

Bayes Inference for the Spatial Bilinear Time Series Model with Application to Epidemic Data

Sung Duck Lee¹ · Duk Ki Kim²

¹Department of Information & Statistics, Chungbuk National University

²Department of Information & Statistics, Chungbuk National University

(Received May 12, 2012; Revised May 14, 2012; Accepted June 13, 2012)

Abstract

Spatial time series data can be viewed as a set of time series simultaneously collected at a number of spatial locations. This paper studies Bayesian inferences in a spatial time bilinear model with a Gibbs sampling algorithm to overcome problems in the numerical analysis techniques of a spatial time series model. For illustration, the data set of mumps cases reported from the Korea Center for Disease Control and Prevention monthly over the years 2001~2009 are selected for analysis.

Keywords: Spatial time series data, STARMA, STBL, Bayesian, MCMC, Gibbs sampling, Mumps data.

1. 서론

일반적으로 공간자료(Spatial Data)는 특정 위치 또는 특정 지역에서의 관측 값이나 측정값으로 구성되는데 이러한 공간자료 분석의 주요 문제는 n 개의 공간의 정보에 기초해서 관측되지 않았거나 측정되지 않은 공간의 값을 예측하는데 있다. 그러나 일반적으로 공간시계열자료(Spatial Time Series Data)는 특정 공간의 위치와 시간의 흐름에 따라 동시에 얻어지는 관측 값이나 측정값으로 구성된다. 이러한 공간 자료의 분석에 있어서 시계열의 모형과 개념을 적용시키는 공간통계학과 시계열 분석의 융합으로 이루어진 공간시계열 모형(Spatial Time Series model)의 개념을 적용시켜서 모델링을 해야 한다.

공간시계열모형은 시간과 공간을 모두 고려해야 하기 때문에 시간과 공간의 파라미터를 갖고 있는 모형으로 매우 복잡하고 또한 많은 모수를 갖게 된다. 그러므로 모수추정에 있어서 추정의 정도를 높일 수 있는 여러 가지 수치해석적인 알고리즘의 연구가 진행되어 왔으나 기존의 수치해석적인 방법은 초기치 선정문제와 수렴 안정성에 여러 가지 한계가 있다. 기존의 공간시계열 자료에 대한 연구로 Sansó와 Guenni (1999)이 공간시계열모형을 통한 베이시안(Bayesian)방법의 통계적 추론을 연구하였고, Pfeifer와 Deutsch (1980)는 STARMA(공간시계열자기회귀이동평균)모형을 제시하였고, Chen (1992)은 비선형모형인 증선형(BL)모형에서의 모수추정 방법으로 Gibbs sampling방법을 이용한 베이시안 추정에 대해 연구하였다.

This research was supported by Basic Science Research Program through the National Research Foundation of Korea(NRF) funded by the Ministry of Education, Science and Technology(NRF-2010-0021743).

¹Corresponding author: Professor, Department of Information & Statistics, Chungbuk National University, 52 Naesudong-ro, Heungdeok-gu, Cheongju, Chungbuk 361-763, Korea. E-mail: ssdlee@chungbuk.ac.kr

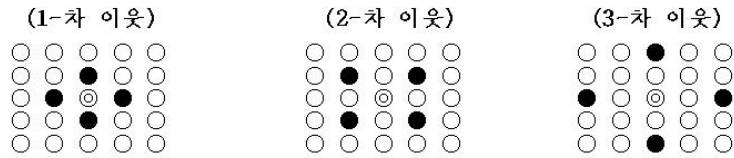


Figure 2.1. The neighbor structure of spatial order in lattice data

$$W^{(1)} = \begin{bmatrix} 0 & \frac{1}{2} & 0 & 0 & 0 & \frac{1}{2} & 0 & 0 & 0 & 0 & \dots \\ \frac{1}{3} & 0 & \frac{1}{3} & 0 & 0 & 0 & \frac{1}{3} & 0 & 0 & 0 & \dots \\ 0 & \frac{1}{3} & 0 & \frac{1}{3} & 0 & 0 & 0 & \frac{1}{3} & 0 & 0 & \dots \\ 0 & 0 & \frac{1}{3} & 0 & \frac{1}{3} & 0 & 0 & 0 & \frac{1}{3} & 0 & \dots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \ddots \end{bmatrix}$$

Figure 2.2. The first-order neighborhood weight matrix

본 논문에서는 공간시계열모형 분석에 비교적 잘 적용된다는 STARMA모형과 비선형모형인 STBL(공간시계열 중선형)모형에 대해 연구하였고, 특히 실증분석에 사용된 Mumps(유행성이하선염)자료가 중선형(BL)모형의 성격을 띠므로 Chen의 연구를 확장시켜 STBL모형에 대해 Gibbs sampling방법을 이용한 베이시안 추정방법에 초점을 두어 연구하였다.

Tierney와 Kadane (1986)에 의해 제안된 MCMC(Markov Chain Monte Carlo)방법 등을 이용하여 앞서 제시한 공간시계열모형에서 모수들에 대한 완전조건부분포를 유도하고, 모수추정에 있어 Gelfand와 Smith (1990)에 의해 제안된 깁스샘플링(Gibbs-Sampling)방법을 적용하여 모의실험과 실증분석으로 한국의 Mumps(유행성 이하선염)자료를 이용해서 STBL모형의 모수들에 대한 사전분포를 정의하고 깁스샘플링을 이용하여 사후분포(Posterior distribution)를 구하고 모수를 추정하는 방법과 추정된 모수의 수렴안정성을 진단해 보았다.

