H.G.B. Alexander Research Foundation
Graduate School of Business
University of Chicago

Rejoinder

by

Arnold Zellner and Tomohiro Ando

Booth School of Business
University of Chicago
5807 South Woodlawn Avenue
Chicago, IL 60637

arnold.zellner@chicagoBooth.edu, andoh@hc.cc.keio.ac.jp

Paper 0904 **December 2009**

Rejoinder

Arnold Zellner* Tomohiro Ando

Booth School of Business, University of Chicago, 5807 South Woodlawn Avenue Chicago, Illinois, USA, 60637

Graduate School of Business Administration, Keio University 4-1-1 Hiyoshi, Kohoku-ku, Yokohama-shi, Kanagawa, 223-8526, Japan

---

Thanks to Editor Kajal Lahiri for giving us an opportunity to contribute to this special issue on Bayesian forecasting models and to respond to Geweke's comments.

It is of the greatest importance that forecasters have accurate and efficient computational procedures to implement their Bayesian forecasting models. That is a major objective of our current and recent papers in which we develop direct Monte Carlo (DMC) methods that have been shown to work well in experimental and applied problems for a variety of models; see references in our paper for these results.

Note that in our approach, we begin with an original model (OM) with its likelihood function and prior and posterior densities for its parameters. Since the posterior density is often difficult to analyze analytically or numerically, we introduce a transformed model (TM) that is a one to one transformation of the OM's observations and its parameters (see, e.g. eq. (3) of our paper) and yields a posterior density, based on the transformed prior of the OM, from which draws of the parameters can be made and converted to draws from the posterior density for the OM by use of the one to one relations connecting parameters of the TM and of the OM. Thus we obtain exact posterior densities, moments, intervals, etc. for parameters of the OM in this DMC approach. Since Geweke in his comment failed to transform back to obtain the posterior density for the parameters of the OM in his example, his results are not relevant results. As mentioned above, in previous experimental and applied problems,

[1] *Corresponding author
Tel.:+1-773-702-7145; fax:+1-773-702-0458.
E-mail address: arnold.zellner@ChicagoBooth.edu andoh@kbs.keio.ac.jp

including those in our present paper, our DMC approach works very well. See below for additional evidence.

Given limited space, we shall now just respond to Geweke's two major comments. His first comment is:

Comment 1: *The DMC method does not produce samples from the posterior distribution of the SUR model for any of the disturbance distributions considered in ZA, including multivariate normal. Current MCMC algorithms are quite adequate for these SUR models, implying that further gains from a hypothetical i.i.d. posterior simulator would be minimal in any event.*

Let us show that our DMC method for the normal SUR model can produce posterior samples correctly. To parallel Geweke's simple analysis, we consider an $m = 2$ equation SUR model. We first investigate the relationship between the original model (OM)

$$\boldsymbol{y}_j = X_j \boldsymbol{\beta}_j + \boldsymbol{u}_j, \quad \text{with} \quad E[\boldsymbol{u}_i \boldsymbol{u}_j'] = \begin{cases} \omega_{ij} I, & (i \neq j), \\ \omega_i^2 I, & (i = j). \end{cases} \tag{1}$$

$j = 1, 2$ and the transformed model (TM)

$$\begin{aligned} \boldsymbol{u}_1 &= \boldsymbol{y}_1 - X_1 \boldsymbol{\beta}_1 = \rho_{12} \boldsymbol{u}_2 + \boldsymbol{e}_1, \\ \boldsymbol{u}_2 &= \boldsymbol{y}_2 - X_2 \boldsymbol{\beta}_2 = \rho_{21} \boldsymbol{u}_1 + \boldsymbol{e}_2, \end{aligned}$$

with

$$E[\boldsymbol{e}_i \boldsymbol{e}_j'] = \begin{cases} O, & (i \neq j) \\ \sigma_i^2 I, & (i = j) \end{cases}, \quad \text{and} \quad \Sigma = \operatorname{diag}\{\sigma_1^2, \sigma_2^2\},$$

and thus,

$$\begin{cases} \boldsymbol{y}_1 = X_1 \boldsymbol{\beta}_1 + \rho_{12} \boldsymbol{u}_2 + \boldsymbol{e}_1, \\ \boldsymbol{y}_2 = X_2 \boldsymbol{\beta}_2 + \rho_{21} \boldsymbol{u}_1 + \boldsymbol{e}_2, \end{cases}$$

