Journal of Statistical and Econometric Methods, vol.3, no.1, 2014, 85-96
ISSN: 2241-0384 (print), 2241-0376 (online)
Scienpress Ltd, 2014

The Bayesian Approach to Multi-equation Econometric Model Estimation

Dorcas Modupe Okewole[1] and Olusanya E. Olubusoye[2]

Abstract

The Bayesian approach to Statistical inference enables making probability statements about parameter (s) of interest in any research work. This paper presents a study of multi-equation Econometric model estimation from the flat or locally uniform prior Bayesian approach. Using a two-equation model containing one just-identified and one over-identified equation, we present a Bayesian analysis of multi-equation econometric model as well as a Monte Carlo study on it, using WinBUGS (windows version of the software: Bayesian analysis Using Gibbs Sampling). In studies involving the use of flat or locally uniform prior, it is the usual practice to specify the prior distribution in such a way that the variance is large. However, the definition of this “large” variance could vary among researchers. This research work considered three different variances (10, 100 and 1000) with particular focus on the Mean squared error of posterior estimates, looking for possible sensitivity to the prior variance specification. The result of the Monte Carlo study showed that a prior variance 10 gave the smallest Mean squared error out of the three stated prior variances. As reflected in the kernel density plots, the distribution of the Posterior estimates from the prior variance 10 was the closest to the t distribution obtained theoretically.

Mathematics Subject Classification: 62J05; 62J07

[1] Department of Mathematical Sciences, Redeemer’s University, Ogun State, Nigeria.
[2] Department of Statistics University of Ibadan, Nigeria.

Article Info: *Received* : November 29, 2013. *Revised* : January 12, 2014.
Published online : February 7, 2014.

Keywords: Bayesian approach; Multi-equation; Prior variance; Mean squared error

1 Introduction

The application of statistics in economic modeling has brought about several research works on multi-equation models such as general linear models (GLM), vector autoregressive models (VAR), seemingly unrelated regression models (SURE), simultaneous equations models (SEM) among several others. Studies on these models have so far been mostly classical, that is, the conventional methods such as least squares, maximum likelihood, generalized method of moments, etc. The Bayesian approach has become more attractive now because of the availability of numerical intensive software and high speed computing technology that make the analysis easier to handle than the usual rigorous and at times, intractable mathematics involved.

There are many approaches to Bayesian analysis; the most common ones are the objective, subjective, robust, frequentist-Bayes and quasi-Bayes approaches. Beginning with the first set of Bayesians, Thomas Bayes (1783) and Laplace (1812), who carried out Bayesian analysis using constant prior distribution for unknown parameters, Bayesian analysis has been taken as an objective theory. The use of uniform or flat prior, more generally known as noninformative, is a common objective Bayesian approach, Jeffrey's prior as presented in Jeffrey (1961) is the most popular in this school of thought. Although these priors are often referred to as noninformative prior, they also reflect certain informative features of the system being analysed, in fact, some Bayesians have argued that it rather be referred to as "weakly informative" prior for example, German et. al (2008).

This paper is focused on the objective Bayesian approach, being the most commonly used. Using a two-equation model, flat prior was stated for the regression coefficients while a Wishart distribution with zero degree of freedom was stated for the variance-covariance matrix of the residual terms of the model. The prior distributions were then combined with the likelihood function to have the posterior. A Monte Carlo study was then carried out to illustrate the use of WinBUGS for a multi-equation model such as the one considered in this paper. Specifically, the variance of the prior distribution of the regression coefficients was stated at three levels as; 10, 100, and 1000 which corresponds to precision 0.1, 0.01 and 0.001.

Section two contains theoretical background; section three contains methodology and design of the Monte Carlo experiment; results and interpretations are presented in section four, while the last section concludes the paper.

2 Preliminary Notes

A two-equation model was considered as follows;

$$y_{1t} = \gamma y_{2t} + \beta_{11} X_{1t} + u_{1t}$$
$$y_{2t} = \beta_{21} X_{1t} + \beta_{22} X_{2t} + \beta_{23} X_{3t} + u_{2t} \quad (2.1)$$

y_{1t} and y_{2t} are each $(T\text{x}1)$ vectors containing observations on the endogenous variables

X_{1t}, X_{2t}, X_{3t} are each $(T\text{x}1)$ vectors of observations on the exogenous variables

γ is the scalar coefficient of the endogenous explanatory variable

$\beta_{11}, \beta_{21}, \beta_{22}, \beta_{23}$ are the scalar coefficients of the predetermined explanatory variables.

u_{1t} and u_{2t} are each $(T\text{x}1)$ random disturbance terms.

