Thailand Statistician
July 2021; 19(3): 583-592
http://statassoc.or.th
Contributed paper

# A Comparative Study of Bayesian and Frequentist Testing for Model Comparison in Nested Model

**Oluwadare Olatunde Ojo**
Department of Statistics, Federal University of Technology Akure, Ondo State, Nigeria.
Corresponding author; e-mail: ojooo@futa.edu.ng

Received: 6 May 2020
Revised: 16 June 2020
Accepted: 11 August 2020

**Abstract**

Comparison of the two modelling approaches; frequentist and Bayesian had been carried out by various researchers in the area of estimation of parameters in regression model. Modelers are also interested in comparing different models to know the correct one. However, the use of relevant prior information about the data in the concept of model comparison for Bayesian can help to select the right model. This work facilitates comparison between frequentist and Bayesian methods in a nested model when all the parameters considered are evaluated at zero through a simulation study. The results from both simulation and real life data suggest that Bayesian approach with the use of Savage-Dickey density ratio (SDDR) provide a reasonable decision for the nested model than frequentist approach. It was also demonstrated that the SDDR is a true representation of Bayes factor.

**Keywords:** Bayes factor, Bayesian, frequentist, prior, Savage-Dickey density ratio.

## 1. Introduction

In statistical modelling, apart from estimation of parameters of models, some of the things a model builder would also wish to do are to compare different models. However, choosing the correct model by model builders and researchers can be difficult. Model comparison is a way by which models are compared to one another and the best model is chosen.

There are so many works that had been carried out on model comparison in Classical point of views. According to Harvey (1990), some classical model comparisons methods are Akaike's Information Criterion (AIC), Mallows's $C_p$ criterion, Amemiyas's Prediction criterion (APC), Davidson-Mackinon test etc. Most of these aforementioned classical methods of model comparison are used purposely for non-nested model.

In Bayesian framework, there are limited applications of model comparison and also Bayesian inference is more recent than classical method. Thus, some prominent works on Bayesian method of model comparison are; Smith and Spiegelhalter (1980), Griffiths and Wan (1994), O'Hagan (1995), Kass and Wasserman (1995), Diciccio et al. (2003), Plummer (2008), Feng and Giles (2009), Wetzels et al. (2010), and Li et al. (2017).

Smith and Spiegelhalter (1980) looked into forms of model choice criteria and compared alternative nested linear models on the basis of their asymptotic properties. Bayesian estimation of normal regression model with an uncertain inequality constraint was considered by Griffiths and Wan (1994). They adopted a non-informative prior and uncertainty concerning the inequality restriction that was represented by prior odds ratio while O'Hagan (1995) proposed a fractional Bayes factor for Bayesian comparison of models. The approach was found to be consistent, simple, robust and coherent.

Carlin and Chib (1995) summarizes a range of computational methods for obtaining estimates of Bayes factor. They provide a convenient method by using a marginal likelihood for comparing models by their fit with less theoretical problems attached to it than encountered when comparing non-nested models in a classical framework.

A generalized method Savage-Dickey Density Ratio (SDDR) for computing a Bayes factor was developed by Verdinelli and Wasserman (1995). It was observed that their methods in terms of computational complexity can be extended to other models. Kass and Wasserman (1995) computed a Bayes factor for testing equality restriction in the presence of nuisance parameter priors. Their results suggested that Schwartz criterion can provide sensible approximate solutions to Bayesian testing problems when the hypotheses are nested.

Several methods of estimating Bayes factors for non- linear models when it is possible to simulate observations from posterior distributions through MCMC and other techniques was examined by Diciccio et al**.** (1997). The simulated versions of Laplace's, Bartlett correction, importance sampling and reciprocal importance sampling techniques were considered. They found out that simulated version of Laplace's method was the most accurate approach among all the techniques.

A deviance-based loss function was derived by Plummer (2008) using a decision-theoretic framework. This approach was developed in order to capture some mixture models. The theoretical properties of the loss functions were examined in normal linear models and exponential family models using this penalized loss function. This approach was applied in mixture modelling and disease mapping.

Feng and Giles (2009) employed a posterior odds analysis to select the correct number of clusters for a Bayesian fuzzy regression analysis. They used a natural conjugate prior for parameters and concluded that the Bayesian posterior odds can provide a very powerful tool for choosing the number of clusters through their results.