2. 가중행렬과 공간시계열모형

2.1. 가중행렬(Weight Matrix)

Figure 2.1은 공간차수에서 격자자료(lattice-data)의 1차~3차 이웃(neighbor)하고 있는 구조의 예를 나타낸 그림이다. 1차 이웃의 경우 한 위치(⊙)에서 가장 가까우며 동일한 유클리드 거리를 갖는 집합(●)이며, 2차 이웃의 경우 1차 이웃보다 먼 동일한 유클리디안 거리를 갖는 집합으로 나타낸다. 3~4차 이웃의 경우도 마찬가지로 방법으로 나타낼 수 있다.

위 5×5 ($n = 25$) 격자자료의 경우 각 위치는 왼쪽 위에서부터 오른쪽 아랫방향으로 가중행렬(weight matrix)을 Figure 2.2와 같이 결정할 수 있다.

2.2. 공간시계열 중선형모형(STBL)

공간시계열 중선형모형(Spatial-Time Bilinear Model)은 비선형 시계열모형인 중선형(BL)모형에 공간적 정보를 반영하여 시간과 공간의 파라미터를 갖는 모형으로 특징지어질 수 있으며, STBL모형은

STBL($p_{\lambda_i}, q_{\eta_j}, r_{\xi_i}, s_{\mu_j}$)로 표현하고 모형은 다음과 같다.

$$Z(t) = \sum_{i=1}^p \sum_{m=0}^{\lambda_i} \phi_m^i W^{(m)} Z(t-i) + \sum_{j=1}^q \sum_{n=0}^{\eta_j} \theta_n^j W^{(n)} e(t-i) + \sum_{i=1}^r \sum_{j=1}^s \sum_{m=0}^{\xi_i} \sum_{n=0}^{\mu_j} \beta_{mn}^{ij} \left[W^{(m)} Z(t-i) \right] \# \left[W^{(n)} e(t-j) \right] + e(t) \quad (2.1)$$

p, q : 최대 자기회귀차수, 최대 이동평균차수

λ_i, η_j : i 번째 자기회귀항의 차수와 j 번째 이동평균항의 차수

ϕ_m^i, θ_n^j : 공간차수가 m, n , 시차가 i, j 인 자기회귀모수, 이동평균모수

$W^{(m)}$: 공간차수가 m 인 $n \times n$ 가중행렬

r, s : 증선형 항의 최대자기회귀차수, 최대이동평균차수

ξ_i, μ_j : 증선형 항의 i, j 번째 시차에서 m, n 번째 공간차수의 증선형모수

β_{mn}^{ij} : 자기회귀, 이동평균의 i, j 번째 시차에서 m, n 번째 공간차수의 증선형모수

$z(t) = [z_1(t), \dots, z_n(t)]^T$: $n \times 1$ 확률벡터과정

$e(t) = [e_1(t), \dots, e_n(t)]^T$: $n \times 1$ 확률잡음벡터

$\#$ 은 행렬 $A = (a_{ij}), B = (b_{ij})$ 에서 $C = A\#B \rightarrow c_{ij} = a_{ij}b_{ij}$ 의 값을 계산하는 연산자

$$E[e(t)] = 0, E[e(t)e(t+j)^T] = \begin{cases} \sigma^2 I_n, & j = 0 \\ 0, & j \neq 0 \end{cases}, E[z(t)e(t+j)^T] = 0, j > 0.$$

3. STBL모형에 대한 베이지안 추정방법

3.1. 김스 샘플링을 이용한 STBL모형의 베이지안 추정

비선형 시계열모형인 STBL모형에 대한 베이지안 추정방법을 살펴보면 다음과 같다. 공간시계열증선형 STBL($p_{\lambda_i}, q_{\eta_j}, r_{\xi_i}, s_{\mu_j}$) 모형에서 STBL($1_1, 1_1, 1_1, 1_1$)로 가정한다.

$$z(t) = y(t)\Theta + e(t) \\ e(t) = [e_1(t), \dots, e_n(t)]^T, t = 1, \dots, T, \quad e(t) = 0, t = 1, \dots, (m-1), \\ (\text{여기서, } m = \max(p, q, r, s) + 1 = 2)$$

여기서, $\Theta = (\phi_0, \phi_1, \theta_0, \theta_1, \beta_{00}, \beta_{01}, \beta_{10}, \beta_{11})^T$ 이고, $y(t)$ 는 다음과 같이 정의된다.

$$y(t) = \left[z(t-1), W^{(1)}z(t-1), e(t-1), W^{(1)}e(t-1), z(t-1)\#e(t-1), \right. \\ \left. z(t-1)\#W^{(1)}e(t-1), W^{(1)}z(t-1)\#e(t-1), W^{(1)}z(t-1)\#W^{(1)}e(t-1) \right]$$

