응용통계연구 제 11권 1호, 1998년, pp. 97-111

회귀모형 오차항의 1차 자기상관에 대한 베이즈 검정법

김혜중[†] 한성실[‡]

요 약

본 논문에서는 회귀모형 오차항의 1차 자기상관에 대한 베이즈 검정법을 제안하였다. 이를 위해 자기상관검정에서 설정된 귀무 및 대립가설간에 베이즈요인을 도출하고, 이를 근사추정하는 방법을 일반화 Savage-Dickey 밀도비와 Gibbs 추출법의 합성을 통해 제시하였다. 또한, 근사추정의 효율 및 제안된 검정법의 검정력을 평가하기 위해서 모의실험과 경험적 자료분석 예를 사용하였다.

1. 서론

일반적으로 오차항들이 서로 독립인 정규확률변수의 가정하에서 이루어지는 회귀모형의 분석에서 모형을 유효하게 추정하기 위해서는 오차항들 사이의 자기상관 존재여부를 검정하는 것이 매우 중요한 문제이다. 특히, 통계학의 응용 분야들 중에서 계량경제 및 생산공정 분야에서는 위에서 언급된 문제에 대한 인식과 더불어 위 문제의 해결에 대한 연구가 활발히 진행되어 왔다(Pindyck과 Rubinfeld(1981) 및 Alwan과 Roberts(1989) 참조).

오차항들 사이의 자기상관(ρ) 존재여부를 검정하는 방법으로 표본이론에서는 Durbin과 Watson(1951)에 의해 제안된 D 통계량이 널리 사용되고 있으며, 이들은 귀무가설 $H_0 : \rho = 0$ 하에서 D 통계량을 이용한 정확한 검정(exact test) 및 유의수준 $\alpha = 0.01, 0.025, 0.05$에 대한 상한과 하한을 이용한 검정법을 제안하였다. 이외에도, Theil(1968), Abrahamse와 Koerts(1971), Phillips와 Harvey(1974) 그리고, Marr와 Quesenberry(1991)에 의해서 귀무가설 $H_0 : \rho = 0$의 검정법이 여러 형태로 개발되었으나, 기존의 연구(Marr와 Quesenberry (1991)참조)에 의하면 타 방법들이 검정력에 있어서 Durbin과 Watson의 검정법보다 뒤떨어지는 것으로 밝혀졌다. 그러나, 잘 알려진 것과 같이 Durbin과 Watson의 검정법을 포함한 기존의 표본이론에 의한 자기상관검정법들은 귀무가설 $H_0 : \rho = 0$이 참이라는 가정하에서 도출된 검정통계량을 사용하고 있어서 일반적인 귀무가설 $H_0 : \rho = \rho_0$ $(-1 \le \rho \le 1)$의 검정에는 이들을 사용할 수 없는 문제점을 갖고 있다.

한편, 베이즈 방법론에서 볼 때 회귀모형 오차항의 자기상관검정에 대한 연구는 아직 행해지지 않은 상태이다(Palmer와 Broemeling(1993) 및 Zellner(1971) 참조). 그 이유는 자기상관계수의 주변사후확률분포가 복잡한 형태를 가지고 있다는데 기인한다. 이 점을 고려하여 본 논문에서는 베이즈 이론에 의한 자기상관 검정법을 개발하고, 제시된 검정법이 귀무가설 $H_0 : \rho = 0$ 뿐 아니라 일반적인 귀무가설 $H_0 : \rho = \rho_0$를 모두 유효하게 검정할

[†] (100-715) 서울시 중구 필동 3가 26, 동국대학교 통계학과 교수
[‡] (100-715) 서울시 중구 필동 3가 26, 동국대학교 통계학과 박사과정 수료

수 있는 것임을 보이고자 한다. 이를 위하여 회귀모형의 모수들에 대한 정보적 사전확률분포(informative prior probability distribution) 가정하에서 자기상관계수를 검정하기 위한 베이즈요인(Bayes factor)을 유도하였으며, Verdinelli와 Wasserman(1995)이 제안한 일반화 Savage-Dickey 밀도비(generalized Savage-Dickey density ratio)와 Gibbs 추출법(Casella와 George(1992) 참조)의 합성을 통하여 베이즈요인을 근사 추정하는 방법을 제시하였다. 또한, 추정된 베이즈요인에 의한 자기상관모수 검정법의 유용성을 모의실험과 경험적 자료의 분석을 통하여 검토하였다.

2. 자기상관검정을 위한 베이즈요인

2.1. 베이즈요인

일반적으로 베이지안 추론에서 대립가설에 대한 귀무가설을 검정하기 위한 방법으로 베이즈요인이 사용된다. 베이즈요인은 대립가설에 대한 귀무가설의 사전확률의 승산비(prior odds)와 사후확률의 승산비(posterior odds)에 의해서 정의된다.

