Journal of Science Research (2014) Vol. 13: 73-83

Bayesian and Classical Estimation Methods in Linear Regression Model: A Monte Carlo Simulation Analysis

***Olubusoye, O.E. and Akanbi, O.B.**

Abstract
In the recent times the Bayesian approach has emerged as a strong competitor to the traditional Classical approach to econometric analysis. The main reason for the vicious controversy between the promoters of the two approaches is the notions of probability they employ. The framework of Bayesian approach allows for subjectivity in the choice of prior distribution. The main concern has been the relative performance of this approach especially in the areas of applications. This study focuses on the general linear regression model and investigates the differences between the Classical and the Bayesian approach in terms of model formulation and estimation procedures. The paper describes the procedure for executing Bayesian regression and the choice of the prior distribution to be employed. The *Prior* distribution is the degree of belief that the researcher has about the distribution of the parameter that he is trying to estimate. Non-Bayesians usually employ this information to lead them to add, drop, or modify variables in an ad hoc search for a "better" result. The Bayesians employ it *ex ante* in an explicit, upfront fashion. The performance of the Bayesian approach and the Classical was assessed through a series of Monte Carlo experiments. A large number of samples are drawn and the Classical and Bayesian estimators were computed for each sample. The approximate sampling distribution of the statistics were then determined. Using bias and mean square error criteria, the results showed that there was nothing to choose from when prior density is non-informative.

Keywords: Prior, Posterior, Likelihood, Mean Squared Error, Credible Interval.

Introduction

There are now two very different approaches to modern econometrics, the "Classical" approach and the "Bayesian" approach. The main difference centered on the notion of probability they employed. The objectivity of Classical approach is obtained by disregarding any prior knowledge about the process being measured. Bayesian econometricians use both sources of information; the prior information about the process and the information about the process contained in the data. The adoption of Bayesian approach in econometrics involves several practical difficulties but in recent years, development of more powerful computers, new software, and computational innovations, these practical difficulties have for the most part been overcome.

***Olubusoye, O.E. and Akanbi, O.B.**
**Department of Statistics, University of Ibadan, Ibadan, Oyo State, Nigeria.*
Corresponding Author
E-mail: oe.olubusoye@ ui.edu.ng, busoye2001@yahoo.com

Bayesian approach is a very vital aspect of statistics to both the econometricians and non econometricians; is discussed comprehensively in literature starting from the elementary level, via the intermediate level, to the excellent advanced stage, where the Bayesian view of econometric problems was covered, [1]. Its application was also widely discussed [2-4]. Koop [5] and Albert [6] worked on computer programming for Bayesian econometrics. Weber [7] examines the history of the Bayesian controversy. Bayesian econometrics is based on a few simple rules of probability. This is one of the basic advantages of the Bayesian approach and for which many are finding it increasingly more attractive. A complete and more persuasive listing of the advantages of the Bayesian approach can be found in Zellner [8] and Poirier [9] which focuses on the methodology and statistical theory underlying Bayesian and the Classical methods. Other important Bayesian books, such as those of Bauwens et al. [10], deal

ISSN 11179333

only with particular areas of econometrics (e.g. time series models). More recently, Lancaster [11] and Koop [12] introduced Bayesian approach to econometrics and cover a wide range of models that are popular in non-Bayesian econometric textbooks. But Koop's [12] submission is an excellent exposition of progress in applied Bayesian econometrics, with particular emphasis on computational considerations.

The centerpiece of the Bayesian methodology is the Bayes theorem. This theorem states that for events A and B, the conditional probability of event A given that B has occurred is

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)} \quad (1)$$

This equation is paraphrased for econometric applications as follows:

$$P(\text{parameters}|\textit{data}) = \frac{P(\text{data}|\text{parameters})P(\text{parameters})}{P(\text{data})} \quad (2)$$

With this arrangement, data are viewed as constants whose distributions do not involve the parameters of interest. Data are treated as a fixed set of information to be used in updating ones beliefs about the parameters. In equation (2) above, if we drop the marginal density of the data, we have

$$P(\text{parameters}|\textit{data}) \propto P(\text{data}|\text{parameters})P(\text{parameters}) \quad (3)$$

The left-hand side is the posterior density of the parameters, given the current body of data. The first term on the right is the joint distribution of the data, given the parameters. It is called the likelihood function. The second term is the prior density which represents the prior beliefs of the analyst about the parameters. Thus, we can re-write the above equation as

$$\text{Posterior } \textit{density} \propto \text{Likelihood function} \times \textit{Prior density} \quad (4)$$