공간(i)와 시간(t)에 대한 표현:

$$z_i(t) = \phi_0 z_i(t-1) + \phi_1 \sum_{j=1}^n w_{ij} z_j + \theta_0 e_i(t-1) + \theta_1 \sum_{j=1}^n w_{ij} e_j(t-1) \\ + \beta_{00} z_i(t-1) e_i(t-1) + \beta_{01} z_i(t-1) \left[\sum_{j=1}^n w_{ij} e_j(t-1) \right]$$

$$\begin{aligned}
& +\beta_{10} \left[\sum_{j=1}^n w_{ij} z_j(t-1) \right] e_i(t-1) + \beta_{11} \left[\sum_{j=1}^n w_{ij} z_j(t-1) \right] \left[\sum_{j=1}^n w_{ij} e_j(t-1) \right] + e_i(t) \\
& i = 1, \dots, n(\text{공간}), \quad t = 1, \dots, T(\text{시간}) \\
& e_i(t) = z_i(t) - y_i(t)\Theta, \quad e_i(t) \sim \text{iid } N(0, \sigma^2), \quad \Theta = (\phi_0, \phi_1, \theta_0, \theta_1, \beta_{00}, \beta_{01}, \beta_{10}, \beta_{11})^T
\end{aligned}$$

$$\begin{aligned}
y_i(t) = & \left[z_i(t-1), \sum_{j=1}^n w_{ij} z_j(t-1), e_i(t-1), \sum_{j=1}^n w_{ij} e_j(t-1), z_i(t-1)e_i(t-1), \right. \\
& \left. z_i(t-1) \left[\sum_{j=1}^n w_{ij} e_j(t-1) \right], \left[\sum_{j=1}^n w_{ij} z_j(t-1) \right] e_i(t-1), \left[\sum_{j=1}^n w_{ij} z_j(t-1) \right] \left[\sum_{j=1}^n w_{ij} e_j(t-1) \right] \right].
\end{aligned}$$

우도함수를 다음과 같이 근사적으로 구할 수 있다.

$$L(\Theta, \sigma^2 | S) \propto \left(\frac{1}{\sigma^2} \right)^{\frac{n(T-m+1)}{2}} \exp \left\{ -\frac{1}{2\sigma^2} \sum_{i=1}^n \sum_{t=n}^T e_i(t)^2 \right\} \quad (3.1)$$

여기서, $S = [z(1), \dots, z(T), e(1), \dots, e(m-1)]^T$ 이고, $e(t-1)$ 이 $\hat{e}(t-1)$ 로 알려져 있다고 가정하면 우도함수를 다음과 같이 근사적으로 구할 수 있다.

$$L^*(\Theta, \sigma^2 | S) \propto \left(\frac{1}{\sigma^2} \right)^{\frac{n(T-m+1)}{2}} \exp \left\{ -\frac{1}{2\sigma^2} \sum_{i=1}^n \sum_{t=n}^T e_i^*(t)^2 \right\}, \quad (3.2)$$

$$\begin{aligned}
e_i^*(t) = & z_i(t) - \theta_0 z_i(t-1) - \phi_1 \sum_{j=1}^n w_{ij} z_j(t-1) - \theta_0 \hat{e}_i(t-1) - \theta_1 \sum_{j=1}^n w_{ij} \hat{e}_j(t-1) \\
& - \beta_{00} z_i(t-1) \hat{e}_i(t-1) - \beta_{01} z_i(t-1) \left[\sum_{j=1}^n w_{ij} \hat{e}_j(t-1) \right] - \beta_{10} \left[\sum_{j=1}^n w_{ij} z_j(t-1) \right] \hat{e}_i(t-1) \\
& - \beta_{11} \left[\sum_{j=1}^n w_{ij} z_j(t-1) \right] \left[\sum_{j=1}^n w_{ij} \hat{e}_j(t-1) \right].
\end{aligned}$$

식 (3.2)에 대한 Θ 와 $\tau = 1/\sigma^2$ 의 공액사전분포(conjugate prior)는 다음과 같다.

$$\pi(\Theta, \tau) = \pi_1(\Theta | \tau) \pi_2(\tau), \quad \Theta | \tau \sim N_k(\Theta_0, V^{-1}), \quad \tau \sim \Gamma(\alpha, \beta) \quad (3.3)$$

여기서, $k = (p+1) + (q+1) + (r+1)(s+1)$, ($k=8$), $\alpha > 0$, $\beta > 0$, V^{-1} 는 공분산행렬, V 는 $(k \times k)$ 양정치행렬이다. 베이지안 기법을 이용해 사후결합 확률분포를 다음과 같이 근사적으로 구할 수 있다.

$$p(\Theta, \tau | z, \hat{e}) \propto L^* \times \pi_1(\Theta | \tau) \times \pi_2(\tau)$$

여기서, $z : (n \times T)$ 자료행렬, $\hat{e} : (n \times T)$ 잔차행렬 $e_i^*(t) = z_i(t) y_i^*(t) \Theta$, $y_i^*(t) : e(t-1) \rightarrow \hat{e}(t-1)$ 대체 Θ_0 는 평균벡터 0 , V^{-1} 의 대각선 성분 τ_j , $j = 1, \dots, k$ 로 가정한다.