We use $\boldsymbol{y}_1 = X_1 \boldsymbol{\beta}_1 + \boldsymbol{u}_1$ as a starting point to get a draw of $\boldsymbol{u}_1$ that we put into the second equation as an "independent" variable along with X_2.

The model parameters in the OM (1) are $\{\boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \omega_1^2, \omega_2^2, \omega_{12} (= \omega_{21})\}$. The unknown parameters in the TM are $\{\boldsymbol{b}_1, \boldsymbol{b}_2, \sigma_1^2, \sigma_2^2\}$. Note that $\boldsymbol{\beta}_1 = \boldsymbol{b}_1$, $\omega_1^2 =$

σ_1^2 and $(\boldsymbol{\beta}_2', \rho_{21})' = \boldsymbol{b}_2'$. The relationship between ρ_{21} in $\boldsymbol{u}_2 = \rho_{21}\boldsymbol{u}_1 + \boldsymbol{e}_2$ and ω_{12} is $\rho_{21} = \omega_{12}/\omega_{11}^2$. The parameter ω_2^2 is expressed as $\omega_2^2 = \sigma_2^2 + \rho_{21}^2\sigma_1^2$.

To make an inference on Ω based on the generated samples $\{\sigma_1^{2(k)}, \sigma_2^{2(k)}, \rho_{21}{}^{(k)}; k = 1, ..., N\}$ from a direct Monte Carlo sampling procedure, we can use the following relations:

$$
\begin{aligned}
\omega_1^{2(k)} &= \sigma_1^{2(k)}, \\
\omega_2^{2(k)} &= \rho_{21}^{2\ (k)} \times \sigma_1^{2(k)} + \sigma_2^{2(k)}, \\
\omega_{12}{}^{(k)} &= \rho_{21}{}^{(k)} \times \sigma_1^{2(k)}.
\end{aligned}
$$

Therefore the posterior samples of $\Omega^{(k)}$, $k = 1, ..., N$ can be easily obtained. From $\boldsymbol{y}_1 = X_1\boldsymbol{\beta}_1 + \rho_{12}\boldsymbol{u}_2 + \boldsymbol{e}_1$, we obtain, $\boldsymbol{\beta}_1 = \hat{\boldsymbol{\beta}}_1 - \rho_{12}(X_1'X_1)^{-1}X_1'\boldsymbol{y}_2 - (X_1'X_1)^{-1}X_1'X_2\boldsymbol{\beta}_2 + (X_1'X_1)^{-1}X_1'\boldsymbol{e}_1$ with $\hat{\boldsymbol{\beta}}_1 = (X_1'X_1)^{-1}X_1'\boldsymbol{y}_1$. Using the draws $\{\boldsymbol{\beta}_2^{(k)}, \rho_{12}^{(k)}, \boldsymbol{e}_1^{(k)}\}$, we can obtain the draws of $\boldsymbol{\beta}_1^{(k)}$. This illustrates well the need to transform back to the parameters of the OM and shows explicitly how it can be done.

We point out that the TM can be also expressed as

$$
\begin{cases}
\boldsymbol{y}_1 = X_1\boldsymbol{b}_1 + \boldsymbol{e}_1, \\
\boldsymbol{y}_2 = X_2\boldsymbol{b}_2 + \boldsymbol{y}_1 b_3 + X_1\boldsymbol{b}_4 + \boldsymbol{e}_2,
\end{cases}
$$