The posterior is a mixture of the prior information and the "current information," that is, the data. Once obtained, this posterior density is available to be the prior density function when the next body of data or other usable information becomes available. The principle involved, which appears nowhere in the Classical analysis, is one of continual accumulation of knowledge about the parameters. The prior density and the likelihood function are crucial elements of the analysis, and both must be fully specified for estimation to proceed. The Bayesian "estimator" is the mean of the posterior density of the parameters, a quantity which is usually obtained either by integration (when closed form exists), approximation of integrals by numerical techniques, or by Monte Carlo methods [13-16].

Studying both Bayesian and Classical methods provide a much better understanding of statistics than that provided by studying only one approach. In recent times, there has been various controversy among the proponents of these two approaches. Efron [17] is suitable for exploring the Classical/Bayesian controversy while Poirier [18] is suitable for associated commentary.

Olubusoye and Osowole [19] compared the Bayesian and the Classical estimators of a binomial proportion using a Monte Carlo experiment. The results showed that the Bayesian estimator outperformed the classical even when judged by the Classical criterion.

The arching question is: what is the gain of Bayesian over the Classical approach? In order to answer in the context of an econometric model, this study focuses on the general linear regression model and investigates the differences between the Classical and the Bayesian estimator of the model parameters. The Bayesian estimators are derived in this paper. The performance of the estimators from the two approaches is judged from a series of simulation experiments with varying sample sizes.

The rest of the paper is divided into four sections. In section II, we discussed the form of the likelihood function, the prior density and the posterior density of a simple linear regression model. Also, the Bayesian credible interval and the classical confidence interval are discussed in this section. In section III, we describe the design of the Monte Carlo experiment. The results of the experiment are discussed in section IV and finally the concluding remark is given in section V.

Bayesian Approach to Linear Regression

The Bayesian approach to linear model specification is not as such different from that of classical except for the elicitation of the prior. For simplicity sake, consider a simple two-variable linear model with dependent variable y_i expressed as a linear combination of one regressor or explanatory variable plus an error term.

Specifying the model explicitly as:

$$y_i = \ _1 + \ _2 x_i + £_i \qquad (5)$$

where, y_i and x_i denote the observed data on the dependent and explanatory variables respectively, for i = 1, 2, ……, N; ε_i is an error term which is independently and identically distributed as N(0, $\pm_\varepsilon^2$). The x_i is fixed (i.e. not a random variable) or, if random, is independent of ε_i with a probability density function, P(x/λ), where λ is a vector parameter space that does not include β_2 and $\pm_\varepsilon^2$. β_1 is the mean value for y given $x = \bar{x}$. β_2 is the slope (regression coefficient) and y_i is normally distributed with mean, $_1 + \ _2\bar{x}$ and variance, $\pm_\varepsilon^2$.

Bayes Theorem for the Regression Model

Bayes' theorem is always summarized by **posterior ∝ likelihood × prior**.

Prior distribution (P($_1$), P($_2$)): is a probability statement about parameters which is expressed as degree of one's belief about the parameter before observing the data. The choice of prior cannot be ignored in Bayesian analysis because of its sensitivity; Olubusoye et al. [20] worked on the prior sensitivity in Bayesian linear regression model. Prior distributions can be informative, non-informative or Hierarchical. The informative priors (also known as conjugate priors) convey information concerning the prior preference for certain values of the parameters. It comes from either the "expert" opinion or the previous experiments of a similar nature. This can often be done formally using meta-analysis or hierarchical Bayesian modelling of the existing data. It is often easier to express probability bounds, from which the parameters can then be obtained rather than trying to directly specify values for the parameters of a prior density. Non informative priors do not convey information concerning prior preference for certain values of the parameters.