$$\begin{aligned}
p(\Theta, \tau | z, \hat{e}) \propto & \tau^{\frac{n(T-m+1)}{2}} \exp \left\{ -\frac{\tau}{2} \sum_{i=1}^n \sum_{t=m}^T e_i^*(t)^2 \right\} \times \tau_1^{-\frac{1}{2}} \exp \left[-\frac{\phi_0^2}{2\tau_1} \right] \\
& \times \tau_2^{-\frac{1}{2}} \exp \left[-\frac{\phi_0^2}{2\tau_2} \right] \times \tau_3^{-\frac{1}{2}} \exp \left[-\frac{\theta_0^2}{2\tau_3} \right] \times \tau_4^{-\frac{1}{2}} \exp \left[-\frac{\theta_1^2}{2\tau_4} \right]
\end{aligned} \quad (3.4)$$

$$\begin{aligned} & \times \tau_5^{-\frac{1}{2}} \exp \left[-\frac{\beta_{00}^2}{2\tau_5} \right] \times \tau_6^{-\frac{1}{2}} \exp \left[-\frac{\beta_{01}^2}{2\tau_6} \right] \times \tau_7^{-\frac{1}{2}} \exp \left[-\frac{\beta_{10}^2}{2\tau_7} \right] \\ & \times \tau_8^{-\frac{1}{2}} \exp \left[-\frac{\beta_{11}^2}{2\tau_8} \right] \times \frac{1}{\Gamma(\alpha)\beta^\alpha} (\tau)^{\alpha-1} \exp \left[-\frac{\tau}{\beta} \right]. \end{aligned}$$

식 (3.5), 식 (3.6)과 같이 모수들의 조건부사후분포를 구할 수 있다.

$$p(\Theta|z, \hat{e}, \tau) \sim N_k(\Theta_*^{-1}) \tag{3.5}$$

여기서,

$$\begin{aligned} \Theta_* &= (A\tau + V)^{-1}(B\tau + V\Theta_0), \\ V_* &= (A\tau + V), \quad A = \sum_{i=1}^n \sum_{t=m}^T y_i^*(t) y_i^*(t)', \quad B = \sum_{i=1}^n \sum_{t=m}^T y_i^*(t) z_i(t), \\ y_i^*(t) &= \left[z_i(t-1), \sum_{j=1}^n w_{ij} z_j(t-1), \hat{e}_i(t-1), \sum_{j=1}^n w_{ij} e_j(t-1), z_i(t-1) \hat{e}_i(t-1), \right. \\ & \quad \left. z_i(t-1) \left[\sum_{j=1}^n w_{ij} e_j(t-1) \right], \left[\sum_{j=1}^n w_{ij} z_j(t-1) \right] \hat{e}_i(t-1), \left[\sum_{j=1}^n w_{ij} z_j(t-1) \right] \left[\sum_{j=1}^n w_{ij} e_j(t-1) \right] \right]^T. \\ p(\tau|z, \hat{e}, \Theta) &\sim \Gamma \left(\frac{(2\alpha + n(T-m+1))}{2}, \frac{(2\beta + \text{SSR})}{2} \right) \end{aligned} \tag{3.6}$$

여기서,

$$\frac{(2\beta + \text{SSR})}{\sigma^2} \sim \chi_{2\alpha+n(T-m+1)}^2, \quad \text{SSR} = \sum_{i=1}^n \sum_{t=m}^T (z_i(t) - \Theta' y_i^*(t))^2.$$

A는 $(k \times k)$ 대칭행렬(여기서, $k = 8$)이고, B는 $(k \times 1)$ 열 벡터, w_{ij} 은 i 번째 공간에 이웃한 j 번째 공간의 1차 가중행렬로 각 원소는 다음과 같다.

$$A = \begin{pmatrix} A_{11} & A_{12} & A_{13} & A_{14} & A_{15} & A_{16} & A_{17} & A_{18} \\ & A_{22} & A_{23} & A_{24} & A_{25} & A_{26} & A_{27} & A_{28} \\ & & A_{33} & A_{34} & A_{35} & A_{36} & A_{37} & A_{38} \\ & & & A_{44} & A_{45} & A_{46} & A_{47} & A_{48} \\ & & & & A_{55} & A_{56} & A_{57} & A_{58} \\ & & & & & A_{66} & A_{67} & A_{68} \\ & & & & & & A_{77} & A_{78} \\ & & & & & & & A_{88} \end{pmatrix}, \quad B = \begin{pmatrix} B_1 \\ B_2 \\ B_3 \\ B_4 \\ B_5 \\ B_6 \\ B_7 \\ B_8 \end{pmatrix},$$

$$W^{(1)} = \begin{pmatrix} 0 & 1/3 & 1/3 & 1/3 & 0 & 0 & 0 & 0 \\ 1/3 & 0 & 0 & 1/3 & 0 & 0 & 1/3 & 0 \\ 1/3 & 0 & 0 & 1/3 & 1/3 & 0 & 0 & 0 \\ 1/5 & 1/5 & 1/5 & 0 & 1/5 & 1/5 & 0 & 0 \\ 0 & 0 & 1/5 & 1/5 & 0 & 1/5 & 1/5 & 1/5 \\ 0 & 0 & 0 & 0 & 1/2 & 0 & 0 & 1/2 \\ 0 & 1/4 & 0 & 1/4 & 1/4 & 0 & 0 & 1/4 \\ 0 & 0 & 0 & 0 & 0 & 1/3 & 1/3 & 1/3 \end{pmatrix}$$