We carried out the Bayesian analysis by working directly on the structural model (2.1) which we write in matrix form as;

$$Y = X\delta + U \quad (2.2)$$

Where $Y = \begin{pmatrix} y_{1t} \\ y_{2t} \end{pmatrix}$, $X = \begin{pmatrix} y_{2t} & X_{1t} & 0 & 0 \\ 0 & X_{1t} & X_{2t} & X_{3t} \end{pmatrix}$, $\delta = \begin{pmatrix} \gamma & 0 \\ \beta_{11} & \beta_{21} \\ 0 & \beta_{22} \\ 0 & \beta_{23} \end{pmatrix}$ and $U = \begin{pmatrix} u_{1t} \\ u_{2t} \end{pmatrix}$

We make the following assumptions about the model (2.2)

(1) The assumption about the exogenous regressors or predetermined variable: it is assumed that the matrix of the exogenous regressors is of full rank, i.e., the number of independent columns of the matrix X, which in the case of model (2.2) is three.

(2). The second assumption is the major one which is about the residual term U, the summary of this assumption is;

$$U \sim NIID(0, \Sigma) \quad (2.3)$$

2.1 Prior Probability density function

The flat or locally-uniform prior is assumed for the parameters of our models. The idea behind the use of this prior is to make inferences that are not greatly affected by external information or when external information is not available. Two rules were suggested by Jeffrey (1961) to serve as guide in choosing a prior distribution. The first one states as, "If the parameter may have any value in a finite range, or from $-\infty$ to $+\infty$, its prior probability should be taken as uniformly distributed". While the second is that if the parameter, by nature, can take any value from 0 to ∞, the prior probability of its logarithm should be taken as uniformly distributed.

For the purpose of this study, we assume that little is known, a priori, about the elements of the parameter δ, and the three distinct elements of Σ. As the prior pdf, we assume that the elements of δ and those of Σ are independently distributed; that is,

$$P(\delta, \Sigma) = P(\delta)P(\Sigma) \tag{2.4}$$

$$P(\delta) = \text{constant} \tag{2.5}$$

$$P(\Sigma) \propto |\Sigma|^{-3/2} \tag{2.6}$$

By denoting $\sigma^{\mu\mu'}$ as the $(\mu, \mu')th$ element of the inverse of Σ, and the Jacobian of transformation of the three variances, $(\sigma_{11}, \sigma_{12}, \sigma_{22})$ to $(\sigma^{11}, \sigma^{12}, \sigma^{22})$ as

$$J = \left| \frac{\partial(\sigma_{11}, \sigma_{12}, \sigma_{22})}{\partial(\sigma^{11}, \sigma^{12}, \sigma^{22})} \right| = |\Sigma|^3 \tag{2.7}$$

The prior pdf in (2.6) implies the following prior pdf on the three distinct elements of Σ^{-1}

$$P(\Sigma^{-1}) \propto |\Sigma^{-1}|^{-\frac{3}{2}} \tag{2.8}$$

This could also be seen as taking an informative prior pdf on Σ^{-1} in the Wishart pdf form and allowing the “degrees of freedom” in the prior pdf to be zero. With zero degrees of freedom, there is a “spread out” Wishart pdf which then serve as a diffuse prior pdf since it is diffuse enough to be substantially modified by a small number of observations. The Wishart distribution is the conjugate for the multivariate normal distribution, which is the distribution of the variance-covariance matrix (Σ).

Hence, our prior p.d.f‘s are (2.5), (2.6), and (2.8), as obtained also by Zellner (1971), Geisser (1965) and others for parameters of similar models as in (2.2).

