h) = \frac{h^{\frac{v_0+k}{2}-1}}{2\pi^{\frac{k}{2}}|\sum_0|^{\frac{1}{2}}\Gamma\frac{v_0}{2}\left(\frac{2s_0{}^{-2}}{v_0}\right)^{\frac{v_0}{2}}}\left\{exp\left[\frac{-h}{2}(\beta-\beta_0)^T(\sum_0)^{-1}(\beta-\beta_0)+\frac{v_0}{s_0{}^{-2}}\right]\right\}$ (3.2)

And

$$p(y|\beta, h) = \frac{h^{\frac{n}{2}}}{(2\pi)^{\frac{n}{2}}}\left\{exp\left[\frac{-h}{2}(vs^2+(b-\beta)^T(X^TX)(b-\beta))\right]\right\} \quad (3.3)$$

Thus

$$p(\beta, h|y) \propto \frac{h^{\frac{v_0+k}{2}-1}}{2\pi^{\frac{k}{2}}|\sum_0|^{\frac{1}{2}}\Gamma\frac{v_0}{2}\left(\frac{2s_0{}^{-2}}{v_0}\right)^{\frac{v_0}{2}}}\left\{exp\left[\frac{-h}{2}(\beta-\beta_0)^T(\sum_0)^{-1}(\beta-\beta_0)+\frac{v_0}{s_0{}^{-2}}\right]\right\}\times$$

$$\frac{h^{\frac{n}{2}}}{(2\pi)^{\frac{n}{2}}}\left\{exp\left[\frac{-h}{2}(vs^2+(b-\beta)^T(X^TX)(b-\beta))\right]\right\} \quad (3.4)$$

Let $c = \frac{1}{(2\pi)^{\frac{k+n}{2}}|\sum_0|^{\frac{1}{2}}\Gamma\frac{v_0}{2}\left(\frac{2s_0{}^{-2}}{v_0}\right)^{\frac{v_0}{2}}}$

Then

$$p(\beta, h|y) \propto h^{\frac{v_0+n+k}{2}-1}\left\{exp\left[\frac{-h}{2}\left((\beta-\beta_0)^T(\sum_0)^{-1}(\beta-\beta_0)+\frac{v_0}{s_0{}^{-2}}+vs^2+(b-\beta)^T(X^TX)(b-\beta)\right)\right]\right\} \quad (3.5)$$

Expanding the terms within the exponent bracket, we have:

$$(\beta-\beta_0)^T(\sum_0)^{-1}(\beta-\beta_0)+(b-\beta)^T(X^TX)(b-\beta) = \quad (3.6)$$

$$\beta^T(\sum_0)^{-1}\beta-\beta^T(\sum_0)^{-1}\beta_0-\beta_0{}^T(\sum_0)^{-1}\beta+\beta_0{}^T(\sum_0)^{-1}\beta_0+b^T(X^TX)b-b^T(X^TX)\beta-$$
$$\beta^T(X^TX)b+\beta^T(X^TX)\beta= \quad (3.7)$$

$$\beta^T[(\sum_0)^{-1}+X^TX]\beta-\beta^T[(\sum_0)^{-1}\beta_0+(X^TX)b]-\left[\beta_0{}^T(\sum_0)^{-1}+b^T(X^TX)\right]\beta+\beta_0{}^T(\sum_0)^{-1}\beta_0+$$
$$b^T(X^TX)b = \quad (3.8)$$

If we let

$$\beta^* = \sum^*(\sum_0{}^{-1}\beta_0+X^TXb)$$
$$\sum^* = (\sum_0{}^{-1}+X^TX)^{-1}$$
$$\beta^T(\sum^*)^{-1}\beta-\beta^T(\sum^*)^{-1}\sum^*[(\sum_0)^{-1}\beta_0+(X^TX)b]-(\sum^*)^{-1}\sum^*\left[\beta_0{}^T(\sum_0)^{-1}+b^T(X^TX)\right]\beta$$
$$+\beta_0{}^T(\sum_0)^{-1}\beta_0+b^T(X^TX)b =$$
$$\beta^T(\sum^*)^{-1}\beta-\beta^T(\sum^*)^{-1}\beta^*-\beta^{*T}(\sum^*)^{-1}\beta+\beta_0{}^T(\sum_0)^{-1}\beta_0+b^T(X^TX)b$$