$$\begin{aligned}
A_{11} &= \sum_{i=1}^n \sum_{t=m}^T z_i(t-1)^2, \dots, & A_{88} &= \sum_{i=1}^n \sum_{t=m}^T \left[\left[\sum_{j=1}^n w_{ij} z_j(t-1) \right] \left[\sum_{j=1}^n w_{ij} \hat{e}_j(t-1) \right] \right]^2, \\
A_{12} &= \sum_{i=1}^n \sum_{t=m}^T \left[z_i(t-1) \left[\sum_{j=1}^n w_{ij} z_j(t-1) \right] \right], & A_{13} &= \sum_{i=1}^n \sum_{t=m}^T [z_i(t-1) e_i(t-1)], \dots, \\
A_{78} &= \sum_{i=1}^n \sum_{t=m}^T \left[\left[\sum_{j=1}^n w_{ij} z_j(t-1) \right] \hat{e}_i(t-1) \right] \left[\sum_{j=1}^n w_{ij} z_j(t-1) \right] \left[\sum_{j=1}^n w_{ij} \hat{e}_j(t-1) \right], \\
B_1 &= \sum_{i=1}^n \sum_{t=m}^T [z_i(t-1) z_i(t)], \dots, & B_8 &= \sum_{i=1}^n \sum_{t=m}^T \left[\left[\sum_{j=1}^n w_{ij} z_j(t-1) \right] \left[\sum_{j=1}^n w_{ij} \hat{e}_j(t-1) \right] z_i(t) \right].
\end{aligned}$$

3.2. 깃스 샘플링 기법의 적용절차

식 (3.5), 식 (3.6)의 완전조건부분포를 이용해 깃스 샘플링 알고리즘을 적용하면 다음과 같다.

$$f_1(\Theta|Z, \hat{e}, \tau), \quad f_2(\tau|Z, \hat{e}, \Theta) \quad (3.7)$$

여기서, \hat{e} 는 모르는 값으로 Stage-1의 Step-1에서 장기-STAR(long STAR)모형의 잔차를 이용해 추정한다.

[Stage-1]

(Step-1) 정상시계열자료 $z(t)$ 를 장기-STAR모형 [즉, $(P^* = p + q + r + s + 1)$, $\text{STAR}(P_\lambda^*) = \text{STAR}(5_1)$]에 적합시켜 $\hat{e}(t-1)$, $t = m, \dots, T$ 을 구하고,

$$z(t) = \sum_{i=1}^p \sum_{m=0}^{\lambda_i} \phi_m^i W^{(m)} z(t-i) + e(t)$$

$\hat{e}(t) = z(t) - y(t)\hat{\Theta}$, $t = m, \dots, T$ (여기서, $m = 6$), $\hat{e}(t) = 0$, $t = 1, \dots, (m-1)$ 로 가정.

$$\begin{aligned}
\Theta &= (\phi_{10}, \phi_{11}, \phi_{20}, \phi_{21}, \phi_{30}, \phi_{31}, \phi_{40}, \phi_{41}, \phi_{50}, \phi_{51})^T \rightarrow \hat{\Theta} \\
y(t) &= \left[z(t-1), W^{(1)} z(t-1), z(t-2), W^{(1)} z(t-2), \right. \\
&\quad \left. z(t-3), W^{(1)} z(t-3), z(t-4), W^{(1)} z(t-4), z(t-5), W^{(1)} z(t-5) \right].
\end{aligned}$$

(Step-2) (Step-1)에서 잔차 분산을 구해 σ^2 의 초기값으로 사용한다.

$$\tilde{\sigma}^2 = \frac{1}{n} (T - m + 1) \sum_{t=m}^T \hat{e}(t)^T \hat{e}(t).$$

(Step-3) 식 (3.7)의 완전조건부분포로부터 최기치 Θ_0 를 설정 후 M 개의 랜덤포본 열을 생성

1-set: $(\Theta^{(1)}, \tau^{(1)})$ 은 $f_1(\Theta|Z, \hat{e}, \tilde{\tau})$, $f_2(\tau|Z, \hat{e}, \Theta^{(1)})$ 로 부터 추출

2-set: $(\Theta^{(2)}, \tau^{(2)})$ 은 $f_1(\Theta|Z, \hat{e}, \tau^{(1)})$, $f_2(\tau|Z, \hat{e}, \Theta^{(2)})$ 로 부터 추출

- 충분히 큰 M 번 반복 후 랜덤포본 셋 $(\Theta^{(M)}, \tau^{(M)})$ 을 얻는다.

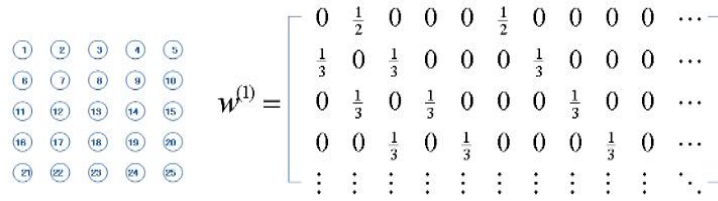


Figure 4.1. The first-order weight matrix of space($n \times n$, $n = 5$)

(Step-4) 충분히 큰 M 번 반복 후 k ($k \geq 5 \sim 30$) 번마다의 랜덤포본을 취하여 N 개의 랜덤포본 열을 얻는다.

$$\left(\Theta^{(M+k)}, \dots, \left(\Theta^{(M+Nk)}, \tau^{(M+Nk)} \right) \rightarrow \left(\Theta^{(M+k)}, \sigma^{2(M+k)} \right), \dots, \left(\Theta^{(M+Nk)}, \sigma^{2(M+Nk)} \right)$$

여기서, $\sigma^2 = \tau^{-1}$.