It can also be categorized as reference priors and vague priors. Reference prior is considered as default prior for the particular model in question. Reference priors are frequently improper and are being guided by Jeffreys's rule. Priors such as P(μ) = 1, P(σ) = 1/σ are improper because they do not

integrate to 1. That is, the area under the prior density is not unity (and, in fact, is infinity). Hierarchical priors are two stage priors, in the sense that a prior is placed on another prior. They are more flexible than non-hierarchical priors, make the posterior less sensitive to the prior and being used in the Hierarchical models to perform meta-analyses, whereby a number of related experiments are performed and it is desired to combine information. Multiplying the joint likelihood by a joint prior, it is proportional to the joint posterior. We used independent priors for each parameter. The joint prior of the two parameters is the product of the two individual priors;

$$P(\beta_1, \beta_2) \propto P(\beta_1) \times P(\beta_2) \quad (6)$$

Likelihood function/distribution: L($_1$:y), L($_2$:y): The expression for the distribution of the data to be observed given the parameter *P(y/ $_2$)*, has two names. When thought of as the probability mass function of y calculated at the point y, conditional on the parameter taking the value $_2$, it is called, the pdf of y given $_2$. But when y is thought of as the actual data it is often denoted by the symbol y^{obs}, it is called the likelihood function (of $_2$) i.e.,

$L(\beta_2{:}y)$: The likelihood function is not, in general, a probability distribution for $_2$ given data *y*, but it is a function of $_2$ with the data values serving as parameters of that function.

In classical approach, inferences are based on likelihood functions following the work of the R.A. Fisher in 1925 and after. Those who follow this approach will typically choose as their estimate of $_2$ the value that provides the maximum (strictly the supremum) of the likelihood over the parameter space, Θ. This is called the maximum likelihood estimator.

In Bayesian paradigm, $L(\beta_2;y)$ is rather called a function of $_2$ when the observed values of the variable(s) of interest are considered. The choice of a likelihood function amounts to choice of a family of probability distributions, one for each $_2$ in the parameter space. The choice must express the economic model that lies at the center of an econometric investigation.

The joint likelihood for $_1$ and $_2$

Using equation (5)

$$\text{Likelihood}_i(\beta_1, \beta_2) \propto exp\left\{-\frac{1}{2\sigma^2}[y_i - \beta_1 - \beta_2 x_i]^2\right\} \quad (7)$$

excluding the constant term, $\left\{(2\pi\sigma^2)^{-1/2}\right\}$. Hence, the expression left is called the Kernel distribution. Then, considering the entire sample n, we have,

$$Likelihood_n(\beta_1, \beta_2) \propto \prod exp\left\{-\frac{1}{2\sigma^2}[y_i - \beta_1 - \beta_2 x_i]^2\right\}$$

$$\propto \exp\left\{-\frac{1}{2\sigma^2}\sum_i^n [y_i - \beta_1 - \beta_2 x_i]^2\right\} \quad (8)$$

The terms in the brackets in the exponent can also be written as:

$$\sum_{i}^{n}\{y_i-\bar{y}+\bar{y}-\ {}_1-\ {}_2x_i\}$$

Breaking the last expression into two sums and multiplying them out gives us;

$$\sum_{i=1}^{n}\left(y_i-\bar{y}\right)^2+2\sum_{i=1}^{n}\left(y_i-\bar{y}\right)\{\bar{y}-\ {}_1-\ {}_2x_i\}+\sum_{i}^{n}\{\bar{y}-\ {}_1-\ {}_2x_i\}^2$$

i.e. $ss_y-2\ {}_2ss_{xy}+\ {}_2{}^2ss_x+\mathrm{n}\left(\ {}_1\text{-}\bar{\mathrm{y}}\right)^2$

where, $ss_y=\sum_{i=1}^{n}\left(y_i-\bar{y}\right)^2$, and $\mathrm{ss_{xy}}=\sum_{\mathrm{i}=1}^{\mathrm{n}}\left(\mathrm{y_i}-\bar{y}\right)\left(x_i-\bar{x}\right)$, and $\mathrm{ss_x}=\sum_{i=1}^{n}\left(\mathrm{x_i}-\bar{x}\right)^2$

Thus,

$$\mathrm{Likelihood_n}\ (\beta_1,\beta_2)\propto exp\left\{-\frac{1}{2\sigma^2}\left[\mathrm{ss_y}-2\beta_2 ss_{xy}+\beta_2^2\mathrm{ss_x}+n(\beta_1-\bar{\mathrm{y}})^2\right]\right\}$$

Breaking down into products of exponent we have;

$$\mathrm{Likelihood_n}\ (\beta_1,\beta_2)\propto exp\left\{-\frac{1}{2\sigma^2}\left[\mathrm{ss_y}-2\beta_2 ss_{xy}+\beta_2^2\mathrm{ss_x}\right]\right\}\times exp\left\{-\frac{1}{2\sigma^2}\left[\mathrm{n}(\beta_1-\bar{\mathrm{y}})^2\right]\right\}$$