장애모수 $\psi \in \Psi$를 포함한 모수공간 $(\Theta \cup \Psi)$의 한 점에 대한 귀무가설 $(H_0 : \theta \in \Theta_0)$과 대립가설 $(H_1 : \theta \notin \Theta_0)$을 고려해 보자. 여기서, $\Theta(\Theta_0 \cup \Theta_1 = \Theta)$는 θ의 모수공간으로 $\Theta_0 = \{\theta; \theta = \theta_0\}$, $\Theta_1 = \{\theta; \theta \neq \theta_0\}$ 이다.

귀무가설의 사전확률(prior probability) $P(H_0)$를 π_0라고 하고, 대립가설의 사전확률 $P(H_1)$을 π_1이라고 하자$(\pi_0+\pi_1 = 1)$. 그리고, 귀무가설하에서 ψ의 사전밀도함수를 $p_0(\psi)$로 가정하고, 대립가설하에서의 θ와 ψ의 사전밀도함수를 $p\ (\theta, \psi)$로 나타내면, 모수 θ와 ψ의 사전확률분포함수는 혼합함수의 형태로서

$$F(\theta, \psi) = \pi_0 I_{[\theta_0, \infty)}(\theta) \int_{-\infty}^{\psi} p_0(t_2) dt_2 + \pi_1 \int_{-\infty}^{\theta} \int_{-\infty}^{\psi} p(t_1, t_2) dt_1 dt_2 \tag{2.1}$$

와 같이 표현될 수 있으며, 대립가설에 대한 귀무가설의 사전확률승산비와 사후확률 승산비간의 비를 나타내는 베이즈요인은

$$B = \frac{P_0/P_1}{\pi_0/\pi_1} = \frac{\pi_1}{\pi_0} \cdot \frac{P_0}{P_1} \tag{2.2}$$

와 같이 정의된다(Lee(1988) 참조). 여기서, $P_0 = P(H_0|x) = P(\theta \in \Theta_0|x) = \frac{\Delta_1}{\Delta_1+\Delta_2}$, $P_1 = P(H_1|x) = 1 - P_0$이며, $\Delta_1 = \pi_0 \int_\Psi p_0(\psi)\ell(\theta_0, \psi) d\psi$, $\Delta_2 = \pi_1 \int_\Psi \int_\Theta p(\theta, \psi) \ell(\theta, \psi) d\theta d\psi$이고, $\ell(\theta, \psi)$는 θ와 ψ의 우도함수를 나타낸다.

따라서, 식 (2.2)에서 정의된 베이즈요인은 다음과 같이 다시 쓸 수 있다.

$$\begin{aligned} B &= \frac{\pi_1}{\pi_0} \cdot \frac{\int_\Psi \pi_0 p_0(\psi) \ell(\theta_0, \psi) d\psi}{\int_\Psi \int_\Theta \pi_1 p(\theta, \psi) \ell(\theta, \psi) d\theta d\psi} \\ &= \frac{\int_\Psi p_0(\psi) \ell(\theta_0, \psi) d\psi}{\int_\Psi \int_\Theta p(\theta, \psi) \ell(\theta, \psi) d\theta d\psi}. \end{aligned} \tag{2.3}$$

이와 같이 구해진 베이즈요인에 의한 가설검정법은 베이즈요인 값이 1보다 크면 귀무가설을 채택하고, 그렇지 않으면 귀무가설을 기각하게 된다.

2.2. 자기상관검정을 위한 베이즈요인

오차항이 다음과 같은 1차 자기회귀 과정을 가지는 단순선형회귀모형을 고려하자.

$$y'_t = \alpha + \beta x'_t + u_t,\ u_t = \rho u_{t-1} + \epsilon_t,\ t = 1, 2, \cdots, T. \tag{2.4}$$

식 (2.4)에서 y'_t는 t번째 종속변수의 관측치이고, α, β는 회귀계수이며, x'_t는 t번째 독립변수의 관측치를 나타낸다. 또한, u_t는 1차 자기상관(first order autocorrelation)을 가지는 오차항이고, ρ는 자기상관모수이며, ϵ_t는 독립이고, 평균이 0, 분산이 $1/\tau$인 정규분포를 따른다고 가정한다. 여기서, $\rho = 0$이면 식 (2.4)는 일반적인 회귀모형과 같게 된다.