where $\boldsymbol{b}_1 = \boldsymbol{\beta}_1$, $\boldsymbol{b}_2 = \boldsymbol{\beta}_2$, $b_3 = \rho$ and $\boldsymbol{b}_4 = -\rho\boldsymbol{\beta}_1$. Then from the definition of the $\boldsymbol{b}$'s, we have $\boldsymbol{b}_1 + \boldsymbol{b}_4 = \boldsymbol{\beta}_1 - \rho\boldsymbol{\beta}_1$. and thus obtain the relation: $\boldsymbol{\beta}_1 = (\boldsymbol{b}_1 + \boldsymbol{b}_4)/(1-\rho)$. Thus since draws $\boldsymbol{b}_1$ depend on $\boldsymbol{y}_1$, etc., and draws of $\boldsymbol{b}_4$ depend on $\boldsymbol{y}_2$, etc., the $\boldsymbol{\beta}_1$ values from the above relation depend on both $\boldsymbol{y}_1$ and $\boldsymbol{y}_2$. Thus we are not just relying on $\boldsymbol{y}_1 = X_1\boldsymbol{b}_1 + \boldsymbol{e}_1$ to obtain draws of $\boldsymbol{b}_1 = \boldsymbol{\beta}_1$.

We emphasize that Geweke failed to transform back from the posterior density for the parameters of the TM to the posterior density for the parameters of the OM. We explicitly pointed out that $\boldsymbol{\beta}_1$ appeared in both equations of the TM in our paper and never claimed that the first equation provided the complete marginal posterior for $\boldsymbol{\beta}_1$. This is Geweke's "incorrect" conclusion.

As shown in the analysis above, all we required is a one-to-one transformation from which parameter draws can be made and transformed back in a one to one fashion to the parameters of the original model's posterior density's parameters.

In addition, when we developed the DMC method, we also transformed the original Jeffreys' prior $\pi_1(\boldsymbol{\beta}, \Omega)$ from the space $\{\boldsymbol{\beta}, \Omega\}$ to $\{\boldsymbol{b}, \Sigma\}$. When $m = 2$, it is

$$\pi_1(\boldsymbol{b}, \Sigma) \propto |\Omega(\boldsymbol{b}, \Sigma)|^{-3/2} \times |J|_2 = (\sigma_1^2\sigma_2^2)^{-3/2} \times \sigma_1^2 = \frac{1}{\sigma_1} \times \frac{1}{\sigma_2^3},$$

where the Jacobian of the transformation from $\{\boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \omega_{11}^2, \omega_{12}^2, \omega_{12}\}$ to $\{\boldsymbol{b}_1, \boldsymbol{b}_2, \sigma_1^2, \sigma_2^2\}$ is given by,

$$|J|_2 = \begin{vmatrix} \frac{\partial \boldsymbol{\beta}_1'}{\partial \boldsymbol{b}_1} & \frac{\partial \boldsymbol{\beta}_2'}{\partial \boldsymbol{b}_1} & \frac{\partial \omega_{12}}{\partial \boldsymbol{b}_1} & \frac{\partial \omega_1^2}{\partial \boldsymbol{b}_1} & \frac{\partial \omega_2^2}{\partial \boldsymbol{b}_1} \\ \frac{\partial \boldsymbol{\beta}_1'}{\partial \boldsymbol{b}_2} & \frac{\partial \boldsymbol{\beta}_2'}{\partial \boldsymbol{b}_2} & \frac{\partial \omega_{12}}{\partial \boldsymbol{b}_2} & \frac{\partial \omega_1^2}{\partial \boldsymbol{b}_2} & \frac{\partial \omega_2^2}{\partial \boldsymbol{b}_2} \\ \frac{\partial \boldsymbol{\beta}_1'}{\partial \sigma_1^2} & \frac{\partial \boldsymbol{\beta}_2'}{\partial \sigma_1^2} & \frac{\partial \omega_{12}}{\partial \sigma_1^2} & \frac{\partial \omega_1^2}{\partial \sigma_1^2} & \frac{\partial \omega_2^2}{\partial \sigma_1^2} \\ \frac{\partial \boldsymbol{\beta}_1'}{\partial \sigma_2^2} & \frac{\partial \boldsymbol{\beta}_2'}{\partial \sigma_2^2} & \frac{\partial \omega_{12}}{\partial \sigma_2^2} & \frac{\partial \omega_1^2}{\partial \sigma_2^2} & \frac{\partial \omega_2^2}{\partial \sigma_2^2} \end{vmatrix} = \begin{vmatrix} I & O & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ O & \begin{pmatrix} I \\ \mathbf{0} \end{pmatrix} & \begin{pmatrix} \mathbf{0} \\ \sigma_1^2 \end{pmatrix} & \mathbf{0} & \mathbf{0} \\ \mathbf{0}' & \mathbf{0}' & \rho_{21} & 1 & \rho_{21}^2 \\ \mathbf{0}' & \mathbf{0}' & 0 & 0 & 1 \end{vmatrix} = \sigma_1^2.$$

