A Direct Monte Carlo Approach for Bayesian Analysis of the Seemingly Unrelated Regression Model

Arnold Zellner

Tomohiro Ando

(Chicago GSB)

Theodore W. Anderson's 90th Birthday Conference, Stanford U.

Outline

- Background, Motivation
- Bayesian Inference for the Seemingly Unrelated Regression Model Using a Direct Monte Carlo Procedure
- Numerical results
- Conclusion and future work

Outline

- Background, Motivation
- Bayesian Inference for the Seemingly Unrelated Regression Model Using a Direct Monte Carlo Procedure
- Numerical results
- Conclusion and future works

Background and motivation

SUR model: Overview

- Introduced by Zellner (1962)

- Many studies have contributed to the analysis of SUR models.

- Many textbooks and journal papers
  - Zellner (1963),Gallant (1975), Rocke (1989), Neudecker and Windmeijer (1991), Mandy and Martins (1993), Kurata (1999), Liu (2002), Ng (2002) and references therein.
  - Zellner (1971), Judge et al. (1988), Greene (2002), Geweke (2005), Lancaster (2004), Rossi, McCulloch and Allenby (2005) and so on.

SUR model: Overview Cont.

Generalized least squares

- Zellner (1962, 1963), GLS, Iterative GLS, finite sample
- Madansky (1964), Iterative GLS and ML

Bayesian Methods

Zellner (1971), et al. , approx. finite sample posteriors, Stein shrinkage, etc.

van der Merwve and Viljoen, (1998), Bayesian MOM

Bayesian Markov Chain Monte Carlo estimation

- Percy (1992, 1996), Chib and Greenberg (1995), Smith and Kohn (2000) and Rossi et al. (2005))

SUR model : Motivation

- MCMC methods are rather complicated and involve many decisions

- Initial parameter value
- Choice of an appropriate proposal density
- The length of the burn in period
- Check for convergence

SUR model : Motivation

- Several drawbacks of MCMC

- Can we estimate the Bayesian SUR model by using a direct Monte Carlo (DMC) procedure?

Comparison of DMC and MCMC

User friendly

Items	DMC	MCMC (Gibbs)
Need to fix the number of samples drawn	Yes	Yes
100% acceptance of draws	Yes	No (Yes)
Require initial parameters value	No	Yes
Burn-in period setting	No	Yes
Check for convergence	No	Yes
Select convergence check criteria	No	Yes
Selection of a proposal density	No	Yes (No)
Use of a proposal density	No	Yes (No)

Outline

- Background, Motivation
- Bayesian Inference for the Seemingly Unrelated Regression Model Using a Direct Monte Carlo Procedure
- Numerical results
- Conclusion and future works

The standard SUR models

The standard SUR models (1/2)

- A set of m equations (E.g.,Zellner (1971))

$$\underset{(n\times 1)}{\mathbf{y}_j} = \underset{(n\times p_j)}{X_j}\underset{(p_j\times 1)}{\boldsymbol{\beta}_j} + \underset{(n\times 1)}{\mathbf{u}_j}, \quad j = 1,\ldots,m,$$

$$E\left[\mathbf{u}_i\mathbf{u}_j{}'\right] = \begin{cases} \omega_{ij} I \ (i \neq j) \\ \omega_i^2 I \ (i = j) \end{cases}.$$

The standard SUR models (1/2)

- The likelihood function

$$L(D \mid \boldsymbol{\beta}, \Omega) = \frac{1}{(2\pi)^{nm/2} \mid \Omega \mid^{n/2}} \exp\left[-\frac{1}{2} \left(\mathbf{y} - X\boldsymbol{\beta} \right)'(\Omega^{-1} \otimes I) \left(\mathbf{y} - X\boldsymbol{\beta} \right) \right]$$

$$= \frac{1}{(2\pi)^{nm/2} \mid \Omega \mid^{n/2}} \exp\left[-\frac{1}{2} \mathrm{tr} \left\{ R\Omega^{-1} \right\} \right].$$

$$\mathbf{y}' = (\mathbf{y}_1', \ldots, \mathbf{y}_m') \quad X = \mathrm{diag}\{X_1, \ldots, X_m\}$$

$$R = (r_{ij}); \quad r_{ij} = (\mathbf{y}_i - X_i\boldsymbol{\beta}_i)'(\mathbf{y}_j - X_j\boldsymbol{\beta}_j)$$

$$\boldsymbol{\beta}' = (\boldsymbol{\beta}_1', \ldots, \boldsymbol{\beta}_m')$$

Bayesian estimation

$$\{\boldsymbol{\beta}, \boldsymbol{\Omega}\}$$

Bayesian approach (1/3)

- Diffuse prior

$$\pi_1(\boldsymbol{\beta}, \Omega) = \pi_1(\boldsymbol{\beta})\pi_1(\Omega) \propto |\,\Omega\,|^{-\frac{m+1}{2}}$$

- Joint posterior distribution

$$g_1(\boldsymbol{\beta}, \Omega \mid D) \propto |\,\Omega\,|^{-(n+m+1)/2} \exp\left[-\frac{1}{2}\mathrm{tr}\left\{R\Omega^{-1}\right\}\right]$$

Bayesian approach (2/3)

- Conditional posteriors

$$\begin{cases} g_1(\boldsymbol{\beta} \mid \Omega, D) = N\left(\hat{\boldsymbol{\beta}}, \hat{\Omega}_{\beta}\right) \\ g_1(\Omega \mid \boldsymbol{\beta}, D) = IW(R, n) \end{cases}$$

with $\hat{\boldsymbol{\beta}} = \{X'(\Omega^{-1} \otimes I)X\}^{-1} X'(\Omega^{-1} \otimes I)\mathbf{y}$

$$\hat{\Omega}_{\beta} = (X'(\Omega^{-1} \otimes I)X)^{-1}$$

Bayesian approach (3/3)

- Normal/inverse Wishart priors

$$\pi_2(\boldsymbol{\beta}, \Omega) = \pi_2(\boldsymbol{\beta})\pi_2(\Omega)$$

$$\pi_2(\boldsymbol{\beta}) = N(\boldsymbol{\beta}_0, A_\beta^{-1}), \quad \pi_2(\Omega) = IW(\Lambda_0, \nu_0)$$