Simplifying further:

$$\beta^T(\sum^*)^{-1}\beta - \beta^T(\sum^*)^{-1}\beta^* - \beta^{*T}(\sum^*)^{-1}\beta + \beta_0{}^T(\sum_0)^{-1}\beta_0 + b^T(X^TX)b =$$
$$\beta^T(\sum^*)^{-1}\beta - \beta^T(\sum^*)^{-1}\beta^* - \beta^{*T}(\sum^*)^{-1}\beta^* + \beta^{*T}(\sum^*)^{-1}\beta^* - \beta^{*T}(\sum^*)^{-1}\beta + \beta_0{}^T(\sum_0)^{-1}\beta_0 + b^T(X^TX)b =$$
$$(\beta-\beta^*)^T(\sum^*)^{-1}(\beta-\beta^*) - \beta^{*T}(\sum^*)^{-1}\beta + \beta_0{}^T(\sum_0)^{-1}\beta_0 + b^T(X^TX)b \quad (3.9)$$

The last three terms can be further combined to yield:

$$\beta_0{}^T(\sum_0)^{-1}\beta_0 + b^T(X^TX)b - \beta^{*T}(\sum^*)^{-1}\beta = (b-\beta_0)^T\left[\sum_0{}^{-1} + (X^TX)^{-1}\right]^{-1}(b-\beta_0) \quad (3.10)$$

Thus the posterior can be written as:

$$p(\beta,h|y)$$
$$\propto h^{\frac{v_0+n+k}{2}-1}\left\{exp\left[\frac{-h}{2}\left((\beta-\beta^*)^T(\sum^*)^{-1}(\beta-\beta^*) + v_0s_0{}^2 + vs + (b-\beta_0)^T\left[\sum_0{}^{-1} + (X^TX)^{-1}\right]^{-1}(b-\beta_0)\right)\right]\right\} =$$
$$\propto h^{\frac{v_0+n+k}{2}-1}\left\{exp\left[\frac{-h}{2}((\beta-\beta^*)^T(\sum^*)^{-1}(\beta-\beta^*))\right]\right\} * exp\left[\frac{-h}{2}\left(v_0s_0{}^2 + vs + (b-\beta_0)^T\left[\sum_0{}^{-1} + (X^TX)^{-1}\right]^{-1}(b-\beta_0)\right)\right]$$

$$\propto h^{\frac{k}{2}}\left\{exp\left[\frac{-h}{2}((\beta-\beta^*)^T(\sum^*)^{-1}(\beta-\beta^*))\right]\right\} * h^{\frac{v^*}{2}-1}exp\left[\frac{-hv^*}{2s^{-2*}}\right] \quad (3.12)$$

The above is indeed the kernel of a Normal-gamma distribution.

Therefore

$$\beta,h|y \sim NG\left(\beta^*,\sum{}^*,v^*,s^{-2^*}\right) \quad (3.13)$$

Where

$$\beta^* = \sum^*(\sum_0{}^{-1}\beta_0 + X^TXb) \quad (3.14)$$
$$\sum^* = (\sum_0{}^{-1} + X^TX)^{-1} \quad (3.15)$$
$$v^* = v_0 + n \quad (3.16)$$
$$s^{-2^*} = \frac{v^*}{v_0s_0{}^2 + vs + (b-\beta_0)^T(\sum_0{}^{-1} + (X^TX)^{-1})^{-1}(b-\beta_0)} \quad (3.17)$$
$$v^*s^{2*} = v_0s_0{}^2 + vs + (b-\beta_0)^T(\sum_0{}^{-1} + (X^TX)^{-1})^{-1}(b-\beta_0) \quad (3.18)$$