This analysis also indicates that our DMC method is invariant to the ordering of the equations.

Next, as regards his comments on the SUR model with Student-t errors, as shown by the form of the likelihood function $L(\boldsymbol{y}|\boldsymbol{b}, \Sigma, \nu)$ after equation (7), the likelihood function contains information about the degrees of freedom parameter ν. This is obvious evidence that our formulation is NOT equivalent to the normal model. Noting that the posterior distribution of the parameters is proportional to the likelihood times the prior, the posterior distribution of the degrees of freedom parameter ν is "MORE" informative than the prior distribution, and it will become more informative as $n \to \infty$. Therefore, the proposed method does address the Bayesian analysis of the SUR model with Student-t errors.

One of the motivations of our research is that the direct Monte Carlo approach is computationally much more efficient than the MCMC approach. In his comments, Geweke stated *"While generic criticisms of MCMC methods could be relevant for some applications of MCMC, they do not pertain to the SUR model. The reason is that $\boldsymbol{\beta}$ and Ω are nearly independent in the posterior distribution, implying that Percy's Gibbs sampler produces draws that are nearly i.i.d."* Unfortunately, this comment is not correct .

To make this point clear, we shall provide a simple example that violates his conclusion by simulating a set of $n = 50$ observations from the $m = 2$

dimensional normal SUR model. The number of predictors for each of the equations is to be $p_j = 40$, $j = 1, 2$. This model can thus be written as follows:

$$\begin{pmatrix} \boldsymbol{y}_1 \\ \boldsymbol{y}_2 \end{pmatrix} = \begin{pmatrix} X_1 & O \\ O & X_2 \end{pmatrix} \begin{pmatrix} \boldsymbol{\beta}_1 \\ \boldsymbol{\beta}_2 \end{pmatrix} + \begin{pmatrix} \boldsymbol{u}_1 \\ \boldsymbol{u}_2 \end{pmatrix}, \quad \Omega = \begin{pmatrix} 0.4 & -0.2 \\ -0.2 & 0.8 \end{pmatrix},$$

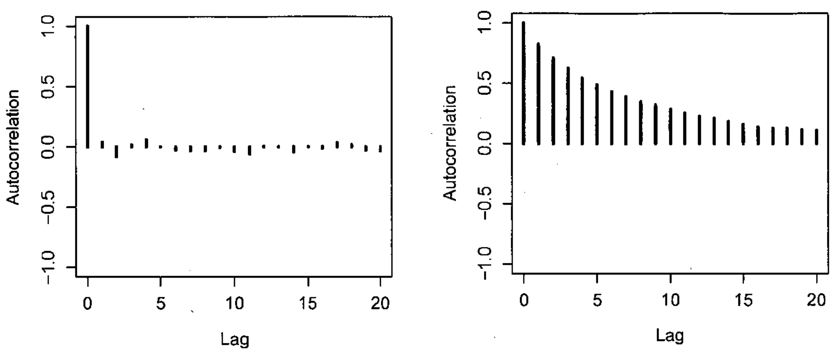
where $\boldsymbol{y}_j$ and $\boldsymbol{u}_j$ are $n \times 1$ vectors, X_j is the $n \times p_j$ matrix and $\boldsymbol{\beta}_j$ is the p_j-dimensional vector. The covariate matrices X_j $j = 1, 2$ were generated from a uniform density over the interval $(-1, 1)$ The coefficient vectors $\boldsymbol{\beta}_1$ and $\boldsymbol{\beta}_2$ were randomly set to between the range -4 to 4, by using uniform random variables.