Then:

$$\mathrm{Likelihood_n}\ (\beta_1,\beta_2)\propto exp\left\{-\frac{1}{2\sigma^2/_{\mathrm{ss_x}}}\left[\beta_2-\frac{\mathrm{ss_{xy}}}{\mathrm{ss_x}}\right]^2\right\}\times exp\left\{-\frac{1}{2\sigma^2/_{\mathrm{n}}}\left[(\beta_1-\bar{\mathrm{y}})^2\right]\right\} \quad (9)$$

If B= $\frac{\mathrm{ss_{xy}}}{\mathrm{ss_x}}$, the least square slope; $\bar{y}=A$, the least squares estimate of the intercept of the vertical line $x=\bar{x}$

Thus;

$$\mathrm{Likelihood_n}(\beta_1,\beta_2)\propto \mathrm{Likelihood_n}(\beta_1)\times \mathrm{Likelihood_n}(\beta_2)$$

But;

$$\mathrm{Likelihood_n}\ (\beta_2)\propto exp\left\{-\frac{1}{2\sigma^2/_{\mathrm{ss_x}}}[\beta_2-\mathrm{B}]^2\right\};\ \beta_2 \text{ is normally distributed as } (\mathrm{B},\ \frac{\sigma^2}{\mathrm{ss_x}}) \quad (10)$$

and

$$\mathrm{Likelihood_n}\ (\beta_1)\propto exp\left\{-\frac{1}{2\sigma^2/_{\mathrm{n}}}[(\beta_1-\mathrm{A})^2]\right\};\ \beta_1 \text{ is normally distributed as } (\mathrm{A},\ \frac{\sigma^2}{\mathrm{n}}) \quad (11)$$

Posterior distribution P($_2$/y); P($_1$/y)

The posterior density represents the belief about $_2$ given prior belief about $_2$ and the belief embodied in its likelihood from the data. In many applications the posterior is the culmination of an empirical analysis. The best way to report the posterior distribution is by its plot after stating the model (if possible) mathematically.

The Joint Posterior for $_1$ and $_2$

The joint posterior is proportional to the joint prior times the joint likelihood.

$$P\left(\beta_1,\beta_2/_{data}\right) \propto P(\beta_1,\beta_2) \times Likelihood_n(\beta_1,\beta_2) \tag{12}$$

Where the data is the set of ordered pair $(x_1,y_1)\ldots.,(x_n, y_n)$. Also, the joint posterior above can be re-arranged as;

$$P\left(\beta_1,\beta_2/_{data}\right) \propto P\left(\beta_1/_{data}\right) \times \left(\beta_2/_{data}\right) \tag{13}$$

Joint posterior gives the product of the marginal posteriors, hence they are independent. For instance, if we use a normal $(m_{2,}\ s^2_2)$ prior for $_2$, we have a normal posterior $\left[m^1_2\ \left(s^1_2\right)^2\right]$, where

$$\frac{1}{\left(s^1_2\right)^2} = \frac{1}{s^2_2} + \frac{ss_x}{\pm^2} \tag{14}$$

and

$$m^1_2 = \frac{\frac{1}{s^2_2}}{\frac{1}{\left(s^1_2\right)^2}}(m_2) + \frac{\frac{ss_x}{\pm^2}}{\frac{1}{\left(s^1_2\right)^2}}(B) \tag{15}$$

$$where\ B = \frac{ss_{xy}}{ss_x}$$

Equation (15) implies that the posterior mean equals the weighted average of the prior mean and the likelihood mean where the weights are the proportions of the precisions to the posterior precision. And of course the posterior distribution is normal.

Similarly if, we use a normal $\left(m_{1,}\ s^2_1\right)$ prior for $_1$, we have the posterior distribution which is also normal with $\left[m^1_1,\left(s^1_1\right)^2\right]$ i.e.