3.1.2 Properties of the posterior parameters

3.1.2.1 Marginal distribution of $\boldsymbol{\beta}$

In this setting h is not of immediate interest and is therefore considered as nuisance parameter. It follows that h has to be integrated out to get the marginal distribution of $\boldsymbol{\beta}^*$:

$$p(\beta|y) = \int_0^\infty p(\beta,h|y)d(h) \quad (3.19)$$
$$= \int_0^\infty \frac{1}{(2\pi)^{\frac{k+n}{2}}|\sum_0|^{\frac{1}{2}}\Gamma_{\frac{v_0}{2}}\left(\frac{2s_0{}^{-2}}{v_0}\right)^{\frac{v_0}{2}}}h^{\frac{v_0+n+k}{2}-1}\left\{exp\left[\frac{-h}{2}\left((\beta-\beta^*)^T(\sum^*)^{-1}(\beta-\beta^*) + v_0s_0{}^2 + vs + (b-\beta_0)^T\left[\sum_0{}^{-1} + (X^TX)^{-1}\right]^{-1}(b-\beta_0)\right)\right]\right\}dh$$

$$= \frac{1}{(2\pi)^{\frac{k+n}{2}}|\sum_0|^{\frac{1}{2}}\Gamma_{\frac{v_0}{2}}\left(\frac{2s_0{}^{-2}}{v_0}\right)^{\frac{v_0}{2}}}\int_0^{\infty} h^{\frac{v_0+n+k}{2}-1}\left\{exp\left[\frac{-h}{2}\left((\beta-\beta^*)^T(\sum^*)^{-1}(\beta-\beta^*)+v_0{s_0}^2+vs\right.\right.\right.$$
$$\left.\left.\left.+(b-\beta_0)^T\left[\sum_0{}^{-1}+(X^TX)^{-1}\right]^{-1}(b-\beta_0)\right)\right]\right\}dh$$
$$= \frac{1}{(2\pi)^{\frac{k+n}{2}}|\sum_0|^{\frac{1}{2}}\Gamma_{\frac{v_0}{2}}\left(\frac{2s_0{}^{-2}}{v_0}\right)^{\frac{v_0}{2}}}\int_0^{\infty} h^{\frac{v*+k}{2}-1}\left\{exp\left[\frac{-h}{2}((\beta-\beta^*)^T(\sum^*)^{-1}(\beta-\beta^*)+v^*s^{2*})\right]\right\}dh$$

(3.20)

Using integration by substitution

Let $m = \frac{h(\beta-\beta^*)^T(\sum^*)^{-1}(\beta-\beta^*)+v^*s^{2*}}{2}$ then

$$\frac{dm}{dh} = \frac{(\beta-\beta^*)^T(\sum^*)^{-1}(\beta-\beta^*)+v^*s^{2*}}{2}$$

Therefore:

$$= c\int_0^{\infty}\{exp(-m)\}\left(\frac{2m}{(\beta-\beta^*)^T(\sum^*)^{-1}(\beta-\beta^*)+v^*s^{2*}}\right)^{\frac{v^*+k}{2}-1}\frac{2dm}{(\beta-\beta^*)^T(\sum^*)^{-1}(\beta-\beta^*)+v^*s^{2*}}$$
$$= c\left(\frac{2}{(\beta-\beta^*)^T(\sum^*)^{-1}(\beta-\beta^*)+v^*s^{2*}}\right)^{\frac{v^*+k}{2}}\left[\int_0^{\infty} m^{\frac{v^*+k}{2}-1}\{exp(-m)\}dm\right]$$

Recall from gamma function that

$$\left[\int_0^{\infty} m^{\frac{v'+k}{2}-1}\{exp(-m)\}dm\right] = \Gamma_{\frac{v^*+k}{2}}$$