본 논문에서는 식 (2.4)에서 정의된 회귀모형에서 절편을 나타내는 부분인 α를 고려하지 않은 다음과 같은 대체모형을 고려하였다.

$$y_t = \beta x_t + u_t,\ u_t = \rho u_{t-1} + \epsilon_t,\ t = 1, 2, \cdots, T. \tag{2.5}$$

여기서, $y_t = y'_t - \bar{y}$, $x_t = x'_t - \bar{x}$이고, 오차항 u_t는 평균이 0, 분산이 $1/\tau(1-\rho^2)$인 정규분포를 따르며, 자기상관모수 ρ는 인접하는 두 시점에서의 오차항 u_t와 u_{t-1}사이의 자기상관계수를 나타낸다. 또한, 모형 (2.5)는 다음과 같이 다시 표현할 수 있다.

$$y_t = \rho y_{t-1} + \beta(x_t - \rho x_{t-1}) + \epsilon_t,\ t = 1, 2, \cdots, T. \tag{2.6}$$

여기서, 완전한 관측값의 집합인 $\{x_t, y_t : t = 1, 2, \cdots, T\}$이 $\epsilon_2, \epsilon_3, \cdots, \epsilon_T$에 의해 발생되므로 결합확률분포함수(joint probability distribuion function)는 다음과 같다(Wonnacott와 Wonnacott(1979) 참조).

$$f(\mathbf{y}|\beta, \rho, \tau) \propto \tau^{T/2} exp\left\{-\frac{\tau}{2}\sum_{t=2}^{T}[y_t - \rho y_{t-1} - \beta(x_t - \rho x_{t-1})]^2\right\}. \tag{2.7}$$

단, $\mathbf{y}' = (y_1, y_2, \cdots, y_T)$이며, $\tau > 0$, $-\infty < \beta < \infty$이고, $-1 \le \rho \le 1$ 이다.

위에서 정의된 회귀모형의 자기상관모수 값을 검정하기 위해서 다음과 같은 가설을 고려해 보자.

$$H_0 : \rho = \rho_0 \quad \text{vs.} \quad H_1 : \rho \neq \rho_0. \tag{2.8}$$

여기서, ρ 의 모수공간은 $A = \{\rho; -1 \le \rho \le 1\}$로 제한하였다.

식 (2.7)에서 주어진 모수들에 대한 결합사전확률밀도함수(joint prior probability density function)를 $p_A(\beta, \rho, \tau)$로 나타내면

$$p_A(\beta, \rho, \tau) = p(\beta, \tau|\rho) p_A(\rho) \tag{2.9}$$

이 성립하고, ρ와 (β,τ)가 사전독립(prior independence)이라고 가정하면, 식 (2.9)는 다음과 같이 다시 쓸 수 있다.

$$p_A(\beta,\rho,\tau) = p(\beta,\tau)p_A(\rho). \tag{2.10}$$

(β,τ)와 ρ의 사전분포를 Parlmer와 Broemiling(1985)에서와 같이 정규-감마분포 (normal-gamma distribution)와 균일분포(uniform distribution)로 가정하자. 즉,

$$p(\beta,\tau) \;\propto\; \tau^{\alpha-1}e^{-\tau\gamma}\tau^{1/2}\, exp\left\{-\frac{\tau}{2}(\beta-\mu)^2\xi\right\},\; -\infty<\beta<\infty,\; \tau>0, \tag{2.11}$$

$$p_A(\rho) = \frac{1}{2},\; -1\le\rho\le 1. \tag{2.12}$$

식 (2.11)의 $\alpha,\ \gamma,\ \xi>0$와 $-\infty<\mu<\infty$는 사전확률분포의 모수(hyper-parameter)를 나타낸다.

그러면, 식 (2.7)과 위에서 정의된 사전분포를 이용하여 ρ의 모수공간인 절단된 집합 $A=\{\rho: -1\le\rho\le 1\}$ 에서의 (β,ρ,τ)에 대한 결합사후확률밀도함수(joint posterior probability density function)를 다음과 같이 구할 수 있다.

$$p_A(\beta,\rho,\tau|\mathbf{y}) \propto \tau^{\frac{T+2\alpha+1}{2}-1}\cdot exp\left\{-\frac{\tau}{2}\left(2\gamma+\sum_{t=2}^{T}[y_t-\rho y_{t-1}-\beta(x_t-\rho x_{t-1})]^2+(\beta-\mu)^2\xi\right)\right\},$$
$$-\infty<\beta<\infty,\; \tau>0,\; -1\le\rho\le 1. \tag{2.13}$$

따라서, 식 (2.3)을 이용하여 자기상관모수를 검정하기 위한 베이즈요인을 구하면 다음과 같다.

$$B = \frac{\int_\beta\int_\tau \ell(\rho_0,\beta,\tau)p(\beta,\tau)d\tau d\beta}{\int_{-1}^{1}\int_\beta\int_\tau \ell(\rho,\beta,\tau)p_A(\rho,\beta,\tau)d\tau d\beta d\rho}. \tag{2.14}$$

그러나, 식 (2.14)에서 정의된 베이즈요인은 자기상관모수의 주변사후확률분포가 복잡한 형태를 가지고 있어 실제로 이를 계산하는 문제가 복잡하다. 따라서, 베이즈요인을 Verdinelli와 Wasserman(1995)이 제안한 일반화 Savage-Dickey 밀도비의 형태로 변환시키고, Gibbs 추출법에 의하여 자기상관모수의 존재여부(또는 값)를 검정하기 위해 베이즈요인을 근사적으로 추정하고자 한다.