$\frac{1}{\left(s^1_1\right)^2} = \frac{1}{s^2_1} + \frac{n}{\pm^2}$ and

$$m^1_1 = \frac{\frac{1}{s^2_1}}{\frac{1}{\left(s^1_1\right)^2}}(m_1) + \frac{\frac{n}{\pm^2}}{\frac{1}{\left(s^1_1\right)^2}}(A), \quad \text{where,}$$

$$A = \bar{y} \tag{16}$$

We may want to summarize the entire belief about the slope (β_2) after examining the data by a (1-α) 100% Bayesian credible interval for slope β_2;

$$m^1_2 \pm Z_{\frac{\sim}{2}} \sqrt{\left(s^1_2\right)^2},\ where\ \left(s^1_2\right)^2 = \frac{1}{s^2_2} + \frac{ss_x}{\pm^2} \tag{17}$$

But, if $\pm^2$ is unknown, then we estimate it by:

$$\pm^2 = \frac{\sum\left[y_i - \left(A + Bx_i\right)\right]^2}{n-2}$$

Then, the credible interval becomes;

$$m^1_2 \pm t_{\sim/2} \sqrt{\left(s^1_2\right)^2} \tag{18}$$

When the variance $\pm^2$ also unknown, the (1-α)100% Classical confidence interval for the slope (β_2) is;

$$B \pm t_{\sim/2} \frac{\pm^2}{\sqrt{ss_x}} \tag{19}$$

where,

$\pm^2$ is the estimate of the variance calculated from the residuals in the least squares regression line.

Design of the Monte Carlo Experiment

A Monte Carlo study approximating the sampling distribution of two estimators of β_2, was performed. The classical estimator for β_2 is the sample regression coefficient

$$_2 = \sum_i^n x_i y_i \Big/ \sum_i^n x_i^2 \text{ , where } \sum_{i=1}^n x_i y_i \text{ ,}$$ is the product of the deviation sum of squares of X and Y. $\sum_{i=1}^n x_i^2$ is the deviation sum of squares of X. The Bayesian estimator used is $m^1_{\beta 2}$ which equals the posterior mean when a normal prior for β_2 was used. The sampling distributions (in terms of bias, variance and mean square error) of the two estimators in a Monte Carlo replicated 10,000 for various sample sizes were compared.

The experiment is performed using the following steps:

Step 1. The prior distribution for the model parameters were specified as follows:
$_1 \sim N(1,100000)$ and $_2 \sim N(1,100000)$.
These are vague priors since they allow for normal densities with extremely large variance.

Step 2. The true values of the parameters were assumed as follows:
$_1 = 4.479383$ and $_2 = 1.299635$

Step 3. The sample size n = 10, 20, 30, 50, 100, 150, 200, 300, 500, 1000, were considered.

Step 4. The values of x for each observation were fixed in all replications. The x values were drawn from a uniform distribution, that is U[0,1].

Step 5. The error term ε_i was drawn from normal probability distribution with mean 0, and variance, $\pm_£^2$

Step 6. Given the β_1, β_2, X_i, and ε_i, equation (5) was used to obtain the implied values of Y_i, for i=1, ……,n.

Step 7. Now, using the y_i values thus generated, they were regressed on the x_i values drawn in step 3 to obtain the Classical and Bayesian estimates of $\beta_1$and β_2.

Step 8. This procedure was repeated 9,999 times, each time using the same true values of the parameters β_1, β_2 and x values. The ε_i values varied from sample to sample and therefore, in all the 10,000 samples, thus producing 10,000 values each of β_1 and β_2.

Step 9. The averages of these 10,000 estimates were computed

$$_{i,classical} = \frac{\sum_i^{10000} \hat{}_{i,classical}}{10000},$$

$$_{i,bayesian} = \frac{\sum_i^{10000} \hat{}_{i,bayesian}}{10000}, \quad i = 1,2$$

Step10. For each sample size, n, the bias of the two estimators were calculated;

$$bias\left(\hat{}_i\right) = E\left(\hat{}_i\right) - \quad , \quad i = 1,2$$

Step 11. The mean squares error (MSE) of the estimate was computed using,

$$MSE(\hat{}_i) = [bias(\hat{}_i)]^2 + Var(\hat{}_i), \quad i=1,2$$

The Eviews (7) programming code was written to run the simulations.

Result and Discussion

The posterior density (*Kernel distribution)* which is the density left when the multiplicative constant in the formal distribution is ignored for the estimates of regression slope from the 10,000 replications is plotted for the two methods. The plots are

shown in Figures 1 and 2 below. The average (posterior mean) is approximately 1.3 in both methods. This value is very close to the true value of 1.299635.