2.2 Likelihood Function

The likelihood function for δ and Σ, which follows from the assumption that rows of U in equation 2.2 are normally and independently distributed, each with zero mean vector and 2x2 covariance matrix Σ, is given as;

$$L(\delta, \Sigma / Y, X) \propto |\Sigma|^{-n/2} \exp[-\tfrac{1}{2} tr(Y - X\delta)'(Y - X\delta)\Sigma^{-1} \tag{2.9}$$

This is the same as;

$$L(\delta,\Sigma/Y,X)\propto|\Sigma|^{-n/2}\exp[-\tfrac{1}{2}trS\,\Sigma^{-1}-\tfrac{1}{2}tr(\delta-\hat{\delta})'X'X(\delta-\hat{\delta})\Sigma^{-1}] \qquad (2.10)$$

where $(Y-X\delta)'(Y-X\delta)=(Y-X\hat{\delta})'(Y-X\hat{\delta})+(\delta-\hat{\delta})'X'X(\delta-\hat{\delta})$,

$=S+(\delta-\hat{\delta})'X'X(\delta-\hat{\delta})$

$S=(Y-X\hat{\delta})'(Y-X\hat{\delta})$ and $\hat{\delta}$ is the estimate of δ.

Thus, the likelihood function for the parameters is as given in (2.10).

2.3 The Posterior Pdf

Combining the Prior pdf (2.5) and (2.8) with the likelihood function (2.10), we have the joint posterior distribution for δ and Σ^{-1} given as;

$$P(\delta,\Sigma^{-1}/Y,X)\propto|\Sigma|^{-\left(n+3/2\right)}EXP\{-\tfrac{1}{2}tr\left[S+\left(\delta-\hat{\delta}\right)'Z'Z\left(\delta-\hat{\delta}\right)\right]\Sigma^{-1}\} \qquad (2.11)$$

Integrating (2.11) with respect to Σ^{-1}, we have the marginal posterior pdf for δ given as:

$$P(\delta/Y,X)\propto\left[S+(\delta-\hat{\delta})'Z'Z(\delta-\hat{\delta})\right]^{-T/2} \qquad (2.12)$$

A pdf in the generalized student-t form.

2.4 The Design and methodology of the experiment

2.4.1 Generating data for the experiment

Monte Carlo simulation approach was used in this research work. The data was generated by arbitrarily fixing values for the parameters of the model and stating specific distributions for the predetermined variables and the error terms. We considered two runs, negatively correlated residual terms for the first run and positively correlated residual terms for the second run. Values stated for the parameters are as follows;

$\gamma=3.0,\beta_{11}=1.0,\beta_{21}=2.0,\beta_{22}=0.5,\beta_{23}=1.5$.

The two runs are stated as follows.

RUN ONE:

$X_{1t}:NID(0,1),X_{2t}:NID(0,1),X_{3t}:NID(0,1),(u_{1t}u_{2t}):NID(0,0;\sigma_{11},\sigma_{12},\sigma_{22}),$

$\sigma_{11}=1.0,\sigma_{12}=-1.0,\sigma_{22}=4.0$

RUN TWO:

$X_{1t}:NID(0,1),X_{2t}:NID(0,1),X_{3t}:NID(0,1),(u_{1t}u_{2t}):NID(0,0;\sigma_{11},\sigma_{12},\sigma_{22}),$

$\sigma_{11}=1.0,\sigma_{12}=1.0,\sigma_{22}=4.0$

Also, three prior variances are stated as 10, 100 and 1000, which in the WinBUGS code is usually stated as precision; 0.1, 0.01 and 0.001 respectively.

In each of these runs, $N = 5000$ samples of size $T = 20, 40, 60,$ and 100 were generated, that is, the number of replicates is 5000, making a total of 20,000 samples in one run, and 40,000 samples altogether. We represent number of replicates with N and sample size with T.

2.4.2 Analysis of the data

WinBUGS was used for the analysis. As earlier mentioned, WinBUGS is windows version of the software referred to as Bayesian Analysis using Gibbs sampling. A sub-program is usually written by the researcher specifying the model and the prior distributions, and then WinBUGS uses Markov Chain Monte Carlo simulation to draw samples from the posterior distribution. Here, we first carried out 1000 iterations after which we observed sign of convergence, then a further 5000 iterations were carried out and the first 1000 taken as 'burn in'. See Gilks et. al.(1996), for detailed information on how to check convergence.

2.4.3 Criteria for assessing the performance of the Estimators

There is a number of comparison criteria used in literature, however, we made use of the bias and Mean Squared Error (MSE).