- Conditional posteriors

$$g_2(\boldsymbol{\beta} \mid \Omega, D) = N(\overline{\boldsymbol{\beta}}, \overline{\Omega}_\beta)$$

$$g_2(\Omega \mid \boldsymbol{\beta}, D) = IW(\Lambda_0 + R, n + \nu_0),$$

The DMC procedure for the transformed SUR model

Transformation(1/2)

- The original SUR model

$$\mathbf{y}_j = X_j \boldsymbol{\beta}_j + \mathbf{u}_j \quad j = 1, \dots, m$$

- Transformation

$$\begin{cases} \mathbf{u}_1 = \mathbf{e}_1 \\ \mathbf{u}_2 = \rho_{21} \mathbf{u}_1 + \mathbf{e}_2 \\ \quad \vdots \\ \mathbf{u}_m = \sum_{j=1}^{m-1} \rho_{mj} \mathbf{u}_j + \mathbf{e}_m \end{cases}$$

$$E\left[\mathbf{e}_i \mathbf{e}_j'\right] = \begin{cases} O \ (i \neq j) \\ \sigma_i^2 I \ (i = j) \end{cases}$$

Transformation(2/2)

- The transformed model

$$\begin{cases} \mathbf{y}_1 = X_1\boldsymbol{\beta}_1 + \mathbf{e}_1 \equiv Z_1\mathbf{b}_1 + \mathbf{e}_1 \\ \mathbf{y}_j = X_j\boldsymbol{\beta}_j + \sum_{l=1}^{j-1} \rho_{jl}(\mathbf{y}_l - X_l\boldsymbol{\beta}_l) + \mathbf{e}_j \equiv Z_j\mathbf{b}_j + \mathbf{e}_j, \quad j = 2, \ldots, m, \end{cases}$$

- The likelihood function

$$L(D \mid \mathbf{b}, \Sigma) = \prod_{j=1}^{m} \frac{1}{(2\pi\sigma_j^2)^{n/2}} \exp\left[-\frac{(\mathbf{y}_j - Z_j\mathbf{b}_j)'(\mathbf{y}_j - Z_j\mathbf{b}_j)}{2\sigma_j^2} \right].$$

Bayesian estimation

$$\{\mathbf{b}, \Sigma\}$$

Bayesian analysis of the transformed model (1/3)

- Diffuse prior

$$\pi_3(\mathbf{b}, \Sigma) = \pi_3(\mathbf{b})\pi_3(\Sigma) \propto \prod_{j=1}^{m} (\sigma_j)^{-1}.$$

- Joint posteriors

$$g(\mathbf{b}, \Sigma \mid D) \propto \prod_{j=1}^{m} (\sigma_j)^{-(n+1)} \exp\left[-\frac{(\mathbf{y}_j - Z_j \mathbf{b}_j)'(\mathbf{y}_j - Z_j \mathbf{b}_j)}{2\sigma_j^2} \right].$$

Bayesian analysis of the transformed model (2/3)

- The conditional posteriors

Normal

$$g(\mathbf{b}_j \mid \mathbf{b}_{j-1},\ldots,\mathbf{b}_1,\sigma_j^2,D) = N\left(\hat{\mathbf{b}}_j, \sigma_j^2\left(Z_j{}'Z_j\right)^{-1}\right),$$

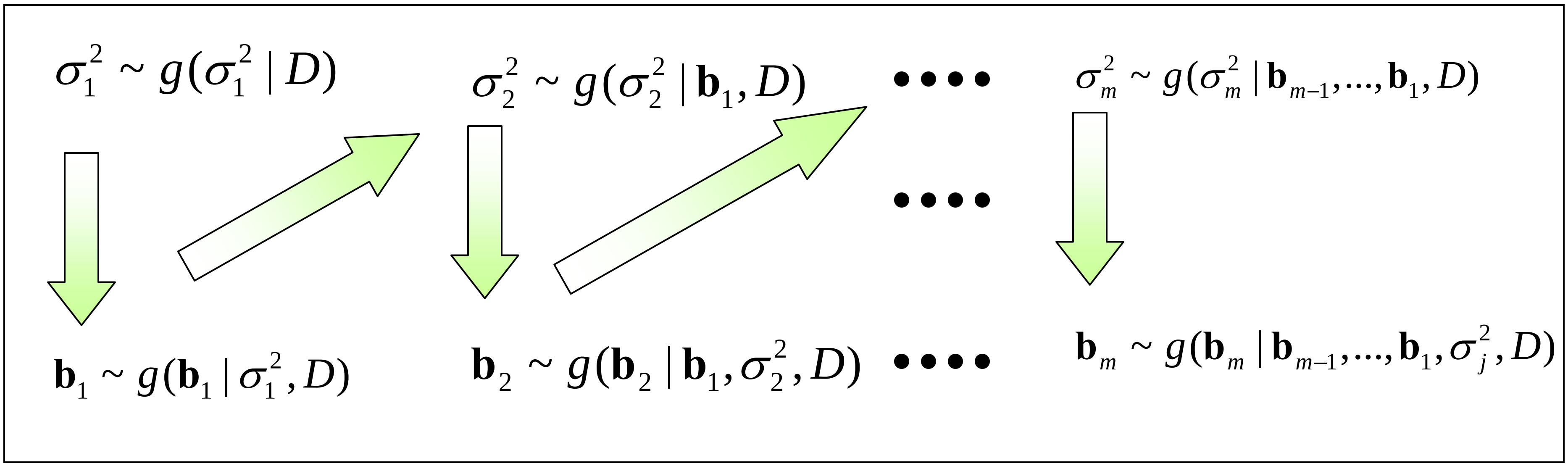
Inverse gamma

$$g(\sigma_j^2 \mid \mathbf{b}_{j-1},\ldots,\mathbf{b}_1,D) = IG\left(\frac{1}{2}\left(\mathbf{y}_j - Z_j\hat{\mathbf{b}}_j\right)'\left(\mathbf{y}_j - Z_j\hat{\mathbf{b}}_j\right), \frac{1}{2}\left(n - p_j - j + 1\right)\right),$$

with $\hat{\mathbf{b}}_j = \left(Z_j{}'Z_j\right)^{-1} Z_j{}'\mathbf{y}_j.$

Bayesian analysis of the transformed model (3/3)

Graphical presentation of the conditional posteriors



A direct Monte Carlo (DMC) sampling procedure:

1: Fix the order of a set of m equations. Set the number of samples N to be generated. Set $j = 1$

2 Generate ${\sigma_1^2}^{(k)}$, $k = 1, \ldots, N$, and insert the drawn values in $\pi(\mathbf{b}_1 \mid \sigma_1^2, D)$. Then make a draw $\mathbf{b}_1^{(k)}$, from $\pi(\mathbf{b}_1 \mid {\sigma_1^2}^{(k)}, D)$, for $k = 1, \ldots, N$.