Thus

$$= c\left(\frac{2}{(\beta-\beta^*)^T(\sum^*)^{-1}(\beta-\beta^*)+v^*s^{2*}}\right)^{\frac{v^*+k}{2}}\Gamma_{\frac{v^*+k}{2}}$$
$$= \frac{\Gamma_{\frac{v^*+k}{2}}}{(2\pi)^{\frac{k+n}{2}}|\sum_0|^{\frac{1}{2}}\Gamma_{\frac{v_0}{2}}\left(\frac{2s_0{}^{-2}}{v_0}\right)^{\frac{v_0}{2}}}\left(\frac{(\beta-\beta^*)^T(\sum^*)^{-1}(\beta-\beta^*)+v^*s^{2*}}{2}\right)^{-\frac{v^*+k}{2}} \quad (3.21)$$

Hence $p(\boldsymbol{\beta}|\boldsymbol{y})$follows the multivariate t-distribution defines as follows:

$(\boldsymbol{\beta}|\boldsymbol{y})\sim\boldsymbol{t}(\beta^*, s^{2*}\sum^*, v^*)$ (3.22)

$\boldsymbol{E}(\beta) = \beta^*$ (3.23)

$\boldsymbol{var}(\beta) = \frac{v^*s^{2*}}{v^*-2}\sum^*$ (3.24)

3.1.2.2 Marginal distribution of $\boldsymbol{h}$

To derive the marginal posterior density for *h*, we can use the "reversed" version of bayes rule:

$p(h|y) = \frac{p(\beta,h|y)}{p(\beta|h,y)}$ (3.25)

The above follow from the definition of a Normal-gamma distribution, which is a product of a conditional Normal distribution, and gamma distribution.

From the posterior definition:

$$\beta, h|y \sim NG(\beta^*, \sum^*, v^*, s^{-2^*})$$

One can easily define the conditional normal distribution of β as:

$$\beta|h, y \sim N(\beta^*, h^{-1}\sum^*) \quad (3.26)$$

Therefore

$$p(\beta|h,y) \propto h^{\frac{k}{2}} exp\left\{\frac{-h}{2}((\beta-\beta^*)^T(\sum^*)^{-1}(\beta-\beta^*))\right\} \quad (3.27)$$

Thus

$$p(h|y) \propto h^{\frac{v*+k}{2}-1}\left\{exp\left[\frac{-h}{2}((\beta-\beta^*)^T(\sum^*)^{-1}(\beta-\beta^*)+v^*s^{2*})\right]\right\} * \left\{exp\frac{-h}{2}((\beta-\beta^*)^T(\sum^*)^{-1}(\beta-\beta^*))\right\}^{-1}$$

$$\propto h^{\frac{k}{2}}\left\{exp\left[\frac{-h}{2}((\beta-\beta^*)^T(\sum^*)^{-1}(\beta-\beta^*))\right]\right\} * h^{\frac{v*}{2}-1}exp\left[\frac{-hv^*}{2s^{-2*}}\right] * \left\{h^{\frac{k}{2}}exp\frac{-h}{2}((\beta-\beta^*)^T(\sum^*)^{-1}(\beta-\beta^*))\right\}^{-1}$$

$$p(h|y) \propto h^{\frac{v*}{2}-1}exp\left[\frac{-hv^*}{2s^{-2*}}\right] \quad (3.28)$$

Hence h follows a Gamma distribution define as $h|y \sim G(s^{-2*}, v^*)$ (3.29)

$$E(h) = s^{-2*} \quad (3.30)$$

$$var(h|y) = \frac{2(s^{-2*})^2}{v^*} \quad (3.31)$$

The above definition follows from gamma distribution with parameters v^* degree of freedom and mean s^{-2*}.