(Step-5) 앞에서 얻은 N 개의 사후분포 표본으로 모수를 추정 $\rightarrow (\hat{\Theta}, \hat{\sigma}^2)$.

[Stage-2]

(Step-1) Stage-1에서 추정된 $(\Theta, \hat{\sigma}^2)$ 을 이용하여 잔차를 다시 추정한다.

$$\text{STBL}(1_1, 1_1, 1_1, 1_1) : \hat{e}(t) = z(t) - y(t)\Theta, \quad t = m, \dots, T, \quad \hat{e}(t) = 0, \quad t = 1, \dots, (m - 1).$$

(Step-2) $(\Theta, \hat{\sigma}^2)$ 을 초기치로 설정 후 Stage-1 (3)~(5)를 통해 최종모수 추정 $\rightarrow (\hat{\Theta}, \hat{\sigma}^2)$.

4. 모의실험 및 실증분석

4.1. STBL 모형에 대한 모의실험 및 결과

공간시계열모형에 대한 모의실험을 위해 STBL모형의 시간, 공간차수를 1로 설정하였고, 실증분석을 위해 앞서 식별한 모형을 고려하였다.

Model : $\text{STBL}(1_1, 1_1, 1_1, 1_1) \rightarrow \Theta = (\phi_0, \phi_1, \theta_0, \theta_1, \beta_{00}, \beta_{01}, \beta_{10}, \beta_{11})$

$$\begin{aligned} z(t) &= \phi_0 z(t - 1) + \phi_1 W^{(1)} z(t - 1) + \theta_0 e(t - 1) + \theta_1 W^{(1)} e(t - 1) \\ &+ \beta_{00} z(t - 1) \# e(t - 1) + \beta_{01} z(t - 1) \# [W^{(1)} e(t - 1)] \\ &+ \beta_{10} [W^{(1)} z(t - 1)] \# e(t - 1) + \beta_{11} [W^{(1)} z(t - 1)] \# [W^{(1)} e(t - 1)] + e(t), \end{aligned}$$

$e_i(t) \sim \text{iid } N(0, 1)$ 을 가정하고, 각 모형에서 1-차 가중행렬을 고려하여 300개의 모의자료를 생성하고 초기 100개의 자료는 버리고, 200개의 모의자료를 고려하여 자료행렬 $A : (N \times T) = (5 \times 200)$ 을 이용하여 분석하였다. STBL모형의 경우 $z(t)$ 을 장기-공간시계열자기회귀모형(STAR(5₁))에 적합시켜 잔차 \hat{e} 을 구하고, 잔차의 분산($\hat{\sigma}^2$)을 구하여 σ^2 의 초기치로 사용하고, 아래와 같이 모수의 초기치를 설정하여 분석하였다.

$$\Theta_0 = 0.0, \quad V = 0.001I_8(\text{STBL}), \quad \alpha = \beta = 0.001.$$

Table 4.1. The estimated posterior statistics in each stage(STBL) in simulation

모수	True value	Stage-1		Stage-2			Percentile	
		Mean	Median	Mean	Median	SD	5%	95%
		ϕ_0	0.40	0.2886	0.2880	0.3666	0.3661	0.0285
ϕ_1	-0.30	-0.2233	-0.2231	-0.3188	-0.3186	0.0402	-0.3868	-0.2521
θ_0	-0.30	-0.1984	-0.1986	-0.2709	-0.2711	0.0357	-0.3300	-0.2102
θ_1	0.20	0.1030	0.1062	0.1832	0.1859	0.0512	0.0980	0.2636
β_{00}	0.30	0.2143	0.2141	0.2813	0.2810	0.0082	0.2683	0.2948
β_{01}	-0.20	-0.2880	-0.2872	-0.1990	-0.1983	0.0354	-0.2584	-0.1425
β_{10}	-0.20	-0.0567	-0.0552	-0.1994	-0.1981	0.0362	-0.2591	-0.1398
β_{11}	0.20	0.1556	0.1547	0.1894	0.1883	0.0238	0.1506	0.2293
σ^2	1.00	1.0692	1.0686	1.0033	1.0033	0.0207	0.9686	1.0379

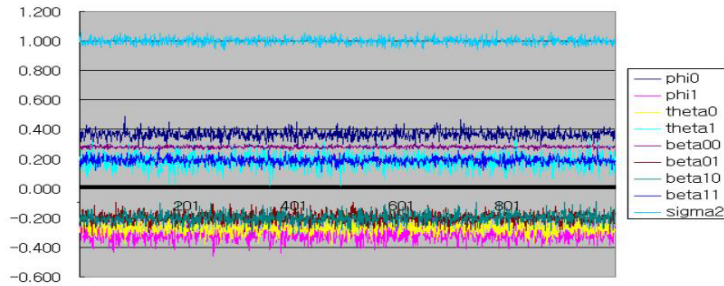


Figure 4.2. The time series plot of posterior sample about parameter of STBL in simulation