We generated a set of 1,000 posterior samples. In Percy's Gibbs sampler, the first 1,000 samples are discarded as a burn-in period. As shown in Figure 1, there is high autocorrelation in the generated values of the covariance parameter ω_{12}. We see that the autocorrelations at the first several lags are very large. Even for the higher order autocorrelations, the autocorrelation function is not close to zero. Therefore, in practical use of the posterior samples from MCMC output, researchers must take account of the autocorrelations, as is obvious. One of the most popular approaches in use is to employ every k'th sample and to discard all remaining samples. The number k is usually determined by considering the magnitudes of the estimated autocorrelation coefficients, with large autocorrelations leading to use of large values of k. An ideal situation would be one in which all autocorrelations are equal to zero and $k = 1$, a situation in which no samples are discarded, as in our DMC approach where the autocorrelations are all theoretically equal to zero (see Fig. 1 for an estimated autocorrelation function using the output of our DMC procedure).

Further, we shall provide another problem with the MCMC procedure below in connection with Geweke's second comment:

Comment 2: *In the model proposed by Lange et al. (1989) the m disturbances at each of the n sample points are i.i.d. with m-variate multivariate Student-t distributions with common denominator. This model readily produces excess kurtosis in disturbances, and Bayesian inference can be conducted reliably and efficiently using a very modest extension of the MCMC algorithm of Geweke (1993).*

We have to point out that there are many ways to make posterior inference about the SUR model with Student-t errors. Combining the MCMC algorithm of Geweke (1993) and the Gibbs algorithm of Percy (1992), Geweke proposed



Our DMC method. Percy (1992).

Fig. 1. Autocorrelation function of successive draws of the covariance parameter ω_{12} from the output of our DMC method and from that of Percy (1992). The vaules of the estimated autocorrelations in Figure 1 are much smaller for our DMC method than the corresponding estimated autocorrelations associated with the MCMC method of Percy (1992).

a new MCMC algorithm for computing Bayesian results for the SUR model with Student-t errors. However, this algorithm is not computationally efficient as we have seen in the previous example that use of the Gibbs algorithm of Percy (1992) led to very highly auto-correlated output draws.

We have to point out that while much work has been done on procedures for checking the convergence of MCMC procedures to make sure not only that they have converged but also converged to the correct value, this is still a very difficult and important area of research. Indeed in a paper to appear in the J. of Econometrics, S. Chib and S. Ramamurthy, "Multiple-Block MCMC Methods for Analysis of DSGE Models," show that MCMC results reported in a recent article published in the American Economic Review are incorrect because the MCMC algorithm employed converged to the wrong answer, apparently due to a failure to integrate over the complete parameter space and to employ appropriate convergence checks. It is our hope that use of our DMC approach will help to avoid such errors and provide accurate and efficient procedures for forecasters and others.

Lastly, we would like to point out that our method can be used for on-line forecasting. In the two sample case, we use the first sample to analyze the first model and its parameters, $\boldsymbol{y}_1 = X_1\boldsymbol{\beta}_1 + \boldsymbol{u}_1$ in the first period. Then in the second period we receive data for the second model, $\boldsymbol{y}_2 = X_2\boldsymbol{\beta}_2 + \boldsymbol{u}_2$, with elements of $\boldsymbol{u}_1$ and $\boldsymbol{u}_2$ correlated and then jointly analyze the two models and data sets JOINTLY. In the paper, we did not make the sequential on line nature of our problem explicit. However, the sequential or "on line" version of the problem is very relevant to what practical forecasters actually do and our DMC algorithm will provide them with an exact procedure for computing sequentially updated optimal estimates for the parameters of the regression equations period by period. Hopefully, we think that our DMC approach will provide a nice online forecasting tool.