3.1.2.3 Interpretation of the Estimators

β is now the posterior mean for β, which is the Bayesian estimator for the unknown regression coefficient and thus interpreted as the weighted average of the prior mean β_0 and OLS estimator b where the weight reflect the strength of information by prior $(\sum_0)^{-1}$ and data X^TX . The latter of these reflects the confidence in the prior. For instance, if the prior variance selected is high, that implies we are very uncertain about what likely values of β are. As a result, $(\sum_0)^{-1}$ will be small and little weight will be attached to β_0; the best prior guess at what β is. The term X^TX plays a similar role with respect to databased information. Loosely speaking, it reflects the degree of confidence that the data have in its best guess for β; the OLS estimate b. According to frequentist econometrics, we recognize $(X^TX)^{-1}$ as being proportional to the variance of β. Note that, for both prior mean and the OLS estimate, the posterior mean attaches weight proportional to their precisions (i.e. the inverse of their variances). Hence, Bayesian methods combine data and prior information in a sensible way.

In frequentist econometrics, the variance of the OLS estimator for the regression model given in (2.9) is $s^2(X^TX)^{-1}$. The Bayesian analogue is the posterior variance of β given in (3.24), which has a very similar form, but incorporates both prior and data information. For instance, (3.14)

can be informally interpreted as saying "posterior precision is an average of prior precision $(\sum_0)^{-1}$and data precisionX^TX . Similarly, (3.18) has an intuitive interpretation of posterior sum of squared errors (v^*s^{2*}) is the sum of prior sum of squared errors ($v_0{s_0}^2$), OLS sum of squared errors (vs), and a term which measures the conflictbetween prior and data information".

The other equations above also emphasize the intuition that the Bayesian posterior combines data and prior information. Furthermore, the natural conjugate prior implies that the prior can be interpreted as arising from a fictitious dataset (e.g. v and$n$play the same role in (3.16) and (3.18) and, hence,$v$can beinterpreted as a prior sample size).

It is useful to draw out the similarities, differences between what a Bayesian would do, and what a frequentist would do. The latter might calculate b and its variance, $s^2(X^TX)^{-1}$, andestimate σ^2 by s^2. The former might calculate the posterior mean and variance ofβ (i.e. $\beta^*$and $\frac{v^*s^{2*}}{v^*-2}\sum^*$) and estimate $h=\sigma^{-2}$ by its posterior mean,s^{-2*} . These arevery similar strategies, except for two important differences. First, the Bayesian formulae all combine prior and data information. Secondly, the Bayesian interpretsβ as a random variable, whereas the frequentist interprets b as a randomvariable.

The fact that the natural conjugate prior implies prior information enters inthe same manner as data information helps with prior elicitation. For instance,when choosing particular values for$\beta_0, \sum_{\mathbf{0}}, {s_0}^2$ and $v_0$it helps to know that β_0 isequivalent to the OLS estimate from an imaginary data set of v_0 observations withan imaginaryX^TXequal to$(\sum_{\mathbf{0}})^{-1}$ and an imaginary s^2 given by${s_0}^2$.

4.0 Results and Discussion

4.1 Results

The variance inflation factor for all the variables is as follows;

$$Vif(X_1)=5.807767, Vif(X_2)=109.514563, Vif(X_3)=92.23301$$

Table 4.1: Summary of estimate of coefficient and standard error at sample size 20 - 120

Sample size	**BAYES ESTIMATE**			**OLS ESTIMATE**		**RIDGE ESTIMATE**	
		coef	Std deviation	coef	Std Error	coef	Std Error
$n=2020$	$\beta_0=25$	**25**	**7.1**	698.15	20676.02	9502.78	100.34
	$\beta_1=165$	**164.96**	**2.34**	164.91	3.17	164.48	2.82
	$\beta_2=150$	**150.06**	**1.16**	150.47	12.45	155.68	1.36
	$\beta_3=345$	**345.02**	**1.26**	344.72	9.41	340.67	1.51
$n=4040$	$\beta_0=25$	**25**	**10.08**	-264.07	18962.12	6063.21	131.09
	$\beta_1=165$	**165**	**2.36**	165.02	2.91	164.71	2.59
	$\beta_2=150$	**150.04**	**1.15**	149.87	11.42	153.62	1.25