The criteria used for assessing the estimators of both the Classical and the Bayesian approaches are bias and the mean squared errors (MSEs). The results on the simulation experiment based on these criteria are presented in Table 1 and the Figures 1-4 below. Table 1 shows the bias and the mean squared errors of the two estimators respectively for both the Classical and the Bayesian over 10 different sample sizes. From Table 1, it is observed that the bias and MSE for the estimators in both the Classical and the Bayesian are the same for both the small and the large samples. This phenomenon is captured vividly in Figures 3 and 4 below. When the sample size is small ($n < 20$), the slope's bias for the two estimators reduced drastically, while the sample size between 20 and 30, the bias increased slowly. Also, at sample size $n = 50$ the slope's bias for the two estimators is zero. Furthermore, it is confirmed that the larger the sample size, the smaller the biases and MSEs for the two estimators in both cases (Bayesian and Classical). Similarly, for the two estimators, the bias is very close to zero over the entire sample sizes. Considering, the Figures 3 and 4, bias of the slope within the sample size 300 and 500 shows high fluctuation. The plot becomes more stable as the sample size increases from 500 upward for both the Classical and Bayesian estimates. Figure 4 shows the mean square errors of the estimators in both methods approaching zero when sample size becomes very large. Mean squared errors (MSE) of the slopes (estimators) reduced sharply for sample sizes $n < 150$ but at n greater or equal to 150, their MSEs gradually reduce to zero.

Table 1: The Bias and MSE values of the Classical and Bayesian Estimators

	BIAS				MSE			
Sample size(n)	**Bayesian B1**	**Classical B1**	**Bayesian B2**	**Classical B2**	**Bayesian B1**	**Classical B1**	**Bayesian B1**	**Classical B2**
10	0.0036	0.0036	0.0039	0.0039	0.5974	0.5974	1.5776	1.5776
20	0.0050	0.0050	-0.0043	-0.0043	0.2134	0.2134	0.6579	0.6579
30	0.0083	0.0083	-0.0011	-0.0011	0.0927	0.0927	0.3641	0.3641
50	0.0011	0.0011	0.0000	0.0000	0.0954	0.0954	0.2763	0.2763
100	0.0002	0.0002	-0.0003	-0.0003	0.0369	0.0369	0.1094	0.1094
150	-0.0002	-0.0002	0.0002	0.0002	0.0266	0.0266	0.0773	0.0773
200	0.0001	0.0001	0.0003	0.0003	0.0188	0.0188	0.0590	0.0590
300	-0.0010	-0.0010	0.0023	0.0023	0.0133	0.0133	0.0383	0.0383
500	0.0011	0.0011	-0.0018	-0.0018	0.0082	0.0082	0.0253	0.0253
1000	0.0006	0.0006	-0.0016	-0.0016	0.0041	0.0041	0.0123	0.0123

Conclusion

In this study, a Monte Carlo experiment and a general linear regression model were used to investigate the differences between the Classical approach and the Bayesian approach in terms of model formulation and estimation. Using bias and mean square errors criteria, the two exhibited similar characteristics and therefore gives no room for a choice. This result may be due to the use of vague priors which reflect the ignorance of the researcher about the parameter. In this case, the two methods will yield similar results. As more knowledge about the parameters is utilized in specifying the prior (informative prior), it is expected that the Bayesian will tend to outperform the Classical.