There are 5000 replicates so estimated bias $= \frac{1}{5000}\sum_{i=1}^{N=5000} \hat{\theta}_i - \theta$

The mean squared error, for an estimator of a parameter θ, is given as;

$MSE(\hat{\theta}) = E(\theta - \hat{\theta})^2$ = Var($\hat{\theta}$) + (Estimated bias)^2

where $Var(\hat{\theta}) = \frac{1}{N_r}\sum_{1}^{N_r}\hat{\theta}^2 - (\frac{1}{N_r}\sum_{1}^{N_r}\hat{\theta})^2$ and N_r is number of replications. The kernel density of the estimates of the regression coefficients was also plotted and compared for the three prior variances with the density of the t distribution obtained theoretically.

3 Main Results

The Kernel of the Posterior mean was plotted for γ with the three prior variance levels (10, 100, and 1000). The choice of γ was because it is the only parameter mostly affected by the other equation of the model. The t distribution was then imposed on the plots to show how close they are to the t distribution obtained theoretically; these plots are presented in Figures 1 to 3.

The results of the Monte Carlo experiment are presented in Tables 2 to 5 in the appendix. Tables 2 and 3 contain the results from run 1 while the results for run 2 are contained in Tables 4 and 5. However a summary of these results, in terms of the number of times each prior variance level produced the least bias and MSE, is as shown in Table 1.

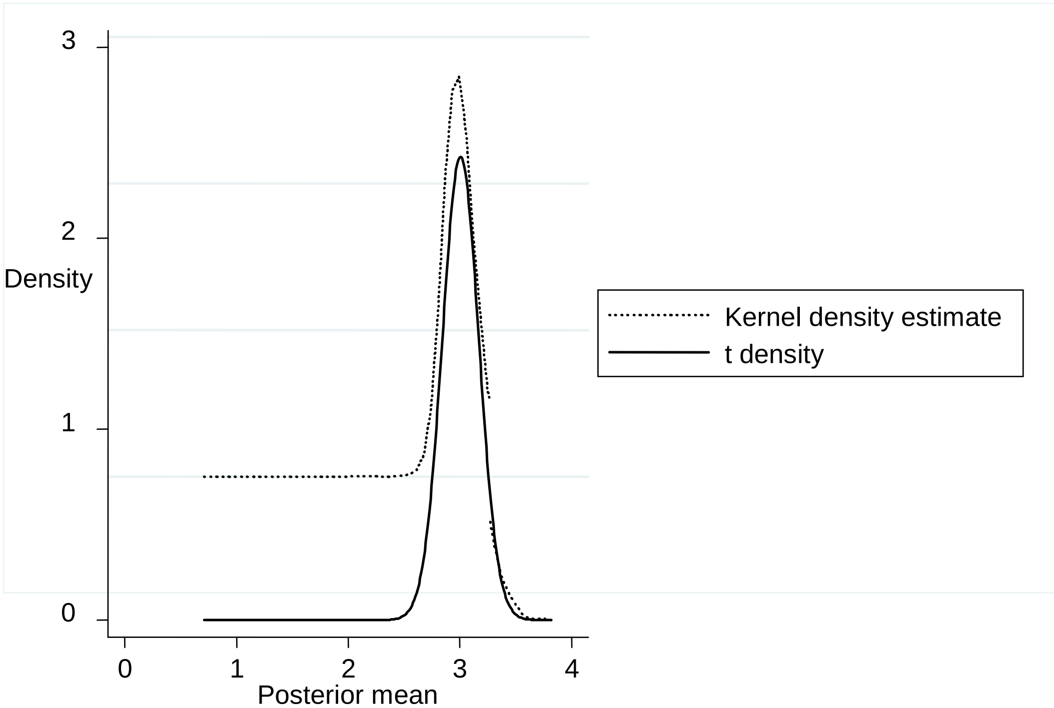

Figure 1: Kernel density plot with prior variance 10

From the three plots as shown in Figure 1, 2, 3, the Kernel density plot with prior variance 10 is the closest to the t distribution, this is a suggestion that the use of variance 10 might produce better posterior estimates than the other two higher variances (100 and 1000).