	$\beta_3 = 345$	**345**	**1.27**	345.13	8.63	342.22	1.39
$n =$ 6060	$\beta_0 = 25$	**25**	**8.53**	447.47	12432.24	5426.41	109.99
	$\beta_1 = 165$	**165.08**	**1.63**	165.06	1.9	164.81	1.7
	$\beta_2 = 150$	**149.99**	**0.79**	150.24	7.49	153.19	0.82
	$\beta_3 = 345$	**344.95**	**0.88**	344.76	5.66	342.47	0.91
$n =$ 8080	$\beta_0 = 25$	**25**	**9.56**	-604.39	12234.45	4043.53	116.41
	$\beta_1 = 165$	**165.07**	**1.59**	165.12	1.87	164.89	1.67
	$\beta_2 = 150$	**149.97**	**0.77**	149.59	7.37	152.34	0.81
	$\beta_3 = 345$	**344.96**	**0.85**	345.24	5.57	343.11	0.9
$n =$ 10000	$\beta_0 = 25$	**25**	**8.25**	5.39	9349.61	3823.23	108.36
	$\beta_1 = 165$	**164.95**	**1.22**	164.95	1.43	164.76	1.28
	$\beta_2 = 150$	**149.99**	**0.59**	149.97	5.63	152.23	0.62
	$\beta_3 = 345$	**345.03**	**0.66**	345.04	4.25	343.28	0.69
$n =$ 12020	$\beta_0 = 25$	**25**	**7.75**	-76.8	7910.44	3645.4	94.15
	$\beta_1 = 165$	**165.04**	**1.05**	165.05	1.21	164.87	1.08
	$\beta_2 = 150$	**149.99**	**0.51**	149.93	4.76	152.13	0.52
	$\beta_3 = 345$	**344.98**	**0.56**	345.02	3.6	343.31	0.58

Table 4.2: Summary of estimate of coefficient and standard error at sample size 140 –200 500, 1000.

Sample size	**BAYES ESTIMATE**			**OLS ESTIMATE**		**RIDGE ESTIMATE**	
		Coef	Std deviation	Coef	Std error	Coef	Std error
$n =$ 14040	$\beta_0 = 25$	**25**	**9.71**	-113.28	9193.91	3325.15	118.49
	$\beta_1 = 165$	**164.99**	**1.22**	165	1.41	164.83	1.25
	$\beta_2 = 150$	**150.01**	**0.59**	149.92	5.54	151.96	0.61
	$\beta_3 = 345$	**345**	**0.65**	345.07	4.18	343.48	0.68
$n =$ 16060	$\beta_0 = 25$	**25**	**9.03**	166.19	7929.53	3244.19	114.06
	$\beta_1 = 165$	**165.02**	**1.06**	165.01	1.21	164.86	1.08
	$\beta_2 = 150$	**150.02**	**0.51**	150.11	4.78	151.93	0.53
	$\beta_3 = 345$	**344.99**	**0.57**	344.92	3.61	343.51	0.58
$n =$ 18080	$\beta_0 = 25$	**25**	**9.18**	-206.9	7612.49	2768.65	113.31
	$\beta_1 = 165$	**164.99**	**1.02**	165.01	1.17	164.86	1.04
	$\beta_2 = 150$	**150**	**0.49**	149.86	4.59	151.63	0.51
	$\beta_3 = 345$	**345**	**0.55**	345.11	3.46	343.74	0.56

$n =$ 20000	$\beta_0 = 25$	**25**	**8.82**	-227.56	6931.2	2727.32	103.92
	$\beta_1 = 165$	**165.02**	**0.93**	165.04	1.06	164.89	0.95
	$\beta_2 = 150$	**150.02**	**0.45**	149.86	4.17	151.61	0.46
	$\beta_3 = 345$	**344.99**	**0.5**	345.1	3.15	343.74	0.51
$n =$ 50000	$\beta_0 = 25$	**25**	**8.96**	25	4393.89	1893.03	110.45
	$\beta_1 = 165$	**165.02**	**0.6**	165.02	0.67	164.92	0.6
	$\beta_2 = 150$	**149.99**	**0.29**	149.99	2.65	151.1	0.3
	$\beta_3 = 345$	**344.99**	**0.32**	344.99	2	344.13	0.33
$n =$ 100000	$\beta_0 = 25$	**25**	**9.06**	25	3134.79	1311.35	103.97
	$\beta_1 = 165$	**164.98**	**0.43**	164.98	0.48	164.91	0.43
	$\beta_2 = 150$	**150**	**0.21**	150	1.89	150.76	0.22
	$\beta_3 = 345$	**345.01**	**0.23**	345.01	1.43	344.42	0.23