3 Increase the index $j+1 \leftarrow j$. Draw ${\sigma_j}^{(k)}$ from the conditional inverse gamma density $g(\sigma_j^2 \mid \mathbf{b}_{j-1}^{(k)}, \ldots, \mathbf{b}_1^{(k)}, D)$, and then generate $\mathbf{b}_j^{(k)}$ from $g(\mathbf{b}_j \mid \mathbf{b}_{j-1}^{(k)}, \ldots, \mathbf{b}_1^{(k)}, {\sigma_j}^{(k)}, D)$ for $k = 1, \ldots, N$

4 Repeat Step 3 sequentially until j=m.

Some remarks (1/3)

- Only set the number of sampling
- Improper prior… Bayes factor?
  - BPIC (Ando, 2007)
- Informative prior?
- Inference on the original SUR model?

$$\{\boldsymbol{\beta}, \boldsymbol{\Omega}\} \longleftarrow \{\mathbf{b}, \boldsymbol{\Sigma}\}$$

?

Some remarks (2/3)

- One to one relationship

$$\omega_1^2 = \sigma_1^2,$$

$$\omega_j^2 = \sum_{k=1}^{j-1} \rho_{jk}^2 \omega_k^2 + \sum_{k,l=1,\, k<l}^{j-1} \rho_{jk}\rho_{jl}\omega_{lk} + \sigma_j^2, \quad (j \neq 1),$$

$$\omega_{ji} = \sum_{k=1, k\neq i}^{j-1} \rho_{jk}\omega_{ki} + \rho_{ji}\omega_i^2, \quad (j \neq 1).$$

- Transform a set of posterior samples

$$\{\sigma_1^{2\,(k)}, \sigma_2^{2\,(k)}, \cdots; k = 1, \ldots, N\} \longrightarrow \{\hat{\Omega}_\beta^{(k)}; k = 1, \ldots, N\}$$

- Generate $\boldsymbol{\beta}$ from the conditional posterior

$$\boldsymbol{\beta} \sim N\left(\hat{\boldsymbol{\beta}}^{(k)}, \hat{\Omega}_\beta^{(k)}\right)$$

Some remarks (3/3)

- The original model

$$\mathbf{y}_j = X_j \boldsymbol{\beta}_j + \mathbf{u}_j$$

$$E\left[\mathbf{u}_i \mathbf{u}_j{}'\right] = \begin{cases} \omega_{ij} I \ (i \neq j) \\ \omega_i^2 I \ (i = j) \end{cases}.$$

$$\pi_1(\boldsymbol{\beta}, \Omega) \propto |\,\Omega\,|^{-\frac{m+1}{2}}$$

- Transformed model

$$\mathbf{y}_j = Z_j \mathbf{b}_j + \mathbf{e}_j$$

$$E\left[\mathbf{e}_i \mathbf{e}_j{}'\right] = \begin{cases} O \ (i \neq j) \\ \sigma_i^2 I \ (i = j) \end{cases}$$

$$\pi_3(\mathbf{b}, \Sigma) = \pi_3(\mathbf{b}) \pi_3(\Sigma) \propto \prod_{j=1}^{m} (\sigma_j)^{-1}.$$

$$\pi_1(\mathbf{b}, \Sigma) \propto |\,\Omega(\mathbf{b}, \Sigma)\,|^{(m+1)/2} \times |\,J\,|_m = \prod_{j=1}^{m} (\sigma_j^2)^{\frac{m-1}{2} - j}.$$

Inference on the original model

- The conditional posteriors

$$g(\mathbf{b}_j \mid \mathbf{b}_{j-1},\ldots,\mathbf{b}_1,\sigma_j^2,D) = N\left(\hat{\mathbf{b}}_j, \sigma_j^2\left(Z_j{}'Z_j\right)^{-1}\right),$$

$$g(\sigma_j^2 \mid \mathbf{b}_{j-1},\ldots,\mathbf{b}_1,D) = IG\left(\frac{1}{2}\left(\mathbf{y}_j - Z_j\hat{\mathbf{b}}_j\right)'\left(\mathbf{y}_j - Z_j\hat{\mathbf{b}}_j\right), \frac{1}{2}\left(n - p_j - j + 1\right)\right),$$

$$g(\mathbf{b}_j \mid \mathbf{b}_{j-1},\ldots,\mathbf{b}_1,\sigma_j^2,D) = N\left(\hat{\mathbf{b}}_j, \sigma_j^2\left(Z_j{}'Z_j\right)^{-1}\right),$$

$$g(\sigma_j^2 \mid \mathbf{b}_{j-1},\ldots,\mathbf{b}_1,D) = IG\left(\frac{1}{2}\left(\mathbf{y}_j - Z_j\hat{\mathbf{b}}_j\right)'\left(\mathbf{y}_j - Z_j\hat{\mathbf{b}}_j\right), \frac{1}{2}\left(n - m - p_j - j + 1\right)\right),$$

A direct Monte Carlo (DMC) sampling procedure:

1: Fix the order of a set of m equations. Set the number of samples N to be generated. Set $j = 1$

2 Generate $\sigma_1^{2\,(k)}$, $k = 1, \ldots, N$, and insert the drawn values in $\pi(\mathbf{b}_1 \mid \sigma_1^2, D)$. Then make a draw $\mathbf{b}_1^{(k)}$, from $\pi(\mathbf{b}_1 \mid \sigma_1^{2\,(k)}, D)$, for $k = 1, \ldots, N$.