Table 1: Summary on cases with least bias and MSE for each prior variance level

	Variance 1000		Variance 100		Variance 10	
	RUN 1	RUN 2	RUN 1	RUN 2	RUN 1	RUN 2
ABS BIAS	5	12	5	4	11	5
MSE	0	2	1	1	17	14

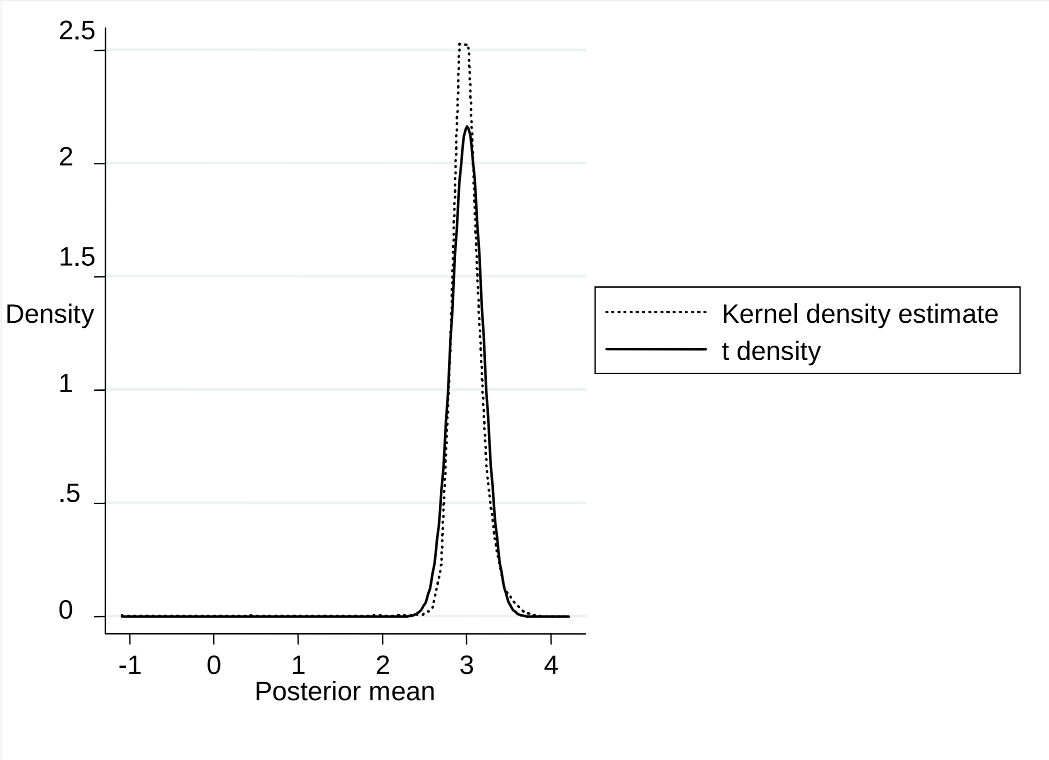

Figure 2: Kernel density plot with prior variance 100

In terms of bias, prior variance 10 produced the best result in run 1, having the highest number of cases (11) in which it gave the smallest bias out of the three prior variances. The result from run 2 however showed prior variance 1000 as the best in terms of bias, having 12 cases in which it produced the smallest bias. Run 1 is the case of negatively correlated residual terms while run 2 is the case of positively correlated residual terms. Hence the results suggest that, when the residual terms of a multi-equation model are negatively correlated, a prior variance 10 is most likely to produce posterior estimates with the least bias while in the case of positively correlated residuals, prior variance 1000 is the most likely to produce posterior estimates with the least bias. However, if the two runs are not considered separately, there is an indication that the two prior variances have similar performance in terms of bias. Concerning the mean squared error (MSE) of estimates, unlike the bias, there is a consistent clear difference in the performance of these prior variances between the two runs of the experiment. As expected, Since the Bayesian approach can be seen as updating the prior information about a parameter of interest, the least prior variance gave the least MSE of the posterior estimates.