Table 4.3: Summary of 95% confidence and credible interval for the estimators

Sample size	**BAYES 95% CREDIBLE INTERVAL**			**OLS 95% CONFIDENCE INTERVAL**		**RIDGE 95% CONFIDENCE INTERVAL**	
		Lower	Upper	Lower	Upper	Lower	Upper
$n =$ 2020	$\beta_0 = 25$	**14.16597**	**36.17456**	-63301.8	24360.65	9290.082	9715.488
	$\beta_1 = 165$	**159.1038**	**168.078**	157.4136	170.8445	158.5006	170.4588
	$\beta_2 = 150$	**145.903**	**150.4816**	110.5315	163.3343	152.7959	158.5682
	$\beta_3 = 345$	**343.4001**	**348.1928**	335.1006	374.9946	337.4551	343.8773
$n =$ 6060	$\beta_0 = 25$	**10.71535**	**42.39604**	-48860.8	948.7104	5206.075	5646.74
	$\beta_1 = 165$	**163.8568**	**169.2768**	165.0413	172.6727	161.4142	168.2089
	$\beta_2 = 150$	**149.3212**	**151.8048**	121.1434	151.1458	151.5473	154.8344
	$\beta_3 = 345$	**342.6347**	**345.5541**	343.2246	365.8922	340.6474	344.3002
$n =$ 10000	$\beta_0 = 25$	**10.76345**	**40.3885**	-14766.5	22351.19	3608.133	4038.33
	$\beta_1 = 165$	**162.28**	**167.3016**	161.7387	167.4256	162.2273	167.2907
	$\beta_2 = 150$	**149.4397**	**151.6747**	141.4799	163.8375	151.0066	153.4618
	$\beta_3 = 345$	**343.762**	**346.4126**	334.9215	351.8133	341.9207	344.6457
$n =$ 20000	$\beta_0 = 25$	**7.558032**	**43.71414**	-16502.7	10835.94	2522.372	2932.274
	$\beta_1 = 165$	**163.5901**	**167.1153**	163.4495	167.6381	163.0267	166.7561
	$\beta_2 = 150$	**149.1336**	**150.986**	140.0205	156.4877	150.7072	152.5223
	$\beta_3 = 345$	**343.8519**	**345.7796**	339.8815	352.323	342.7376	344.748

Table 4.4: Mean square error of the estimators at various sample sizes

Sample Sizes	BAYES MSE	OLS MSE	RIDGE MSE
20	**3.33**	1.32×10^{8}	67235991
40	**1.62**	60055660	30204822
60	**1.11**	40863372	22146422
80	**0.86**	29947007	13710192
100	**0.71**	22829325	11558297
120	**0.55**	21264974	9871460
140	**0.48**	17336409	8464848
160	**0.42**	15317101	7756157
180	**0.38**	12908635	6250229
200	**0.34**	12767788	6159415
500	**0.14**	5109822	2668017
1000	**0.07**	2603927	1290459

Table 4.5: Mean square error of prediction for the estimators at various sample sizes

Sample Sizes	BAYES MSEP	OLS MSEP	RIDGE MSEP
20	**7213.086**	7303.68	8029.317
40	**4635.332**	4650.295	4522.359
60	**8911.436**	8919.751	9087.504
80	**12461.58**	12468.18	12288.64
100	**11809.74**	11802.38	12185.98
120	**11360.94**	11369.32	11229.35
140	**11345.92**	11338.87	11496.92
160	**10553.39**	10551.4	10511.69
180	**13365.38**	13364.45	13314.26
200	**11596.22**	11586.63	11464.2