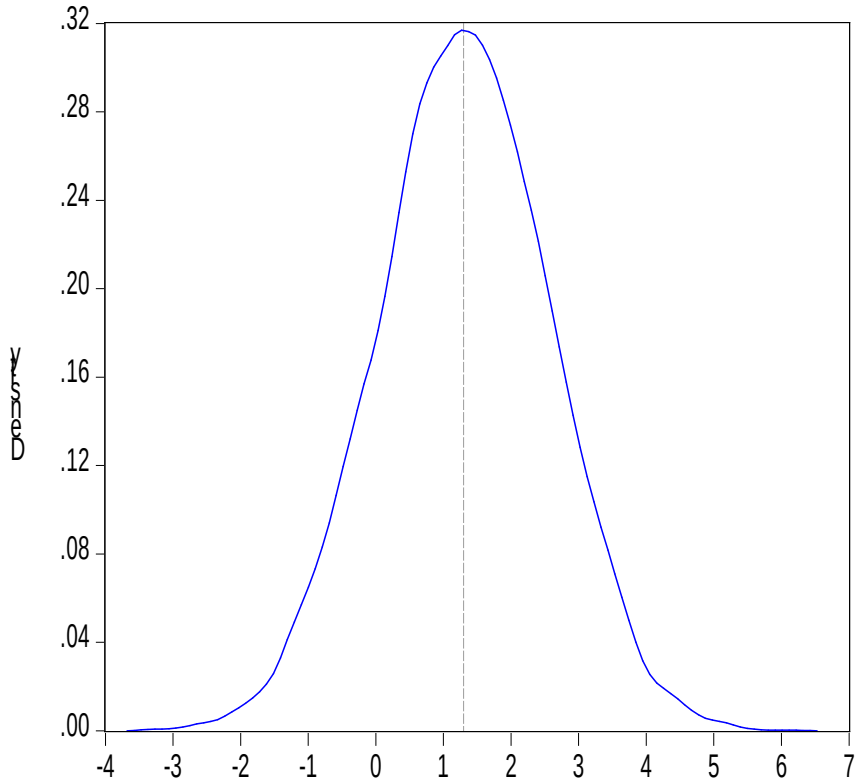

Fig. 1: Density of Estimates of Regression Slope Coefficient (Bayesian)

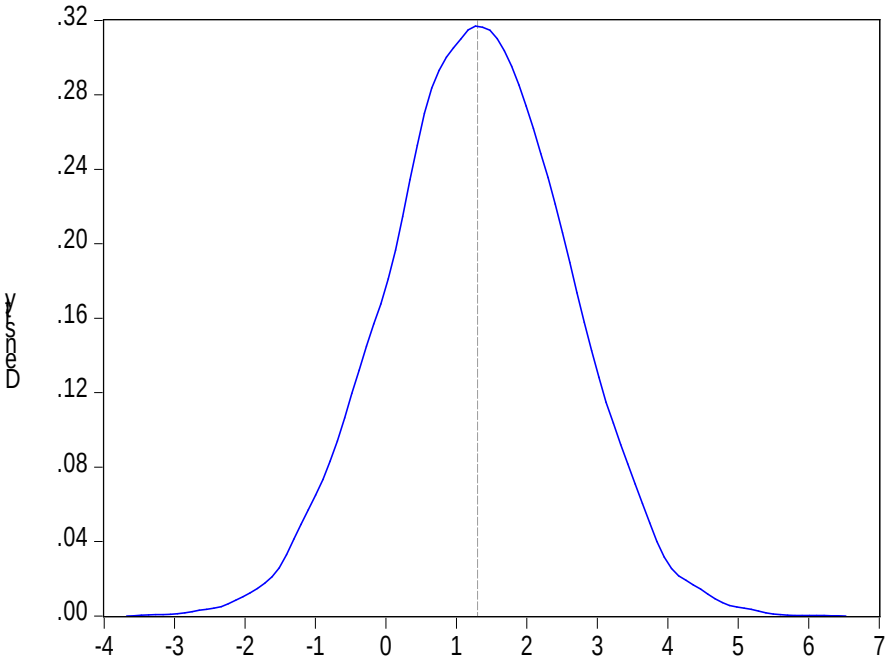

Fig. 2: Density of Estimates of Regression Slope Coefficient (Classical)

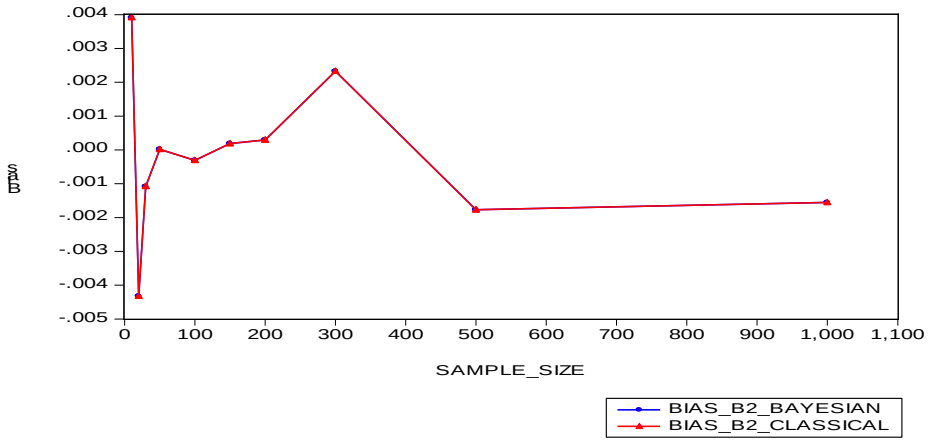

Fig. 3: Bias of Classical and Bayesian Estimates for Regression Slope Coefficient