3 Increase the index $j+1 \leftarrow j$. Draw $\sigma_j^{(k)}$ from the conditional inverse gamma density $g(\sigma_j^2 \mid \mathbf{b}_{j-1}^{(k)}, \ldots, \mathbf{b}_1^{(k)}, D)$, and then generate $\mathbf{b}_j^{(k)}$ from $g(\mathbf{b}_j \mid \mathbf{b}_{j-1}^{(k)}, \ldots, \mathbf{b}_1^{(k)}, \sigma_j^{(k)}, D)$ for $k = 1, \ldots, N$

4 Repeat Step 3 sequentially until completion .

Outline

- Overview of SUR model
- Bayesian Inference for the SUR Model Using a Direct Monte Carlo Procedure
- Numerical results
- Conclusion and future works

Numerical results

Simulation study

Simulation study

- MCMC approach
- Two DMC methods
  - DMC algorithm 1

$$\pi_3(\mathbf{b},\Sigma) = \pi_3(\mathbf{b})\pi_3(\Sigma) \propto \prod_{j=1}^{m} (\sigma_j)^{-1}.$$

  - DMC algorithm 2

$$\pi_1(\mathbf{b},\Sigma) \propto |\Omega(\mathbf{b},\Sigma)|^{(m+1)/2} |J|_m = \prod_{j=1}^{m} (\sigma_j^2)^{\frac{m-1}{2}-j}$$

Simulation study

- The true model

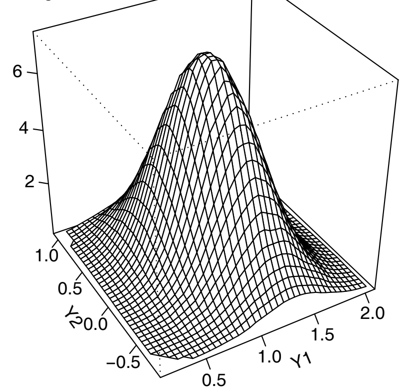
$$\begin{pmatrix} \mathbf{y}_1 \\ \mathbf{y}_2 \end{pmatrix} = \begin{pmatrix} X_1 & O \\ O & X_2 \end{pmatrix} \begin{pmatrix} \boldsymbol{\beta}_1 \\ \boldsymbol{\beta}_2 \end{pmatrix} + \begin{pmatrix} \mathbf{u}_1 \\ \mathbf{u}_2 \end{pmatrix}$$

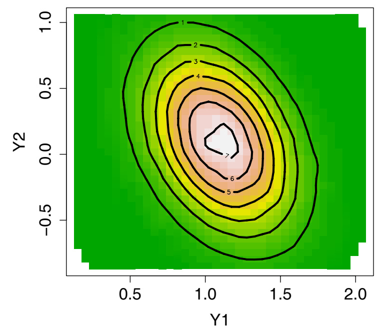
$$\boldsymbol{\beta}_1 = (3,-2)' \quad \boldsymbol{\beta}_2 = (2,1)' \qquad \Omega = \begin{pmatrix} \omega_1^2 & \omega_{12} \\ \omega_{21} & \omega_2^2 \end{pmatrix} = \begin{pmatrix} 0.1 & -0.05 \\ -0.05 & 0.2 \end{pmatrix}.$$

- Dimension of X=2
- n=100
- # of DMC sampling 10,000
- # of MCMC sampling 11,000 (1,000 burn in)

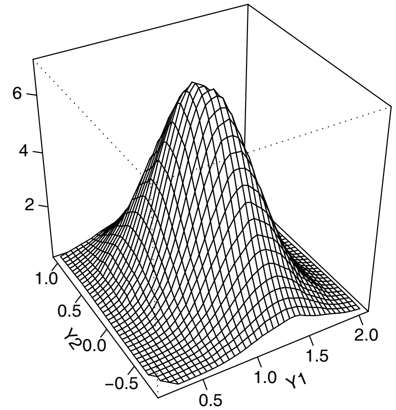
Predictive density

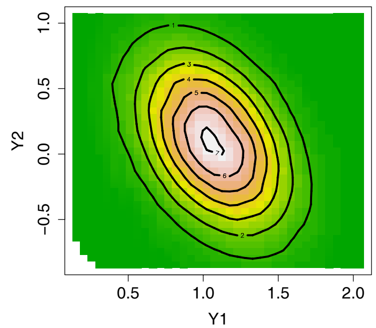
(a)True sampling density.





(b) Estimated predictive density based on the DMC 1.





Summary of the parameter estimates for DMC1

True values		Mean	SD	95%PI	
3.00	β_{11}	3.0581	0.0597	2.9400	3.1750
-2.00	β_{12}	-1.9675	0.0543	-2.0747	-1.8613
2.00	β_{21}	1.9887	0.0822	1.8279	2.1476
1.00	β_{22}	0.9418	0.0816	0.7784	1.0994
0.10	ω_1^2	0.1109	0.0162	0.0834	0.1476
-0.05	ω_{12}	-0.0345	0.0166	-0.0689	-0.0036
0.20	ω_2^2	0.2196	0.0323	0.1641	0.2889

Note: Result from 1 dataset

Summary of the computational time of each methods

	DMC1		DMC2		MCMC	
	Mean	SDs	Mean	SDs	Mean	SDs
$n = 50$	7.5881	0.0704	7.5736	0.0696	14.2223	0.2620
$n = 100$	12.3236	0.1266	12.3309	0.1214	15.1566	0.1609

The variation of the computational time for each method.

For MCMC, the computational times are measured from the initialization of parameters to the end of posterior sampling.

The computational times for our method are measured from the first posterior sampling to the end of posterior sampling.

Real data analysis 1

Incense product sales forecast

- In 2006, the size of the market for incense products in Japan was estimated to be about 30 billion yen.
- In Japan, traditional incense is used differently from lifestyle incense.
- Data consist of the daily sales figures for incense products from April, 2006 to June, 2006.
- The data were collected from two department stores, both located in Tokyo.