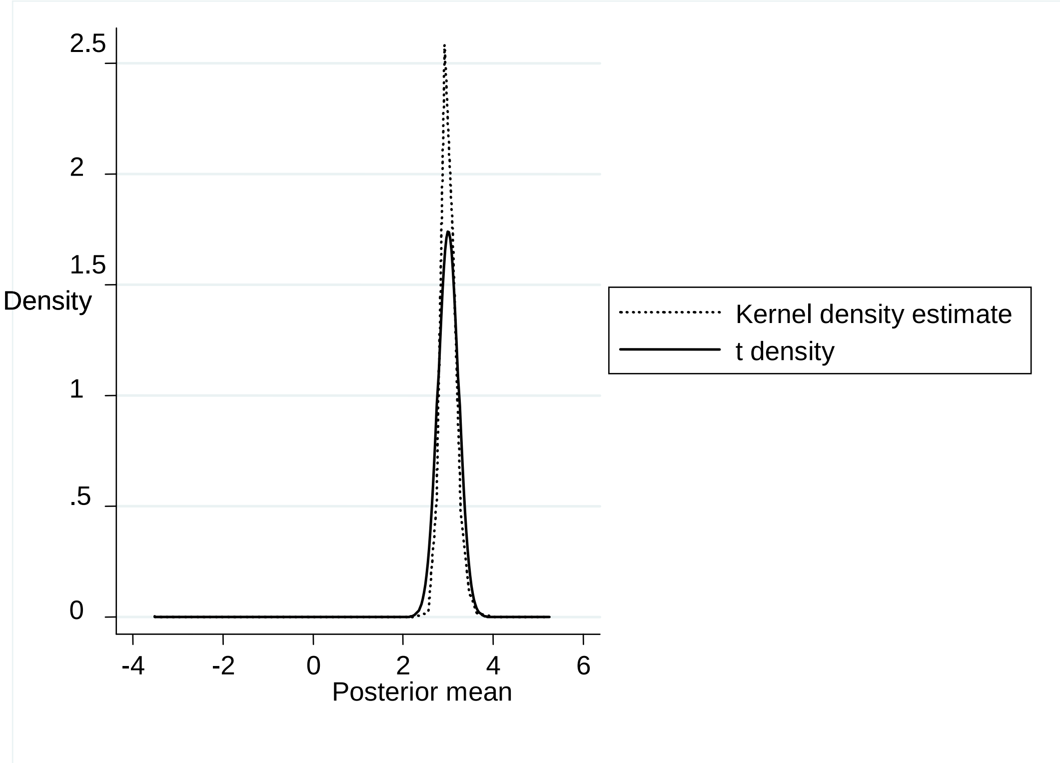

Figure 3: Kernel density plot with prior variance 1000

The results as presented in Table 1 shows prior variance 10 as having the highest number of times (17 from run1 and 14 from run 2) with the least MSE in the whole experiment. In most cases as can be seen in Tables 2 to 5, the MSE reduced with reduction in prior variance. Since there are cases in which prior variance 10 gave the least bias and other cases in which prior variance 1000 gave the least bias, the result from the MSE might be a better suggestion of which prior variance to be used among the three.

5 Conclusion

This research paper was focused on the Bayesian approach in Multi-equation models with some attention on the use of flat or locally-uniform prior. The Monte Carlo experiment carried out on three prior variance levels 10, 100 and 1000 brought about suggestion on which prior variance will produce "best" posterior estimates. From the results, the definition of "best" is in terms of efficiency (using MSE) and closeness of the distribution to the t distribution obtained theoretically. The Bayesian method with the prior variance 10 came up topmost having the highest number of times with the least MSE and also showing the closest kernel density plot to the t distribution. This result thus brings up more questions such as what the result will look like with the use of a prior variance less than 10 or greater than 1000.