Wetzels et al**.** (2010) in their work proposed an Encompassing Prior (EP) approach to facilitate Bayesian model selection for nested models with inequality constraints. EP approach generalizes the Savage-Dickey ratio method and can accommodate both inequality and exact equality constraints. Their EP approach was found to be computationally efficient procedure for calculating Bayes factor for nested models.

The plug-in predictive distribution as an alternative to DIC was also proposed by Li et al. (2017). This method was purposely to provide an asymptotically unbiased estimation to the new expected Kullback-Leibler (KL) divergence under a general framework. It was found out that this alternative DIC is easy for computation from Markov Chain Monte Carlo (MCMC) output and also has a smaller penalty term than the original DIC.

Most authors had considered the estimation of parameters of regression model by facilitating comparison between classical and Bayesian approaches (See Zellner 1976, Kleibergen and Zivot 2003, Adepoju and Ojo 2018 etc**.**). However, researchers are also interested in knowing the best method for choosing among competing models or hypotheses of a given phenomenon.

A model is said to be nested when one model is a special case of another model. Regression in a nested model for model comparison is of two types. The first entails comparison of $M_1$ which imposes R$\beta = r$ to $M_2$ which does not have this restriction while the second involves comparing $M_1 : y = X_1\ \beta_j + \varepsilon_1$ to $M_2 : y = X_2\ \beta_j + \varepsilon_2$, where $X_1$ and $X_2$ contain different explanatory variables, $j = 1,2$ (See Koop 2003).

Bayes factor is kind of method that is useful in the incorporating of external information into evidence about a model. Thus, Bayes factor through a SDDR will be derived for Bayesian to aid comparison with the frequentist approach. Therefore, this work hereby provides a comparative analysis on the performance of Bayesian using a Bayes factor and classical way of model comparison using a nested model to know the strength of these methods.

The paper is organized as follows. Section 2 gives the regression model and different Bayesian method of model comparison most especially the Bayes factor using Savage-Dickey Density Ratio (SDDR). It also gives a simple decision-theoretic justification model comparison. Section 3 provides procedures for numerical analysis and how the data will be analyzed. In Section 4, results of both classical and Bayesian are presented with decisions. Section 5 concludes the paper.

## 2. Materials and Methods

In this section, regression model and model comparison methods in both Classical and Bayesian approaches will be given.

### 2.1. Regression model

Consider the following regression model

$$y = \boldsymbol{x\beta} + u, \tag{1}$$

where $y$ is an $n \times 1$ vector of observations, $\boldsymbol{\beta}$ is a $k \times 1$ vector of unknown regression parameters, $\boldsymbol{x}$ is an $n \times k$ observed matrix of the regression, and $u$ is an $n \times 1$ vector of random errors which is normally distributed with mean zero and constant variance $\sigma^2$. And, $h = 1/\sigma^2$, where $h$ is the precision.

In econometrics, two kinds of models can be used for model comparison: (a) Nested model (b) Non-nested model. The nested model for model comparison involving equality restrictions using $M_1$ which has the form

$$M_1 : Q\beta = q,$$

against $M_2$ which does not imposes restrictions given as

$$M_2 : Q\beta \neq q,$$

will be used in this study, where $Q$ is $P \times k$ matrix and $q$ is P-vector.

Equation (1) allows for any R linear equality restrictions on the regression coefficient $\beta$ and it is also assume that rank (Q) = R. Hence, the two models as suggested by Raftery and Lewis (1996) can be simply defined as

$$H_0 : M_1 : \beta_i = 0 \text{ (reduced model)}, \tag{2}$$

$$H_1 : M_2 : \beta_i \neq 0 \text{ (unreduced model)}, i = 0,1,... \tag{3}$$

## 2.2. Classical model comparison method

Assume the regression model given in (1) is used and also follow the two models given in (2) and (3). Hence, the decision rule for testing the nested model of classical approach is stated below.

Decision rule: if $P < 0.05,$ reject $H_0$ and that means $M_2$ is supported than $M_1$ for data under consideration.

## 2.3. Bayesian model comparison method

There are so many methods that can be use for model comparison in Bayesian context, but some popular methods are however given below:

(i) Deviance Information Criterion (DIC),

(ii) Penalized expected deviance,

(iii) Bayes factor.

**Deviance Information Criterion (DIC):** It is a hierarchical generalization of AIC for models with weak prior information proposed by Spiegelhalter et al. (2002). This method is an approximation to penalized loss function based on deviance. However, this method was criticized due to under-penalizing of complex models especially in disease mapping situation (See Plummer 2008). The method is simply defined as

$$\text{DIC} = \overline{D} + p_D = D(\overline{\theta}) + 2p_D,$$

where $\overline{D}$ is the measure of fit and $p_D$ is the measure of complexity.