Incense product sales forecasting

- Model

$$y_{jt} = \beta_{j0} + \beta_{j1}x_{jt} + \beta_{j2}x_{jt} + \beta_{j3}x_{jt} + \beta_{j4}x_{jt} + u_{jt},$$
$$(t = 1,\ldots,100,\ j = 1,2)$$

Holiday effect

$$x_{j1t} = \begin{cases} 1 & (Sunday, Saturday, National\ holiday) \\ 0 & (Otherwise) \end{cases}$$

Promotion

$$x_{j2t} = \begin{cases} 1 & (Execution) \\ 0 & (Nonexecution) \end{cases}$$

Weather effect

$$x_{j3t} = \begin{cases} 1 & \text{(Fine)} \\ 0 & \text{(Cloudy)} \\ \text{-1} & \text{(Rain)} \end{cases}$$

Event

$$x_{j4t} = \begin{cases} 1 (Holding) \\ 0 (Nonholding) \end{cases}$$

Estimation results

		Mean	SD	95%PI	
	β_{10}	54.5052	5.3536	43.9621	64.9237
Holiday	β_{11}	-4.1032	5.3964	-14.6753	6.3452
Promotion	β_{12}	1.5103	5.3162	-8.8837	12.0405
Weather	β_{13}	-1.8303	3.4758	-8.698	4.9899
Event	β_{14}	17.6856	5.8354	6.2768	29.1508
	β_{20}	25.559	3.399	18.8576	32.1133
Holiday	β_{21}	9.623	4.2288	1.3021	17.8647
Promotion	β_{22}	11.9626	4.0891	3.8663	20.0723
Weather	β_{23}	2.331	2.5327	-2.6756	7.3645
Event	β_{24}	8.4699	4.2123	0.0517	16.6967
	ω_1^2	592.783	89.2884	443.5707	795.0922
	ω_{12}	109.8778	48.2728	20.7837	210.8611
	ω_2^2	331.8955	49.6295	248.3822	442.9569

DMC1 algorithm is used.

Real data analysis 2

Forecasting economic growth

- We forecast the growth rates of real sales of several sectors of the Japanese economy
- Agriculture, Automobile and Service
- 1977~2004, Quarterly data.

growth rate of the real monetary base (M2)

$$y_{j,t+1} = \beta_{j1} SR_t + \beta_{j2} M_t + \beta_{j3} GDP_t + \beta_{j4} GDP_{t-1} + \beta_{j5} GDP_{t-2} + e_{jt},$$

growth rate of TOPIX

the logarithm of quarterly real GDP

Forecasting economic growth

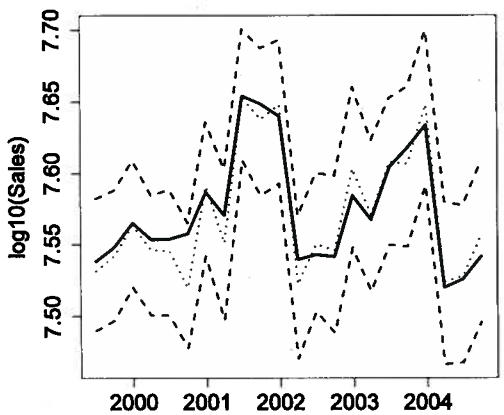
- Forecasting:
  - 1999, 2nd quarter ~2004 the 4th quarter.
- The mean of the predictive densities are used to forecast one quarter ahead growth rates
- These sector growth rate forecasts are then transformed into sales forecasts for each sector

$$\hat{Y}_{j,t+1} = \hat{y}_{j,t+1} \times Y_{j,t}$$

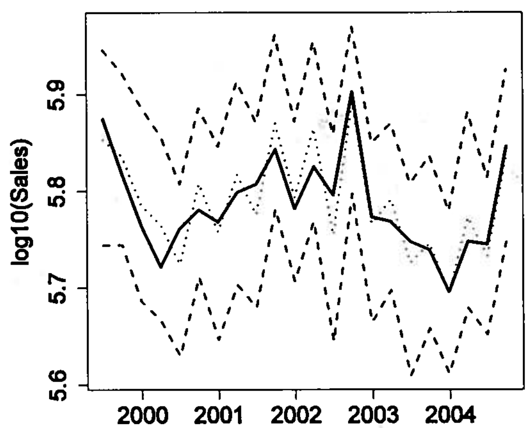
Forecast of one quarter ahead growth ($\hat{y}_{j,t+1}$)

Actual outcome at time t ($Y_{j,t}$)

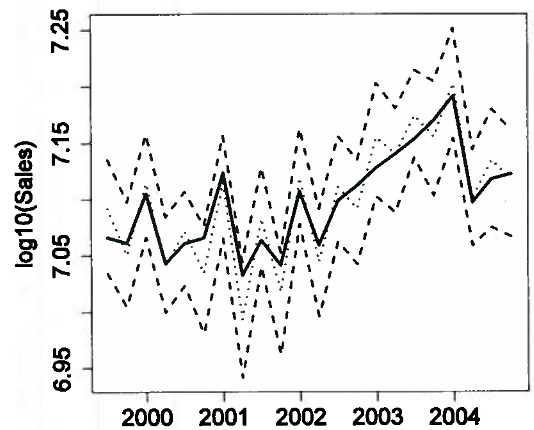
Forecasting economic growth



Agriculture



Service



Automobile

Actual output
Predictive mean
95% Posterior interval

Comparison to AR model

	RMSE		MAE	
	SUR with DMC	AR	SUR with DMC	AR
Automobile	6.1469	7.0990	6.1201	7.0965
Agriculture	5.0991	5.7887	5.0796	5.7836
Service	6.6458	7.5656	6.6365	7.5636

$$\text{RMSE}_j = \sum_{t=t_{min}}^{t_{max}} (Y_{j,t} - \hat{Y}_{j,t})^2,$$

$$\text{MAE}_j = \sum_{t=t_{min}}^{t_{max}} |Y_{j,t} - \hat{Y}_{j,t}|,$$

The SUR model with DMC1 algorithm is used.
An AIC score is used to select the AR model

Outline

- Overview of SUR model
- Bayesian Inference for the SUR Model Using a Direct Monte Carlo Procedure
- Numerical results
- Conclusion and future works

Conclusion

Conclusion

- A direct Monte Carlo (DMC) approach for Bayesian analysis of SUR models.
- Some advantages of DMC
- The method performed well in Monte Carlo experiments and applications using actual data.
- We can recommend our DMC approach for the analysis and use of the SUR model.