Appendix

Table 2: Results from Run 1 sample size 20 and 40

Parameter		Variance 1000		Variance 100		Variance 10	
		T=20	T=40	T=20	T=40	T=20	T=40
$\gamma(3.0)$	Mean	3.0050	2.9925	3.0084	2.9955	3.0067	3.0034
	ABS Bias	0.0050	0.0075	0.0084	0.0045	0.0067	0.0034
	MSE	0.0522	0.0328	0.0339	0.0212	0.0269	0.0091
$\beta_{11}(1.0)$	Mean	1.0042	1.1124	0.9805	1.0594	0.9908	0.9991
	ABS Bias	0.0042	0.1124	0.0195	0.0594	0.0092	0.0009
	MSE	0.5814	0.8094	0.2734	0.3035	0.2133	0.1240
$\beta_{21}(2.0)$	Mean	2.0166	2.0037	2.0058	2.0009	1.9258	1.9807
	ABS Bias	0.0166	0.0037	0.0058	0.0009	0.0742	0.0193
	MSE	0.4698	0.1199	0.4641	0.1196	0.4267	0.1175
$\beta_{22}(0.5)$	Mean	0.4550	0.4811	0.4594	0.4736	0.4917	0.4868
	ABS Bias	0.0450	0.0189	0.0406	0.0264	0.0083	0.0132
	MSE	0.1740	0.0661	0.1728	0.0689	0.1654	0.0642
$\beta_{23}(1.5)$	Mean	1.3745	1.4521	1.3759	1.4434	1.3715	1.4541
	ABS Bias	0.1255	0.0479	0.1241	0.0566	0.1285	0.0459
	MSE	0.2598	0.1154	0.2521	0.1280	0.2213	0.0985

Table 3: Results from Run 1 sample sizes 60 and 100

Parameter		Variance 1000		Variance 100		Variance 10	
		T=60	T=100	T=60	T=100	T=60	T=100
$\gamma(3.0)$	Mean	2.9882	3.0003	2.9918	3.0004	3.0017	3.0012
	ABS Bias	0.0118	0.0003	0.0082	0.0004	0.0017	0.0012
	MSE	0.0571	0.0036	0.0361	0.0036	0.0163	0.0036
$\beta_{11}(1.0)$	Mean	1.1798	1.1036	1.1139	1.0903	1.0198	1.0405
	ABS Bias	0.1798	0.1036	0.1139	0.0903	0.0198	0.0405
	MSE	1.3421	0.7578	0.5870	0.5939	0.1306	0.1644
$\beta_{21}(2.0)$	Mean	2.0051	2.0000	2.0031	1.9990	1.9899	1.9925
	ABS Bias	0.0051	0.0000	0.0031	0.0010	0.0101	0.0075
	MSE	0.0750	0.0379	0.0749	0.0380	0.0740	0.0380
$\beta_{22}(0.5)$	Mean	0.4717	0.4847	0.4731	0.4847	0.4877	0.4874
	ABS Bias	0.0283	0.0153	0.0269	0.0153	0.0123	0.0126
	MSE	0.0479	0.0442	0.0470	0.0441	0.0406	0.0414
$\beta_{23}(1.5)$	Mean	1.4238	1.4705	1.4264	1.4704	1.4437	1.4677
	ABS Bias	0.0762	0.0295	0.0736	0.0296	0.0563	0.0323
	MSE	0.1502	0.0445	0.1447	0.0443	0.1068	0.0445

Table 4: Results from Run 2 sample sizes 20 and 40

Parameter		Variance 1000		Variance 100		Variance 10	
		T=20	T=40	T=20	T=40	T=20	T=40
$\gamma(3.0)$	Mean	2.6942	2.9890	2.9841	2.9890	2.9781	2.9851
	ABS Bias	0.3058	0.0110	0.0159	0.0110	0.0219	0.0149
	MSE	0.1947	0.0141	0.0309	0.0133	0.0293	0.0134
$\beta_{11}(1.0)$	Mean	0.9968	1.0210	1.0239	1.0178	1.0122	1.0129
	ABS Bias	0.0032	0.0210	0.0239	0.0178	0.0122	0.0129
	MSE	0.0884	0.1101	0.1884	0.0787	0.1623	0.0754
$\beta_{21}(2.0)$	Mean	1.3654	1.9893	1.9845	1.9856	1.9264	1.9501
	ABS Bias	0.6346	0.0107	0.0155	0.0144	0.0736	0.0499
	MSE	0.5847	0.1565	0.2567	0.1562	0.2497	0.1542
$\beta_{22}(0.5)$	Mean	0.3187	0.4869	0.4537	0.4867	0.4570	0.4905
	ABS Bias	0.1813	0.0131	0.0463	0.0133	0.0430	0.0095
	MSE	0.1510	0.0997	0.2604	0.0993	0.2503	0.0972
$\beta_{23}(1.5)$	Mean	1.0059	1.4405	1.3859	1.4386	1.3720	1.4224
	ABS Bias	0.4941	0.0595	0.1141	0.0614	0.1280	0.0776
	MSE	0.3995	0.1423	0.2630	0.1416	0.2361	0.1370