**Penalized expected deviance:** It is Bayesian model comparison method proposed by Plummer (2008). This method was proposed to overcome the problem of when the number of parameters is smaller than number of regressors in a regression model. This estimator is given as

$$L^e(Y,Z) = -2\int \log\{p(Y \mid \theta)\} P(\theta \mid Z) d\theta,$$

where $L^e$ is the loss function for expected deviance.

**Bayes Factor:** This approach of model comparison is known to be formal and widely accepted. It involves comparing the Posterior odds (PO) ratio of models. The PO is used for comparing two models.

Given two models, say 1 and 2, the PO can be simply stated as:

$$\text{PO}_{12} = \frac{P(M_1 \mid y)}{P(M_2 \mid y)} = \frac{P(y \mid M_1)P(M_1)}{P(y \mid M_2)P(M_2)} = BF_{12} = \frac{P(M_1)}{P(M_2)} = \frac{P(Q\beta = q \mid y)}{P(Q\beta \neq q \mid y)},$$

where $P(M_1)$ and $P(M_2)$ are prior model probabilities for models $M_1$ and $M_2,$ respectively. $P(y \mid M_1)$ and $P(y \mid M_2)$ are marginal likelihood for models $M_1$ and $M_2,$ respectively. $P(M_1 \mid y)$ and $P(M_2 \mid y)$ are posterior model probabilities for models $M_1$ and $M_2,$ respectively.

The Bayes factor, $BF_{12}$ quantifies the weight of evidence in favor of null hypothesis.

## 2.4. Savage-Dickey Density Ratio (SDDR)

Savage-Dickey Density Ratio (SDDR) as proposed by Dickey (1971) is a generalization of Bayes factor used to compare nested models. It is potentially applicable in variety of applications.

If the priors in two models $M_1$ and $M_2$ satisfy the condition below:

$$P(\gamma \mid \tau = \tau_0, M_2) = P(\tau \mid M_1),$$

then $BF_{12},$ the Bayes factor for comparing $M_1$ to $M_2,$ will has the form:

$$BF_{12} = \frac{P(\tau = \tau_0 \mid y, M_2)}{P(\tau = \tau_0 \mid M_2)},$$

where $P(\tau = \tau_0 \mid y, M_2)$ and $P(\tau = \tau_0 \mid M_2)$ are the unrestricted posterior and prior for $\tau$ evaluated at the point $\tau_0$ (See Koop 2003).

One major advantage of Savage-Dickey density ratio is that, it involves only $M_2$ and there is no need to worry about developing methods for posterior inference using $M_1$.

Thus, the Savage-Dickey Density Ratio (SDDR) as proposed by Dickey (1971), Verdinelli and Wasserman (1995) for this work will take the form

$$BF_{12} = \frac{P(\beta = \beta_0, h \mid y, M_2)}{P(\beta = \beta_0, h \mid M_2)}.$$

For a Bayes Factor (BF) where $M_1$ imposes equality restriction that is, $\beta_i = 0$, we have

$$BF_{12} = \frac{P(\beta_i, h \mid y, M_2)}{P(\beta_i, h \mid M_2)} = \frac{P(\beta_i = 0, h \mid y, M_2)}{P(\beta_i = 0, h \mid M_2)},$$

where

$$P(\beta_i, h \mid y, M_2) = \frac{P(\beta_i, h \mid M_2) L(\beta_i, h)}{P(y \mid M_2)} = \frac{P(\beta_i, h \mid M_2) L(\beta_i, h)}{\iint \pi(\beta_i, h) L(\beta_i, h) d\beta_i dh}. \tag{4}$$

Integrating (4) with respect to $h$ gives:

$$P(\beta_i \mid y, M_2) = \int P(\beta_i, h \mid y, M_2) dh = \int \left[ \frac{\pi(h \mid \beta_i)\pi(\beta_i) L(\beta_i, h)}{\iint (\beta_i, h) L(\beta_i, h) d\beta_i dh} \frac{\pi(h \mid \beta_i)\pi(\beta_i) L(\beta_i, h)}{\iint (\beta_i, h) L(\beta_i, h) d\beta_i dh} \right] dh, \tag{5}$$