Conclusion

User friendly

Items	DMC	MCMC
Need to fix the number of samples drawn	Yes	Yes
100% acceptance of draws	Yes	No
Doesn't require initial parameters value	Yes	No
Doesn't need a burn-in period setting	Yes	No
Doesn't need to check for convergence	Yes	No
Doesn't need to select convergence check criteria	Yes	No
Doesn't require selection of a proposal density	Yes	Yes / No
Doesn't require use of a proposal density	Yes	Yes / No

Future work

- The DMC approach can be applied to more complex variants of the SUR model, for example those involving use of regression splines, wavelet bases, and so on.
- Fat-tailed error.
- Numb. of sampling
- Our method can be applied to the widely used simultaneous equation model.
- Forecasting Japanese economy by MMM model

References (1/5)

- Ando, T. (2007), "Bayesian predictive information criterion for the evaluation of hierarchical Bayesian and empirical Bayes models," Biometrika, 94, 443-458.
- Box, G. E. P. and Tiao, G. C. (1973), Bayesian Inference in Statistical Analysis, Reading, MA: Addison-Wesley.
- Brooks, SP. and Gelman, A. (1997), "General methods for monitoring convergence of iterative simulations," Journal of Computational and Graphical Statistics, 7, 434-455.
- Carlin, B. and Louis, T. (1996), Bayes and empirical Bayes methods for data analysis, New York: Chapman and Hall.
- Carroll, R. J., Doug M., Larry F. and Victor K. (2006), "Seemingly unrelated measurement error models, with application to nutritional epidemiology," Biometrics, 62, 75-84.
- Chib, S. and E. Greenberg (1995), "Hierarchical Analysis of SUR Models with Extensions to Correlated Serial Errors and Time-Varying Parameter Models," Journal of Econometrics, 68, 339-360.
- Frasera, D.A.S., Rekkasb, M. and Wong, A. (2005), "Highly accurate likelihood analysis for the seemingly unrelated regression problem," Journal of Econometrics, 127, 17-33.
- Gallant, R. (1975), "Seemingly unrelated nonlinear regressions," Journal of Econometrics, 3, 35-50.
- Gelman, A and Rubin, D.B. (1992), "Inference from iterative simulation using multiple sequences," Statistical Science, 7, 457-511.

References (2/5)

- Geweke, J. (1992), "Evaluating the accuracy of sampling-based approaches to calculating posterior moments," In Bayesian Statistics 4, eds. J.M. Bernado, J.O. Berger, A.P. Dawid, and A.F.M. Smith, Oxford: Clarendon Press, pp. 169-193.
- Geweke, J. (2005), Contemporary Bayesian Econometrics and Statistics, New York: Wiley.
- Greene, W.H. (2002), Econometric Analysis (5th ed.), New Jersey: Prentice-Hall.
- Gilks, W. R., Richardson, S. and Spiegelhalter, D. J. (1996). Markov Chain Monte Carlo in Practice, New York: Chapman and Hall.
- Heidelberger P and Welch PD. (1983), "Simulation run length control in the presence of an initial transient," Operations Research, 31, 1109-1144.
- Judge, G., Hill, R., Griffiths, W., Lutkepohl, H., Lee, T. (1988), Introduction to the Theory and Practice of Econometrics. New York: Wiley.
- Kim, J. K S., Shephard, N. and Chib, S. (1998), "Stochastic volatility likelihood inference and comparison with ARCH models," Review of Economic Studies, 65, 361-393.
- Kowalski, J. R. Mendoza-Blanco, X. M. Tu, and L. J. Gleser (1999), "On the difference in inference and prediction between the joint and independent t-error models for seemingly unrelated regressions," Communications in Statistics, Part A - Theory and Methods, 28, 2119-2140.
- Kurata, H. (1999), "On the efficiencies of several generalized least squares estimators in a seemingly unrelated regression model and a heteroscedastic model," Journal of Multivariate Analysis, 70, 86-94.

References (3/5)

- Jeffreys, H. (1946), "An Invariant Form for the Prior Probability in Estimation Problems," Proceedings of the Royal Society of London, Series A, 196, 453-461.
- Jeffreys, H. (1961), Theory of Probability (3rd ed.), Oxford: Oxford University Press.
- Lancaster T. (2004), Introduction to Modern Bayesian Econometrics, New Jersey: Cambridge University Press.
- Liu, A. (2002), "Efficient estimation of two seemingly unrelated regression equations," Journal of Multivariate Analysis, 82, 445-456.
- Mandy, D. M. and Martins-Filho, C. (1993), "Seemingly unrelated regressions under additive heteroscedasticity: theory and share equation applications," Journal of Econometrics, 58, 315-346.
- Meyer et al. (2003), "Stochastic volatility: Bayesian computation using automatic differentiation and the extended Kalman filter," Econometrics Journal, 6, 408-420.
- Neudecker, H. and Windmeijer, F. A. G. (1991), " in seemingly unrelated regression equations," Statistica Neerlandica, 45, 405-411.
- Ng, V. M. (2002), "Robust Bayesian Inference for Seemingly Unrelated Regressions with Elliptical Errors," Journal of Multivariate Analysis, 83, 409-414.
- Percy, D. (1992), "Predictions for Seemingly Unrelated Regressions," Journal of the Royal Statistical Society B, 54, 243-252.
- Percy, D.F. (1996), "Zellner's Influence on Multivariate Linear Models," in Bayesian Analysis in Statistics and Econometrics: Essays in Honor of Arnold Zellner, eds. D.A. Berry, K.M Chaloner and J.K. Geweke, New York: John Wiley and Sons, pp. 203-214.

References (4/5)

- Press, S. J. (1972), Applied Multivariate Analysis, New York: Holt, Rinehart and Winston, Inc.
- Raftery, A.E. and Lewis, S.M. (1992), "One long run with diagnostics: Implementation strategies for Markov chain Monte Carlo," Statistical Science, 7, 493-497.
- Rocke, D. M. (1989), "Bootstrap Bartlett adjustment in seemingly unrelated regression," Journal of the American Statistical Association, 84, 598-601.
- Rossi, P.E, Allenby, G. and McCulloch, R. (2005), Bayesian Statistics and Marketing, NJ: John Wiley and Sons.
- Schruben L.W. (1982), "Detecting initialization bias in simulation experiments," Operations Research, 30, 569-590.
- Smith, M. and R. Kohn, (2000), "Nonparametric Seemingly Unrelated Regression," Journal of Econometrics, 98, 257-282.
- 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, 64, 583-639.
- Srivastava, V. K. and Giles, D. E. A. (1987), Seemingly Unrelated Regression Equations Models, New York: Dekker.
- Tierney, L. (1994), "Markov chains for exploring posterior distributions (with discussion)," Annals of Statistics, 22, 1701-1762.
- van der Merwe, A., Viljoen, C. (1988), "Bayesian analysis of the seemingly unrelated regression model," Manuscript, University of the Free State, Department of Mathematical Statistics.

