Bayesian Analysis (2009) **4**, Number 2, pp. 209–212

Comments on Article by Yin

Ming-Hui Chen* and Sungduk Kim†

We would like to congratulate the author for a nice development of the Bayesian Generalized Method of Moments (BGMM). BGMM is a natural extension of the classical GMM. On the one hand, BGMM enjoys asymptotic properties and estimation efficiency of GMM; on the other hand, BGMM has a better computational property due to the recent advance in Markov chain Monte Carlo (MCMC) sampling. Therefore, BGMM is potentially very useful when the parameter estimation is of primary interest especially in statistical analysis of correlated, longitudinal or repeated measurement data.

The BGMM is primarily based on the moment conditions, instead of the likelihood. Thus, the "likelihood" used in constructing the "posterior distribution" in the BGMM is not the usual model-based likelihood function. This may be advantageous when the true likelihood is difficult to derive. However, in the BGMM framework, formal Bayesian model comparisons cannot be carried out as the likelihood function or the predictive distribution is not defined. Since the construction of the BGMM is primarily based on the marginal distribution model, the success of the BGMM in estimating the regression coefficients for the correlated data heavily relies on an adequate specification of the moment conditions. An immediate practical question is: how many moment conditions or what moment conditions need to be specified in order to capture the true correlation matrix? We suspect that when the moment conditions are not correctly specified, the standard deviations of Bayesian estimators based on the BGMM can be over-stated or under-stated especially when the sample size is relatively small. As the models cannot be compared via a usual Bayesian model comparison criterion such as the Bayes factor or the Deviance Information Criterion (Spiegelhalter et al. (2002)) and the true correlation structure is unknown in the BGMM, it becomes quite challenging and difficult to know how many $\mathbf{C}_{(j)}$'s are needed in order to achieve reliable standard deviations of the Bayesian estimators. Although the author has proposed several possible choices of $\mathbf{C}_{(j)}$'s, this issue has not been fully addressed. A BGMM estimator is asymptotically unbiased. However, the BGMM may fail to accurately estimate the certainty of a Bayesian estimator, which may be a major concern for using the BGMM.

To gain a better understanding of the BGMM and to further examine the performance of this method, we have conducted three simulation studies. In all simulations, we consider the similar regression model used in Section 3.3 with $K = 4$ and two covariates (Z_{1ik}, Z_{2ik}), i.e.,

$$Y_{ik} = \beta_0 + \beta_1 Z_{1ik} + \beta_2 Z_{2ik} + \epsilon_{ik}.$$

The covariate distributions for (Z_{1ik}, Z_{2ik}) are the same as those given in Section 3.3. That is, $Z_{1ik} \sim N(0, 1)$ and $Z_{2ik} \sim$ Bernoulli(0.5). The true parameter values are $\beta_0 = 0.2$, $\beta_1 = 0.5$ and $\beta_2 = -0.5$. Also, 500 data sets of sample size $n = 50$ were

*Department of Statistics, University of Connecticut, Storrs, CT, mailto:mhchen@stat.uconn.edu
†Division of Epidemiology, Statistics and Prevention Research, Eunice Kennedy Shriver National Institute of Child Health and Human Development, Rockville, MD, mailto:kims2@mail.nih.gov

 DOI:10.1214/09-BA407A

generated in all simulations.

Simulation I. In this simulation, we basically repeated the simulation conducted in Section 3.3. Instead of $n = 100$, we used $n = 50$. The true correlation matrix is $\mathbb{R} = (1-\rho)\mathbf{I}+\rho\mathbf{1}\mathbf{1}^T$ with $\rho = 0.5$. For the full Bayesian approach, the priors are specified as follows: $\rho \sim U(-1, 1)$, $\sigma^2 \sim IG(0.0001, 0.0001)$, and $\boldsymbol{\beta} = (\beta_0, \beta_1, \beta_2)^T \sim N(0, 1000\mathbf{I})$. The results are shown in Table 1. In addition to "Ave", "ESD", and "ASD" reported in the paper, we also reported "MSE", which was obtained by averaging the squares of the differences between the parameter estimates and the true parameter over 500 replicated data sets. From Table 1, we see that BGMM works well with a single moment condition specified by $\mathbf{C}_{(1)} = \mathbf{I}$ even when the true correlation matrix $\mathbb{R}$ is not an identity matrix. The MSE's are comparable for all three methods.