Divide (5) by $P(\beta_i \mid M_2)$ and evaluate both at $\beta_i = 0$ implies:

$$\frac{P(\beta_i = 0 \mid y, M_2)}{P(\beta_i = 0 \mid M_2)} = \int \left[ \frac{\pi(h \mid \beta_i = 0)\pi(\beta_i = 0) L(h, \beta_i = 0)}{\pi(\beta_i = 0) \iint \pi(\beta_i, h) L(\beta_i, h) \mathrm{d}\beta_i \mathrm{d}h} \right] dh = \int \left[ \frac{\mathrm{P}(h \mid M_1) L(h, \beta_i = 0)}{\iint \pi(\beta_i, h) L(\beta_i, h) \mathrm{d}\beta_i \mathrm{d}h} \right] dh,$$

$$= \frac{P(y \mid M_1)}{P(y \mid M_2)} = BF_{12} = K.$$

Decision rule: According to Jeffreys (1946) and Kass and Raftery (1995), if $K < 1$, reject $H_0$ and that means $M_2$ is supported by the data under consideration than $M_1$.

## 3. Numerical Analysis

### 3.1. Simulation study

The model for this study is given as:

$$y = 0.1 + 0.4x_1 + 2x_2 + 4.5x_3 + u. \tag{6}$$

The following will be used for simulation of data in the study:

(i) The explanatory variables will be generated as: $x_i \sim N(0,1)$, $i = 1, 2, 3$.

(ii) The error components is obtained as: $u \sim N(0, h^{-1})$, where $h = 1$.

(iii) The explanatory variables, error term and regression coefficients specified in (6) will be used to generate the dependent variable.

(iv) The sample sizes are $N$=10, 30, 100, 500, 700 and 1,000.

(v) The sample sizes will be replicated 100,000.

### 3.2. Real data analysis

Here, the frequentist and Bayesian approaches for model comparison in a nested regression model are applied to a real data. Data on US defence budget outlays for 20 years from1962 to 1981 are used to know the effects on Gross National Product (GNP), US military sales/assistance and aerospace industry sales on it. The defence budget-outlay, GNP, US military sales/assistance, and aerospace industry sales are measured in $ billions. The data is sourced from Gujarati and Porter (2005) through various government publications of United States.

Hence, we consider the regression model given as

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + u,$$

where $y$ is the defence budget-outlay, $x_1$ is the GNP, $x_2$ is the US military sales/assistance and $x_3$ is the aerospace industry sales. For the simulation study and real life data, the model to be considered for comparison is given as

$$M_1 : \beta_i = 0 \text{ (reduced model)},$$
$$M_2 : \beta_i \neq 0 \text{ (unreduced model)}, \ i = 1,2,3.$$

## 4. Results and Discussion

The objective of this study is to compare frequentist and Bayesian approaches for model comparison in a nested regression model. This section gives the results obtained from Bayesian and classical approaches with various decisions for parameters. Here, $M_2$ is an unreduced model and represents a good model. The Bayes factor, (K) gives the result for Bayesian using Savage Dickey Density Ratio (SDDR) while the p-value for the t-distribution is also presented for frequentist approach.

**Table 1** Model comparison of nested model for Bayesian and classical methods with sample size of 10

| Parameter | Estimators | Estimates $(K/\Pr(>\|t\|))$ | Decision |
|---|---|---|---|
| $\beta_0$ | Bayesian | 0.000859 | $M_2$ |
| | Classical | 0.151000 | $M_1$ |
| $\beta_1$ | Bayesian | 0.001174 | $M_2$ |
| | Classical | 0.067300 | $M_1$ |
| $\beta_2$ | Bayesian | 0.000801 | $M_2$ |
| | Classical | 1.12E-05 | $M_2$ |
| $\beta_3$ | Bayesian | 0.000123 | $M_2$ |
| | Classical | 1.12E-05 | $M_2$ |

**Table 2** Model comparison of nested model for Bayesian and Classical methods with sample size of 30

| Parameter | Estimators | Estimates $(K/\Pr(>\|t\|))$ | Decision |
|---|---|---|---|
| $\beta_0$ | Bayesian | 2.4E-14 | $M_2$ |
| | Classical | 0.429700 | $M_1$ |
| $\beta_1$ | Bayesian | 3.31E-14 | $M_2$ |
| | Classical | 0.574500 | $M_1$ |
| $\beta_2$ | Bayesian | 2.18E-14 | $M_2$ |
| | Classical | 0.043400 | $M_2$ |
| $\beta_3$ | Bayesian | 1.12E-14 | $M_2$ |
| | Classical | 1.67E-07 | $M_2$ |

Results obtained from both Tables 1 and 2 show that Bayesian method of model comparison picked a good model for all the parameters while the classical method only picked a good model for parameters $\beta_2$ and $\beta_3$ for sample size of 10 and 30.