References (5/5)

- Zellner, A. (1962), "An efficient method of estimating seemingly unrelated regression equations and tests for aggregation bias," Journal of the American Statistical Association, 57, 348-368.
- Zellner, A. (1963), "Estimators for seemingly unrelated regression equations: some exact finite sample results," Journal of the American Statistical Association, 58, 977-992.
- Zellner, A. (1971), An introduction to Bayesian inference in econometrics, New York : Wiley.
- Zellner, A. Bauwens, L. and Van Dijk, H. K. (1988), "Bayesian specification analysis and estimation of simultaneous equation models using Monte Carlo Methods," Journal of Econometrics, 38, 39-72.
- Zellner, A. and Chen, B, (2002), "Bayesian Modeling of Economies and Data Requirements," Macroeconomic Dynamics, 5, 673-700.
- Zellner, A. and Min, C.K. (1995), "Gibbs Sampler Convergence Criteria," Journal of the American Statistical Association, 90, 921-927.

Appendix

Appendix1: MCMC sampling algorithm (1/3)

Sampling $\boldsymbol{\beta}$

At the k-th iteration, we generate a candidate $\boldsymbol{\beta}^{(k)}$ from the proposal density $g(\boldsymbol{\beta}|\Omega_{\mathrm{MLE}}, D)$ in (3), where Ω_{MLE} is set to be the maximum likelihood estimate. Then we accept the candidate draw with the probability α and reject the draw with the probability $1-\alpha$. The probability α is defined as

$$\alpha = \min\left\{1, \frac{h_1(\boldsymbol{\beta}^{(k)}, \Omega^{(k-1)})/g(\boldsymbol{\beta}^{(k)}|\Omega_{\mathrm{MLE}}, D)}{h_1(\boldsymbol{\beta}^{(k-1)}, \Omega^{(k-1)})/g(\boldsymbol{\beta}^{(k-1)}|\Omega_{\mathrm{MLE}}, D)}\right\},$$

with

$$h_1(\boldsymbol{\beta}, \Omega) = |\Omega|^{-(n+m+1)/2} \exp\left[-\frac{1}{2}\mathrm{tr}\left\{R(\boldsymbol{\beta})\Omega^{-1}\right\}\right],$$

where the ijth elements of $m \times m$ matrix $R(\boldsymbol{\beta})$ is $(\boldsymbol{y}_i - X_i\boldsymbol{\beta}_i)'(\boldsymbol{y}_j - X_j\boldsymbol{\beta}_j)$.

Appendix1: MCMC sampling algorithm (2/3)

Sampling Ω

Generate a candidate $\Omega^{(k)}$ from the Wishart proposal density $g(\Omega|\boldsymbol{\beta}_{\mathrm{MLE}}, D)$ in (3), where $\boldsymbol{\beta}_{\mathrm{MLE}}$ is the maximum likelihood estimate. The candidate draw is accepted with the probability

$$\alpha = \min\left\{1, \frac{h_1(\boldsymbol{\beta}^{(k)}, \Omega^{(k)})/g(\Omega^{(k)}|\boldsymbol{\beta}_{\mathrm{MLE}}, D)}{h_1(\boldsymbol{\beta}^{(k)}, \Omega^{(k-1)})/g(\Omega^{(k-1)}|\boldsymbol{\beta}_{\mathrm{MLE}}, D)}\right\},$$

and reject the draw with the probability $1 - \alpha$.

Appendix1: MCMC sampling algorithm (3/3)

1. (Initialization). Initialize $\boldsymbol{\beta}$ and Ω as the maximum likelihood estimates.
2. Sample the coefficient parameter $\boldsymbol{\beta}$.
3. Sample the coefficient parameter Ω.
4. Repeat Steps 2 and 3 for a sufficiently long time.

Summary of the parameter estimates for DMC2

True values		Mean	SD	95%PI	
3.00	β_{11}	3.0580	0.0597	2.9409	3.1737
-2.00	β_{12}	-1.9672	0.0549	-2.0728	-1.8595
2.00	β_{21}	1.9899	0.0821	1.8263	2.1494
1.00	β_{22}	0.9438	0.0814	0.7818	1.1011
0.10	ω_1^2	0.1111	0.0161	0.0840	0.1473
-0.05	ω_{12}	-0.0347	0.0167	-0.0695	-0.0038
0.20	ω_2^2	0.2224	0.0328	0.1676	0.2956

Note: Result from 1 dataset

Summary of the parameter estimates for MCMC

True values		Mean	SD	95%PI		INEF	CD
3.00	β_{11}	3.0387	0.0517	2.9377	3.1418	1.4768	1.2326
-2.00	β_{12}	-1.9790	0.0475	-2.0738	-1.8858	0.4317	0.8065
2.00	β_{21}	1.9264	0.0711	1.7849	2.0660	1.2698	1.7746
1.00	β_{22}	0.9542	0.0704	0.8160	1.0927	1.2881	-1.4312
0.10	ω_1^2	0.1110	0.0160	0.0839	0.1472	0.5902	1.1095
-0.05	ω_{12}	-0.0352	0.0165	-0.0701	-0.0051	0.6408	0.0255
0.20	ω_2^2	0.2176	0.0319	0.1636	0.2861	0.2162	0.4975

Note: Result from 1 dataset
In Appendix, the MCMC details are described

Summary of the results (100 Monte Carlo trials, n=50)

True values	n=50	DMC1		DMC2		MCMC	
		Mean	SDs	Mean	SDs	Mean	SDs
3.00	β_{11}	3.0011	0.0804	3.0010	0.0803	3.0009	0.0801
-2.00	β_{12}	-1.9881	0.0782	-1.9880	0.0783	-1.9878	0.0780
2.00	β_{21}	1.9945	0.1021	1.9944	0.1021	1.9946	0.1025
1.00	β_{22}	0.9911	0.1113	0.9910	0.1115	0.9907	0.1116
0.10	ω_1^2	0.1053	0.0215	0.1052	0.0215	0.1054	0.0215
-0.05	ω_{12}	-0.0507	0.0239	-0.0508	0.0239	-0.0540	0.0251
0.20	ω_2^2	0.2186	0.0484	0.2166	0.0482	0.2149	0.0478