Method	Parameter	True	Ave	ESD	ASD	MSE
BGMM (**I**)	β_0	0.2	0.1958	0.1064	0.1196	0.0113
	β_1	0.5	0.5044	0.0716	0.0784	0.0051
	β_2	-0.5	-0.5026	0.0718	0.0759	0.0052
BGMM (**I** + Exch)	β_0	0.2	0.1932	0.1138	0.1185	0.0130
	β_1	0.5	0.5007	0.0564	0.0618	0.0032
	β_2	-0.5	-0.5040	0.0590	0.0595	0.0035
Full Bayesian	β_0	0.2	0.1954	0.1058	0.1024	0.0112
	β_1	0.5	0.5011	0.0554	0.0667	0.0031
	β_2	-0.5	-0.5023	0.0592	0.0658	0.0035

Table 1: Comparison between the BGMM and full Bayesian approach with the data generated from the model with an exchangeable correlation matrix.

Simulation II. Instead of an exchangeable correlation matrix for $\mathbb{R}$ in Simulation I, we consider

$$\mathbb{R} = \begin{pmatrix} 1.0 & 0.5 & 0.3 & 0.9 \\ 0.5 & 1.0 & -0.5 & 0.7 \\ 0.3 & -0.5 & 1.0 & -0.1 \\ 0.9 & 0.7 & -0.1 & 1.0 \end{pmatrix}$$

and the variances are 0.2, 9, 15, and 5 for $Y_{i1}, \ldots, Y_{i4}$, respectively, in this simulation. Let $\boldsymbol{\Sigma} = \text{Var}(\mathbf{Y}_i)$, where $\mathbf{Y}_i = (Y_{i1}, \ldots, Y_{i4})^T$. For the full Bayesian approach, the priors are specified as follows: $\boldsymbol{\Sigma} \sim \text{Inv-Wishart}(\nu_0, \Lambda_0)$ with $\nu_0 = 1$ and $\Lambda_0 = \text{diag}(0.0001, 0.0001, 0.0001, 0.0001)$, and $\boldsymbol{\beta} = (\beta_0, \beta_1, \beta_2)^T \sim N(\mathbf{0}, 1000\mathbf{I})$. The results are reported in Table 2. Unlike Simulation I, when the true correlation matrix is not exchangeable, the standard deviations based on BGMM (**I**) or BGMM (**I** + Exch) are much larger than those obtained from the full Bayesian approach. In this case, adding $\mathbf{C}_{(2)}$ (Exch) does not seem to improve the standard deviations at all. In addition, we also tried BGMM ($\mathbf{I}$ + Exch + $\mathbf{C}_{(3)}$), where $\mathbf{C}_{(3)} = \begin{pmatrix} 0 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{pmatrix}$. The resulting ESD's are 0.2012, 0.1926, and 0.1868 for β_0, β_1, and β_2, respectively. These ESD's are still much larger than those obtained from the full Bayesian approach. Thus, more moment conditions are needed for this case.

Method	Parameter	True	Ave	ESD	ASD	MSE
BGMM (**I**)	β_0	0.2	0.1917	0.1882	0.2029	0.0354
	β_1	0.5	0.5066	0.1902	0.2036	0.0361
	β_2	-0.5	-0.5061	0.2039	0.2015	0.0415
BGMM (**I** + Exch)	β_0	0.2	0.1862	0.2007	0.2015	0.0404
	β_1	0.5	0.5031	0.2013	0.2010	0.0405
	β_2	-0.5	-0.5039	0.2088	0.2006	0.0435
Full Bayesian	β_0	0.2	0.1998	0.0153	0.0148	0.0002
	β_1	0.5	0.5003	0.0115	0.0110	0.0001
	β_2	-0.5	-0.5008	0.0121	0.0110	0.0002

Table 2: Comparison between the BGMM and full Bayesian approach with the data generated from the model with a general correlation matrix.