**Table 3** Model Comparison of nested model for Bayesian and classical methods with sample size of 100

| Parameter | Estimators | Estimates $(K/\Pr(>\|t\|))$ | Decision |
|---|---|---|---|
| $\beta_0$ | Bayesian | 0.00E+00 | $M_2$ |
| | Classical | 0.783300 | $M_1$ |
| $\beta_1$ | Bayesian | 0.00E+00 | $M_2$ |
| | Classical | 0.001900 | $M_2$ |
| $\beta_2$ | Bayesian | 0.00E+00 | $M_2$ |
| | Classical | 2.00E-16 | $M_2$ |
| $\beta_3$ | Bayesian | 0.00E+00 | $M_2$ |
| | Classical | 0.00E+00 | $M_2$ |

In Tables 3-6, Bayesian method of model comparison give more evidence against the reduced model $M_1$ for all the parameters considered. The classical approach gives evidence of support for most especially parameter $\beta_2$ and $\beta_3$ in Tables 3-6. It was observed that parameter $\beta_0$ does not give support for the unreduced model for classical method of model comparison for all the sample sizes as shown in Tables 1-6.

**Table 4** Model Comparison of nested model for Bayesian and classical methods with sample size of 500

| Parameter | Estimators | Estimates ($K/\Pr(>\|t\|)$) | Decision |
|---|---|---|---|
| $\beta_0$ | Bayesian | 4.04E-67 | $M_2$ |
| | Classical | 0.866080 | $M_1$ |
| $\beta_1$ | Bayesian | 5.76E-68 | $M_2$ |
| | Classical | 0.003390 | $M_2$ |
| $\beta_2$ | Bayesian | 3.31E-67 | $M_2$ |
| | Classical | 4.34E-05 | $M_2$ |
| $\beta_3$ | Bayesian | 8.01E-68 | $M_2$ |
| | Classical | <2.05 | $M_2$ |

**Table 5** Model Comparison of nested model for Bayesian and classical methods with sample size of 700

| Parameter | Estimators | Estimates ($K/\Pr(>\|t\|)$) | Decision |
|---|---|---|---|
| $\beta_0$ | Bayesian | 0 | $M_2$ |
| | Classical | 0.301000 | $M_1$ |
| $\beta_1$ | Bayesian | 0 | $M_2$ |
| | Classical | 0.247000 | $M_1$ |
| $\beta_2$ | Bayesian | 0 | $M_2$ |
| | Classical | 2.00E-16 | $M_2$ |
| $\beta_3$ | Bayesian | 0 | $M_2$ |
| | Classical | 2.00E-16 | $M_2$ |

**Table 6** Model Comparison of nested model for Bayesian and classical methods with sample size of 1,000

| Parameter | Estimators | Estimates ($K/\Pr(>\|t\|)$) | Decision |
|---|---|---|---|
| $\beta_0$ | Bayesian | 0 | $M_2$ |
| | Classical | 0.453120 | $M_1$ |
| $\beta_1$ | Bayesian | 0 | $M_2$ |
| | Classical | 0.005230 | $M_1$ |
| $\beta_2$ | Bayesian | 0 | $M_2$ |
| | Classical | 2.00E-16 | $M_2$ |
| $\beta_3$ | Bayesian | 0 | $M_2$ |
| | Classical | 2.00E-16 | $M_2$ |

**Table 7** Model Comparison of nested model for Bayesian and classical methods in real life data application

| Parameter | Estimators | Estimates $(K/\Pr(>\|t\|))$ | Decision |
|---|---|---|---|
| $\beta_0$ | Bayesian | 0.0000 | $M_2$ |
| | Classical | 0.0000 | $M_2$ |
| $\beta_1$ | Bayesian | 0.0008 | $M_2$ |
| | Classical | 0.0300 | $M_1$ |
| $\beta_2$ | Bayesian | 0.0000 | $M_2$ |
| | Classical | 0.1450 | $M_1$ |
| $\beta_3$ | Bayesian | 0.0000 | $M_2$ |
| | Classical | 0.0000 | $M_2$ |

In results obtained from the real life data application, Bayesian method of model comparison give more evidence against the reduced model $M_1$ for all the parameters considered. This also agreed with the results obtained in simulation.