앞의 모형에 대하여 각 Stage에서 총 반복수(Total iteration) = 3,000, $M = 10,000$ (초기추출표본(Burn-in period)), $k = 20$ (증분)씩 증가시키며 랜덤포본 $N = 1000$ 개 추출하여 Stage-1, Stage-2에서 N 개의 사후분포 표본을 통하여 모수를 추정된 결과가 Table 4.1과 같다. 결과를 살펴보면 STBL모형의 1-단계에서 추정된 평균을 2-단계 모수의 초기치로 사용하여 추정된 결과 참값에 좀 더 가깝게 추정이 되는 것을 확인할 수 있다. 그리고 추정된 사후평균들이 불편성(unbiasedness)을 만족하며 percentile이 추정된 모수들을 모두 포함하고 있다. STBL모형의 모수에 대한 수렴안정성진단을 위해 N 개의 사후분포 표본을 이용해 순차도표를 그린 것이 Figure 4.1과 같다. 결과를 살펴보면, STBL모형의 참값 근방에서 랜덤하게 오르내리며 안정적으로 수렴해 가는 것을 확인할 수 있다. 이는 공간시계열모형에서 깃스 샘플링 기법을 이용한 베이지안 분석방법이 매우 유용하게 활용될 수 있음을 나타낸다.

4.2. 공간시계열모형에 대한 실증분석 결과

$$\text{Model : STBL}(1_1, 1_1, 1_1, 1_1) \rightarrow \Theta = (\phi_0, \phi_1, \theta_0, \theta_1, \beta_{00}, \beta_{01}, \beta_{10}, \beta_{11})$$

위의 STBL의 모형식을 서술하면 다음과 같다.

$$z(t) = \phi_0 z(t-1) + \phi_1 W^{(1)} z(t-1) + \theta_0 e(t-1) + \theta_1 W^{(1)} e(t-1) + \beta_{00} z(t-1) \# e(t-1) + \beta_{01} z(t-1) \# [W^{(1)} e(t-1)] + \beta_{10} [W^{(1)} z(t-1)] \# e(t-1) + \beta_{11} [W^{(1)} z(t-1)] \# [W^{(1)} e(t-1)] + e(t)$$

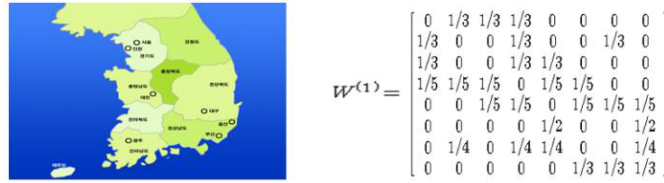


Figure 4.3. The first order weight matrix of eight-do and space($n \times n$, $n = 8$)

Table 4.2. The estimated posterior statistics in each stage(STBL) in mumps data

모수	Stage-1		Stage-2				
	Mean	Median	Mean	Median	SD	Percentile	
						5%	95%
ϕ_0	0.6228	0.6260	0.6561	0.6596	0.0659	0.5447	0.7587
ϕ_1	0.1234	0.1216	0.1288	0.1267	0.0965	-0.0248	0.2928
θ_0	-0.1680	-0.1699	-0.1962	-0.1982	0.0784	-0.3264	-0.0722
θ_1	-0.2424	-0.2407	-0.2139	-0.2119	0.1259	-0.4158	-0.0006
β_{00}	0.0037	0.0039	0.0067	0.0069	0.0252	-0.0348	0.0495
β_{01}	-0.0306	-0.0309	-0.0643	-0.0646	0.0995	-0.2229	0.1028
β_{10}	-0.0137	-0.0146	0.0331	0.0323	0.0910	-0.1134	0.1833
β_{11}	-0.0566	0.0558	-0.0758	-0.0751	0.0722	0.1895	0.0426
σ^2	0.7344	0.7332	0.7395	0.7378	0.0442	0.6695	0.8136

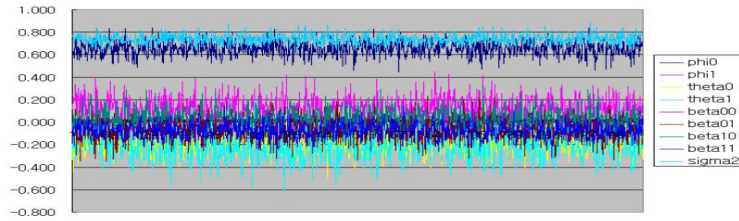


Figure 4.4. The time series plot of posterior sample about parameter of STBL in mumps data