Summary of the results (100 Monte Carlo trials, n=100)

True values	n=100	DMC1		DMC2		MCMC	
		Mean	SDs	Mean	SDs	Mean	SDs
3.00	β_{11}	2.9977	0.0484	2.9977	0.0485	2.9972	0.0483
-2.00	β_{12}	-2.0045	0.0520	-2.0045	0.0521	-2.0048	0.0519
2.00	β_{21}	1.9939	0.0749	1.9940	0.0749	1.9937	0.0748
1.00	β_{22}	0.9919	0.0704	0.9920	0.0705	0.9918	0.0704
0.10	ω_1^2	0.1067	0.0155	0.1067	0.0155	0.1069	0.0155
-0.05	ω_{12}	-0.0527	0.0150	-0.0526	0.0150	-0.0544	0.0155
0.20	ω_2^2	0.2071	0.0253	0.2062	0.0252	0.2056	0.0252

Summary of the results (100 trials; the same data)

n=50	DMC1		DMC2		MCMC	
	Mean	SDs	Mean	SDs	Mean	SDs
β_{11}	2.9886	0.00076	2.9876	0.00066	2.9926	0.00078
β_{12}	-1.9118	0.00076	-1.9113	0.00071	-1.9141	0.00079
β_{21}	1.9800	0.00085	1.9800	0.00091	1.9810	0.00108
β_{22}	0.9057	0.00078	0.9055	0.00071	0.9046	0.00107
ω_1^2	0.1091	0.00024	0.1092	0.00020	0.1087	0.00030
ω_{12}	-0.0667	0.00022	-0.0661	0.00021	-0.0673	0.00025
ω_2^2	0.1934	0.00035	0.1930	0.00037	0.1925	0.00045

n=100	DMC1		DMC2		MCMC	
	Mean	SDs	Mean	SDs	Mean	SDs
β_{11}	3.0030	0.00056	3.0033	0.00057	3.0033	0.00067
β_{12}	-1.9578	0.00052	-1.9571	0.00060	-1.9573	0.00061
β_{21}	2.0045	0.00066	2.0043	0.00064	2.0035	0.00085
β_{22}	1.0263	0.00059	1.0264	0.00070	1.0260	0.00093
ω_1^2	0.1081	0.00014	0.1081	0.00014	0.1084	0.00020
ω_{12}	-0.0352	0.00016	-0.0350	0.00016	-0.0344	0.00020
ω_2^2	0.1945	0.00026	0.1926	0.00027	0.1903	0.00033

Variable selection

- Improper prior ~~⇒~~ Bayes factor

- Information criteria

$$\mathrm{IC} = -2\int \log L(D \mid \mathbf{b}, \Sigma) g(\mathbf{b}, \Sigma \mid D) d\mathbf{b} d\Sigma + 2 \times s,$$

- BPIC (Ando, 2007) $s = \dim\{\mathbf{b}\} + \dim\{\Sigma\}$

- DIC (Spiegelhalter et al. 2002)

$$s = \log L(D \mid \bar{\mathbf{b}}, \bar{\Sigma}) - \int \log L(D \mid \mathbf{b}, \Sigma) g(\mathbf{b}, \Sigma \mid D) d\mathbf{b} d\Sigma$$

Variable selection results (100 Monte Carlo trials)

	$n = 50$		$n = 100$	
	BPIC	DIC	BPIC	DIC
Model 1 $\mathbf{x}_1 = x_{11}, \mathbf{x}_2 = x_{21}$	0	0	0	0
Model 2 $\mathbf{x}_1 = x_{11}, \mathbf{x}_2 = x_{22}$	0	0	0	0
Model 3 $\mathbf{x}_1 = x_{11}, \mathbf{x}_2 = (x_{21}, x_{22})'$	0	0	0	0
Model 4 $\mathbf{x}_1 = x_{21}, \mathbf{x}_2 = x_{21}$	91	80	94	83
Model 5 $\mathbf{x}_1 = x_{21}, \mathbf{x}_2 = x_{22}$	0	0	0	0
Model 6 $\mathbf{x}_1 = x_{21}, \mathbf{x}_2 = (x_{21}, x_{22})'$	4	9	2	11
Model 7 $\mathbf{x}_1 = (x_{11}, x_{12})', \mathbf{x}_2 = x_{21}$	5	10	2	4
Model 8 $\mathbf{x}_1 = (x_{11}, x_{12})', \mathbf{x}_2 = x_{22}$	0	0	0	0
Model 9 $\mathbf{x}_1 = (x_{11}, x_{12})', \mathbf{x}_2 = (x_{21}, x_{22})'$	0	1	0	2

Summary of the parameter estimates from other MCMC setting

Meyer et al. (2003)

	Mean	SDs	95%PIs	
β_{11}	2.9342	0.0054	2.9236	2.9452
β_{12}	-1.9871	0.0050	-1.997	-1.9771
β_{21}	1.9943	0.0076	1.9796	2.0092
β_{22}	0.8994	0.0075	0.8845	0.9141
ω_1^2	0.1041	0.0151	0.0787	0.1379
ω_{12}	-0.0430	0.0159	-0.0769	-0.0133
ω_2^2	0.2186	0.0316	0.1655	0.2887

Previous table

	Mean	SD	95%PI		INEF	CD
β_{11}	3.0387	0.0517	2.9377	3.1418	1.4768	1.2326
β_{12}	-1.9790	0.0475	-2.0738	-1.8858	0.4317	0.8065
β_{21}	1.9264	0.0711	1.7849	2.0660	1.2698	1.7746
β_{22}	0.9542	0.0704	0.8160	1.0927	1.2881	-1.4312
ω_1^2	0.1110	0.0160	0.0839	0.1472	0.5902	1.1095
ω_{12}	-0.0352	0.0165	-0.0701	-0.0051	0.6408	0.0255
ω_2^2	0.2176	0.0319	0.1636	0.2861	0.2162	0.4975

Following the suggestion for the MCMC method of Meyer et al. (2003), we set the proposal density for beta as normal with mean equal to the posterior mode and the variance equal to 1/10 of the previous one