Table 5: Results from Run 2 sample sizes 60 and 100

Parameter		Variance 1000		Variance 100		Variance 10	
		T=60	T=100	T=60	T=100	T=60	T=100
$\gamma(3.0)$	Mean	2.9926	2.9942	2.9924	2.9940	2.9895	2.9925
	ABS Bias	0.0074	0.0058	0.0077	0.0060	0.0105	0.0075
	MSE	0.0081	0.0046	0.0081	0.0046	0.0082	0.0046
$\beta_{11}(1.0)$	Mean	1.0130	1.0117	1.0129	1.0116	1.0111	1.0106
	ABS Bias	0.0130	0.0117	0.0129	0.0116	0.0111	0.0106
	MSE	0.0463	0.0286	0.0462	0.0285	0.0455	0.0283
$\beta_{21}(2.0)$	Mean	1.9921	1.9997	1.9899	1.9985	1.9681	1.9871
	ABS Bias	0.0079	0.0003	0.0101	0.0015	0.0319	0.0129
	MSE	0.0967	0.0541	0.0966	0.0540	0.0958	0.0536
$\beta_{22}(0.5)$	Mean	0.4882	0.4926	0.4875	0.4925	0.4833	0.4920
	ABS Bias	0.0118	0.0074	0.0125	0.0075	0.0167	0.0080
	MSE	0.0705	0.0329	0.0703	0.0329	0.0694	0.0327
$\beta_{23}(1.5)$	Mean	1.4676	1.4767	1.4664	1.4760	1.4549	1.4700
	ABS Bias	0.0324	0.0233	0.0336	0.0240	0.0451	0.0300
	MSE	0.0646	0.0446	0.0645	0.0446	0.0645	0.0444

References

[1] Thomas Bayes, *An Essay Towards Solving a Problem in the Doctrine of Chances,* (1783) In: Berger, James O. 2000.Bayesian Analysis: A look at Today and Thoughts of Tomorrow Journal of the American Statistical

Association, **95**(452), (2000), 1269-1276. Accessed on 03/02/2013 at http://www.jstor.org/stable/2669768

[2] James O. Berger, Bayesian Analysis: A look at Today and Thoughts of Tomorrow, *Journal of the American Statistical Association*, **95**(452), (2000), 1269-1276. Accessed on 03/02/2013 at http://www.jstor.org/stable/2669768

[3] J.C. Chao and P.C.B. Phillips, Posterior distributions in limited information analysis of the simultaneous equations model using the Jeffreys prior, *Journal of Econometrics*, JSTOR, **87**, (1998), 49-86.

[4] M.H.Chen, Q.M. Shao and J.G. Ibrahim, *Monte Carlo Methods in Bayesian Computation*, New York, Springer-Verlag, 2000.

[5] Seymour Geisser, A Bayes Approach for Combining Correlated Estimates, *Journal of the American Statistical Association*, **60**(310), (1965), 602- 607.

[6] A. German, A. Jakulin, M.G. Pittau and Su, Yu-Sung, A weakly informative default Prior Distribution for Logistic and other Regression Models, *Annals of Applied Statistics*, **2**(4), (2008), 1360-1383.

[7] W.R. Gilks, S. Richardson and D.J. Spiegelhalter, *Markov Chain Monte Carlo in Practice*, Chapman & Hall/CRC, 1996.

[8] H. Jeffreys, *Theory of Probability*. 3rd edition, London: Oxford University Press, 1961.

[9] T. Lancaster, *An Introduction to modern Bayesian Econometrics*, Blackwell Publishing Ltd, 2004.

[10] P.S. Laplace, *Théorie Analytique des Probabilités.* (1812), In: Berger, James O. Bayesian Analysis: A look at Today and Thoughts of Tomorrow, *Journal of the American Statistical Association*, **95**(452), (2000), 1269-1276. Accessed on 03/02/2013 at http://www.jstor.org/stable/2669768

[11] A. Zellner, *An Introduction to Bayesian Inference in Econometrics*, John Wiley & sons, Inc. 1971.