## 5. Conclusion

Overtime, different work had been carried out by facilitating comparison between the two modelling approaches, that is, frequentist and Bayesian methods in the area of parameter estimation. It is of interest for model builders to know the strength of the two methods in the area of model comparison.

In this paper, comparison between the Bayesian and frequentist approaches using a nested model was carried out to know the strength of the two methods. Bayes factor through a Savage-Dickey density ratio when parameters are evaluated at zero was provided for Bayesian method in order to facilitate comparison.

Data were simulated for different sample sizes and each of the sample sizes was replicated while a real data set was also applied. For all the sample sizes considered, Bayesian performs well than the frequentist approach by giving an overwhelming evidence support for model $M_2$ for all the parameters.

Therefore, the Bayesian method of model comparison outperformed the frequentist method for nested model. Thus, Bayesian method using a Savage-Dickey density ratio is preferred to frequentist approach where a researcher has choice between the two approaches.

## References

Adepoju AA, Ojo OO. Bayesian method for solving the problem of multicollinearity in regression. Afrika Statistika. 2018; 13(3): 1823-1834.

Carlin BP, Chib S. Bayesian model choice via Markov chain Monte Carlo methods, J Roy Stat Soc B Met. 1995; 57(3): 473-484.

Diciccio TJ, Kass RE, Raftery A, Wasserman L. Computing Bayes factors by combining simulation and asymptotic approximations. J Am Stat Assoc. 1997; 92(439): 903-915.

Dickey JM. The weighted likelihood ratio, linear hypotheses on normal location parameters. Ann Math Stat. 1971; 42(1): 204-223.

Feng H, Giles DE. Bayesian fuzzy regression analysis and model selection: theory and evidence. Econometrics working paper EWP 0903. 2009.

Griffiths WE, Wan ATK. A Bayesian estimator of the linear regression model with an uncertain inequality constraint. Discussion paper 74. 1994.

Gujarati DN, Porter DC. Basic econometrics, New York: McGraw Hill; 2005.

Harvey AC. The econometric analysis of time series, Cambridge: The MIT press; 1990.

Jeffreys H. An invariant form for the prior probability in estimation problems. Proceedings of the Royal Statistical Society of London series A. 1946; 186(1007): 453-461.

Kass R, Raftery A. Bayes Factors. J Am Stat Assoc. 1995; 90(430): 773-795.

Kass RE. Wasserman L. A reference Bayes test for nested hypothesis and its relationship to the schwarz criteria. J Am Stat Assoc. 1995; 90(431): 928-934.

Kleibergen F, Zivot E. Bayesian and classical approaches to instrumental variable regression. J Econometrics. 2003; 114(1): 29-72.

Koop G. Bayesian Econometrics, Wiley, Chichester. 2003.

Li Y, Yu J, Zeng T. Deviance information criterion for Bayesian model selection: Justification and variation. Research Collection School of Economics, Singapore Management University. 2017; 1-40.

O'Hagan A. Fractional Bayes factors for model comparison. J Roy Stat Soc B Met. 1995; 57(1): 99-138.

Plummer M. Penalized loss functions for Bayesian model comparison. Biostatistics. 2008; 9(3): 523-539.

Raftery AE, Lewis SM. Implementing MCMC. In: Gilks WR, Spiegelhalter DJ, Richardson S. editors. Markov Chain Monte Carlo in Practice. London: Chapman and Hall; 1996. p. 115-130.

Smith AFM, Spiegelhalter DJ. Bayes factors and choice criteria for linear models. J. R. Statist. Soc. B 1980; 42(2): 213-220.

Spiegelhalter DJ, Best NG, Carlin BP, Van Der Linde. A Bayesian measures of model complexity and fit. J Roy Stat Soc B Met. 2002; 64(4): 583-639.

Verdinelli I, Wasserman L. Computing Bayes factors using a generalization of the Savage-Dickey density ratio. J Am Stat Assoc. 1995; 90(430): 614-618.

Wetzels R, Grasman RPPP, Wagenmakers EJ. An encompassing prior generalization of the Savage-Dickey density ratio. Comput Stat Data An. 2010; 54(9): 2094-2102.

Zellner A. Bayesian and non-Bayesian analysis of the regression model with multivariate student-t errors terms. J Am Stat Assoc. 1976; 71(354): 400-405.