4.2 Discussion of Results

In this research, two frequentist methods (Ordinary Least Square (OLS) & Ridge Regression (RR)) and Bayesian Regression were used to fit a set of collinear data. OLS performed as expected due to the effect of collinearity existing between the predictors, RR also produces a fairly precise estimate but the estimates were totally different from the true value (Bias). The Bayesian regression which uses the prior as an ingredient to solve the problem of collinearity produces the closest estimates to the true value and also precise. The above results can be found in table 4.1. Increasing the sample size in table 4.2 improves the estimate of the three estimators, OLS estimates and Bayesian estimates now converges to be the same, RR and Bayesian precision also converges to be the same. Table 4.3 presents the 95% confidence and credible interval as in the case of frequentist and Bayesian respectively. Due to the imprecise standard errors of the frequentists estimate, the OLS and RR produces wide confidence intervals i.e if we are to test the hypothesis of significance of β_o, the hypothesis of non-

significance would not be rejected. The Bayesian credible intervals interpreted as the probability that the unknown parameter will fall in the interval were narrower compared to the frequentist counterparts.

Furthermore, Table 4.4 presents the Mean Square Error MSE which is used to measure the average closeness of the estimators to the true values, it was observed that the Bayesian estimator produce << lower MSE compared to OLS and RR. Although RR MSE is approximately half of the OLS MSE, but its better is not the best.

To access the predictive ability of the estimators, Mean Square Error of Prediction was used, the results were presented in table 4.5. It was observed that the three estimators performed extremely the same, with slight better performance from the Bayesian estimators in some cases. All the results discussed above were also confirmed using box & whisker plots and line graphs.

5.2 Conclusion

In this study a simple way of modelling collinear data under simulation approach was presented. It was observed that modelling collinearity in a full Bayesian using a Normal-Gamma conjugate prior have improved the precision of the estimates and the efficiency of the inferences about the parameters.

References

Anscombe F, Aumann R (1963). "A Definition of Subjective Probability", The Annals of Mathematical Statistics, 34(1), 199{205.

Bayes T, Price R (1763). "An Essay Towards Solving a Problem in the Doctrine of Chances" Berger, J., (2006). "The Case for Objective Bayesian Analysis", Bayesian Analysis, 1(3).

Bernardo, J, Smith A (2000); "Bayesian Theory". John Wiley & Sons, West Sussex, England.

Bernardo, J., (2008). "Comment on Article by Gelman", Bayesian Analysis, 3(3), 451{454.

Gelman A (2006), "Prior Distributions for Variance Parameters in Hierarchical Models" Bayesian Analysis, 1(3), 515{533}.

Greene, W. (2000); "Econometric Analysis, fourth edition". New Jersey: Prentice-Hall, p. 7)

Gujarati (2004), "Basic Econometrics", McGraw-hill Companies

Koop, .G., (2003), "Bayesian Econometrics", John Wiley & Sons Ltd

Lindley, D. V., and Smith, A. F. M., (1972); "Bayes Estimates for the Linear Model". In: Journal of the Royal Statistical Society. Series B (Methodological)34 (1972), Nr. 1, S. 1–41. – ISSN 00359246

Raifa, H, Schlaifer, R., (1961). Applied Statistical Decision Theory. Division of Research, Graduate School of Business Administration, Harvard University.

Ramsey., F (1926). Truth and Probability." In R~Braithwaite (ed.), The Foundations of Mathematics and other Logical Essays, pp. 156{198.Harcourt, Brace and Company, NewYork.

Roberts., C., (2007). The Bayesian Choice. 2nd edition. Springer, Paris, France

Simon, K., (2009), The Bayesian Linear model with unknown variance, Seminar for Statistics ETH Zurich p. 9-12