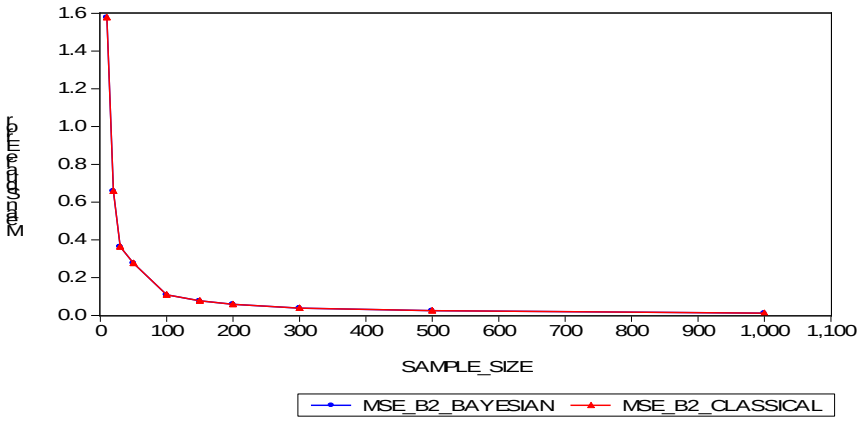

Fig. 4: Mean Square Error of Classical and Bayesian Estimates of Regression Slope Coefficient

References

[1] Zellner, A. 1971. An Introduction to Bayesian Analysis Inference in Econometrics, New York: John Willey & Sons.

[2] Michael, J.R and Sedransk, J. 2010. Bayesian and frequentist methods for provider profiling using Risk–Adjusted Assessments of medical outcomes. *Journal of the American Statistical Association* 105(489).

[3] Andrzej, .K. 2010. A prior for Impulse Responses in Bayesian Structural VAR models. *Journal of Business and Economics Statistics* 28(1).

[4] Berger, J.O. 2000. A look at Today and thought of Tomorrow. *Journal of the American Statistical Association* 95(452): 1269- 1276.

[5] Koop, G. 1999. Bayesian analysis, Computation and Communication Software. *Journal of Applied Econometrics* 14: 677-689.

[6] Albert, J. 2007. Bayesian Computation with R. Springer. ISBN 9780-387-71384-7.

[7] Weber, J.D. 1973. Historical Aspects of the Bayesian Controversy. Tucson: Division of Economic and Business Research, University of Arizona.

[8] Zellner, A. 1974. The Bayesian Approach and Alternatives in Econometrics. In S.E. Poirier, D. 1995. Intermediate Statistics and

Econometrics: A Comparative Approach. Cambridge: The MIT Press.

[9] Bauwens, L., Lubrano, M. and Richard J.F. 1999. Bayesian Inference in Dynamic Econometric Models. Oxford: Oxford University Press.

[10] Lancaster Tony. 2004. An Introduction to modern Bayesian Econometrics, First edition, Oxford, U.K..

[11] Koop, G. 2005. Bayesian Econometrics. John Wiley & Sons, Ltd, England

[12] Greene, W.H. 2005. Econometric Analysis, fifth edition, Pearson Education, Inc.

[13] Bilstad, W.M. 2004. Introduction to Bayesian Statistics. John Wiley & Sons, Inc New York.

[14] Gujarati, D. 1995. Basic Econometrics, third edition, New York; McGraw-Hill.

[15] Kennedy, P. 2003. A Guide to Econometrics. The MIT Press, Cambridge. Massachusetts, UK.

[16] Efron, B. 1986. Why isn't everyone a Bayesian. *The American Statistician* 40: 1-11.

[17] Poirier, D.J. 1988. Frequentist and Subjectivist Perspectives on the Problems of Model Building in Economic. *Journal of Economic Perspectives* 2: 121-144 (commentary, 145-170).

[18] Olubusoye, O.E. and Osowole, I.O. 2007. A Monte Carlo Based Comparison of Bayesian and Frequentist Estimators of the Binomial proportion. JESA, pp. 44-49.

[19] Olubusoye, O.E. and Okewole, D.M. 2009. Prior Sensitivity in Bayesian Linear regression Model. *Int. Journal (Sciences)* 3(1): 21-29.