Simulation III. In the first two simulations, the true correlation matrix is the same across all 50 observations. In this simulation, we consider an usual special case in which the data were generated from a Brownian motion process model. Specifically, for observation i, we generated t_{i1}, t_{i2}, t_{i3}, and t_{i4} independently from $U(0, 100)$. For notional convenience, we assume $t_{i1} < t_{i2} < t_{i3} < t_{i4}$. Then, the true covariance matrix of $\mathbf{Y}_i$ was specified as

$$\boldsymbol{\Sigma}_i = \sigma^2 \begin{pmatrix} t_{i1} & t_{i1} & t_{i1} & t_{i1} \\ t_{i1} & t_{i2} & t_{i2} & t_{i2} \\ t_{i1} & t_{i2} & t_{i3} & t_{i3} \\ t_{i1} & t_{i2} & t_{i3} & t_{i4} \end{pmatrix}$$

with $\sigma^2 = 1$. Under this setting, all true correlation matrices are different across different observations. We suspect that it may be more difficult for the BGMM to capture such observation-varying correlation matrices. Similar to Simulation I, for the full Bayesian approach, we specify the priors as follows: $\sigma^2 \sim IG(0.0001, 0.0001)$ and $\boldsymbol{\beta} = (\beta_0, \beta_1, \beta_2)^T \sim N(\mathbf{0}, 1000\mathbf{I})$. Table 3 shows the results. From this table, we see that the BGMM performs much worse than the full Bayesian approach. However, we can see an improvement of BGMM (**I** + Exch) over BGMM (**I**) in terms of ESD, ASD and MSE except for β_0.

In all three simulations, we have also tried $n = 100$. Except for Simulation I, the ESD's and ASD's obtained from the BGMM remain much larger than those obtained from the full Bayesian approach. Therefore, a much larger sample size n and more moment conditions may be needed for the linear regression models with the true correlation matrices given in Simulations II and III. Finally, we would like to mention that when the dimension of regression coefficients increases, we have experienced slow convergence of the Metropolis algorithm. We suspect that the Metropolis algorithm within the Gibbs sampler may not work well when the dimension of the parameters is relatively large

Method	Parameter	True	Ave	ESD	ASD	MSE
BGMM (**I**)	β_0	0.2	0.1567	0.7991	0.8302	0.6391
	β_1	0.5	0.5114	0.4833	0.4998	0.2332
	β_2	-0.5	-0.5278	0.4970	0.4865	0.2473
BGMM (**I** + Exch)	β_0	0.2	0.1515	0.8180	0.8133	0.6702
	β_1	0.5	0.4892	0.3311	0.3280	0.1095
	β_2	-0.5	-0.5283	0.3371	0.3134	0.1142
Full Bayesian	β_0	0.2	0.1851	0.2823	0.1918	0.0798
	β_1	0.5	0.5021	0.1232	0.0841	0.0152
	β_2	-0.5	-0.5045	0.1233	0.0829	0.0152

Table 3: Comparison between the BGMM and full Bayesian approach with the data generated from a Brownian motion process model.

due to the complexity of the posterior distribution in the BGMM. Some recent development on MCMC sampling such as Kou et al. (2006) and Liang et al. (2007) and general MCMC sampling strategies discussed in Chen et al. (2000) may all be useful in order to develop a more efficient MCMC sampling algorithm from the posterior distribution in the BGMM.

References

Chen, M.-H., Shao, Q.-M., and Ibrahim, J. G. (2000). *Monte Carlo Methods in Bayesian Computation*. New York: Springer-Verlag. 212

Kou, S. C., Zhou, Q., and Wong, W. H. (2006). "Equi-energy Sampler with Applications to Statistical Inference and Statistical Mechanics (with Discussion)." *Annals of Statistics*, 34: 1581–1619. 212

Liang, F., Liu, C., and Carroll, R. J. (2007). "Stochastic Approximation in Monte Carlo Computation." *Journal of the American Statistical Association*, 102: 305–320. 212

Spiegelhalter, D. J., Best, N. G., Carlin., B. P., and van der Linde, A. (2002). "Bayesian Measures of Model Complexity and Fit (with Discussion)." *Journal of the Royal Statistical Society, Series B*, 62: 583–639. 209