STBL모형에 대해 정상시계열로 변환하여 얻은 84개의 자료를 이용하여 $z: (n \times T)(8 \times 84)$ 의 자료행렬로 분석하였다. STBL모형의 경우 $z(t)$ 를 장기-공간시계열자기회귀모형 ($STAR(p_\lambda) = STAR(5_1)$)에 적합시켜 잔차 $\hat{\epsilon}$ 을 구하고 잔차의 $MSE = \hat{\sigma}^2$ 을 구하여 σ 의 초기치로 사용하고, $\Theta_0 = 0.0, V = 0.001I_8, \alpha = \beta = 0.001$ 로 설정하였다. 두 모형에 대하여 각 Stage에서 총 반복수 = 3,000, $M = 10,000, k = 20$ 씩 증가시키며 랜덤포본 $N = 1000$ 개 추출하여 Stage-1, Stage-2에서 N 개의 사후분포 표본을 통하여 모수를 추정한 결과를 Table 4.2와 같다. 1-단계에서 추정된 평균(Mean)대신 중위수(Median)를 2-단계의 모수의 초기치로 사용하여 추정하는 것이 더 바람직하며 본 논문에서도 중위수를 이용하여 각 단계의 모수의 초기치로 사용하였다. 그 결과 추정의 정도가 개선되었고, 추정된 사후평균들은 근사적으로 불편성(unbiasedness)을 만족하고 percentile이 추정된 모수들을 잘 포함하는 것을 알 수 있다.

다음은 추정된 모수의 수렴안정성진단을 위해 N 개의 사후분포 표본을 이용해 순차도표를 그린 것으로 Figure 4.4와 같다. STBL모형에 대해서 살펴보면 추정된 모수 값의 근방에서 랜덤하게 오르내리며 안정적으로 수렴해 가는 것을 확인할 수 있다. 이는 매우 복잡한 공간시계열모형에서 시간과 공간차수가

조금만 커져도 매우 많은 모수를 갖게 되고 이를 추정할 때 초기치의 선정문제 및 수렴성이 떨어지는 기존의 문제에 대하여 Gibbs sampling을 이용한 베이지안 분석방법이 매우 유용하게 적용 될 수 있음을 보여주는 것이라 할 수 있다.

이는 앞에서도 본 바와 같이 Mumps 자료가 증선형 형태를 띠는 것을 반영한 STBL모형이 더 적절한 모형이라 할 수 있으며, 복잡한 공간시계열모형에 대한 기존의 수치해석적 방법이 갖고 있는 여러 한계를 극복할 수 있고 공간과 시간을 동시에 고려해야 하는 다양한 분야에서 매우 유용한 도구로써 활용될 수 있겠다.

5. 결론

공간시계열모형의 경우 시간과 공간의 차수가 커지거나 비선형성이 커지면 모형이 너무 복잡하여 기존의 방법으로 모수를 추정할 때 초기치 선정의 문제와 수렴성이 떨어진다. 이와 달리 사용 방법이 비교적 간단한 깁스 샘플링 기법을 이용한 베이지안 방법은 복잡한 비선형공간시계열모형에도 모수추정이 수월하며 수렴의 안정성을 확보할 수 있다. 또한 증선형(BL)모형에 대한 베이지안 추정에 있어 잔차를 미리 추정해서 사용하여야 하는데 Chen (1992)은 장기시차를 고려한 AR모형을 활용할 것을 제안했다. 이러한 Chen의 제안을 확장하여 비선형공간시계열모형에 대한 베이지안 분석의 경우 비교적 간단한 공간시계열자기회귀(STAR)모형의 장기시차를 고려하여 잔차를 추정하는데 활용하였다. 복잡한 비선형공간시계열모형에 대한 베이지안 분석방법이 모수추정에 있어 기존의 방법이 갖고 있는 여러 한계점을 극복할 수 있음을 모의실험과 실증분석을 통해 보였고, 비선형공간시계열모형에 있어 현재까지 연구가 미진한 베이지안 추정방법에 있어 다양한 분야에서 현실적으로 활용할 수 있는 manual을 제시하였다는데 본 논문에 의의를 둘 수 있다.

References

- Chen, C. W. S. (1992). Bayesian analysis of bilinear time series models: A Gibbs sampling approach, *Communications in Statistics - Theory and Methods*, **21**, 3407–3425.
- Gelfand, A. and Smith, A. F. M. (1990). Sampling based approaches to calculating marginal densities, *Journal of the American Statistical Association*, **85**, 398–409.
- Gelman, A. and Rubin, D. B. (1992). Inference from iterative simulation using multiple sequences, *Statistical Science*, **7**, 457–472.
- Pfeifer, P. E. and Deutsch, S. J. (1980). Identification and Interpretation of First order space-time ARMA Models, *Technometrics*, **22**, 397–408.
- Raftery, A. L. and Lewis, S. (1992). One long run with diagnostics: Implementation strategies for Markov chain Monte Carlo, *Statistical Science*, **7**, 493–507.
- Ritter, C. and Tanner, M. A. (1992). Facilitating the Gibbs sampler: The Gibbs stopper and the griddy Gibbs sampler, *Journal of the American Statistical Association*, **85**, 398–409.
- Geman, S. and Geman, D. (1984). stochastic Relaxation, Gibbs Distributions, and the Bayesian Restoration of Images, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **6**, 721–741.
- Sansó, B. and Guenni, L. (1999). Venezuelan rainfall data analysed by using a Bayesian space-time model, *Applied Statistics*, **48**, 345–362.
- Tanner, M. A. and Wong, W. H. (1987). The calculation of posterior distributions by data augmentation, *Journal of the American Statistical Association*, **82**, 528–550.
- Tierney, L. and Kadane, J. B. (1986). Accurate approximations for posterior moments and marginal densities, *Journal of American Statistical Association*, **81**, 82–86.