3. 베이즈요인의 근사 추정

3.1. 일반화 Savage-Dickey 밀도비

일반적으로 베이즈요인을 계산하기 어려운 경우 이를 해결하기 위한 방법으로 Laplace 방법에 의한 근사적 적분방법이 주로 사용되어 왔으나, 최근에는 사후분포로부터 추출된

표본을 이용하여 베이즈요인을 근사적으로 추정하는 방법이 많이 연구되고 있다(Newton과 Raftery(1994), Gelfand와 Dey(1994), Verdinelli와 Wasserman(1995) 참조).

이러한 방법들 중에서 Verdinelli와 Wasserman이 제안한 일반화 Savage-Dickey 밀도비를 이용하면, 식 (2.14)의 베이즈요인을 쉽게 계산할 수 있음을 이 장에서 보이고자 한다.

식 (2.14)에서 $m = \int_{-1}^{1} \int_\beta \int_\tau \ell(\rho, \beta, \tau) p_A(\rho, \beta, \tau) d\tau d\beta d\rho$ 로 나타내면, 베이즈요인 B는 다음과 같이 다시 쓸 수 있다.

$$\begin{aligned} B &= \frac{\int_\beta \int_\tau \ell(\rho_0, \beta, \tau) p(\beta, \tau) d\tau d\beta}{m} \\ &= p_A(\rho_0|\mathbf{y}) \int_\beta \int_\tau \frac{\ell(\rho_0, \beta, \tau) p(\beta, \tau)}{p_A(\rho_0|\mathbf{y}) m} d\tau d\beta \\ &= p_A(\rho_0|\mathbf{y}) \int_\beta \int_\tau \frac{\ell(\rho_0, \beta, \tau) p(\beta, \tau) p(\beta, \tau|\rho_0, \mathbf{y})}{p_A(\rho_0, \beta, \tau|\mathbf{y}) m} d\tau d\beta. \end{aligned} \tag{3.1}$$

여기서, $p_A(\rho|\mathbf{y})$는 절단된 집합(truncated set) $A = \{\rho : -1 \le \rho \le 1\}$에서의 ρ의 주변사후확률밀도함수(marginal posterior probability density function)로서 결합사후확률밀도함수인 식 (2.13)을 β와 τ에 대해서 차례로 적분하여 다음과 같이 구할 수 있다.

$$p_A(\rho|\mathbf{y}) \propto [A_1(\rho)]^{-1/2} \cdot \left[A_3(\rho) - \frac{A_2^2(\rho)}{A_1(\rho)}\right]^{-\frac{T+2\alpha}{2}}, \; -1 \le \rho \le 1. \tag{3.2}$$

위 식에서, $A_1(\rho) = \sum_{t=2}^{T} (x_t - \rho x_{t-1})^2 + \xi$, $A_2(\rho) = \sum_{t=2}^{T} (y_t - \rho y_{t-1})(x_t - \rho x_{t-1}) + \xi\mu$이고, $A_3(\rho) = 2\gamma + \sum_{t=2}^{T} (y_t - \rho y_{t-1})^2 + \xi\mu^2$이다.

또한, ρ와 β, τ의 절단된 결합사후확률밀도함수(truncated joint posterior probability density function)는 $p_A(\rho_0, \beta, \tau|\mathbf{y}) = \ell(\rho_0, \beta, \tau) p_A(\rho_0, \beta, \tau)/m$ 로 나타낼 수 있으므로, 이를 식 (3.1)에 대입하여 정리하면 베이즈요인 B를 다음과 같이 나타낼 수 있다.

$$\begin{aligned} B &= p_A(\rho_0|\mathbf{y}) \int_\beta \int_\tau \frac{p(\beta, \tau) p(\beta, \tau|\rho_0, \mathbf{y})}{p_A(\rho_0, \beta, \tau)} d\tau d\beta \\ &= \frac{p_A(\rho_0|\mathbf{y})}{p_A(\rho_0)} E\left[\frac{p(\beta, \tau)}{p(\beta, \tau|\rho_0)}\right]. \end{aligned} \tag{3.3}$$

위 식의 $E[\,\cdot\,]$는 조건부 사후확률밀도함수 $p(\beta, \tau|\rho_0, \mathbf{y})$에 대한 기대값을 나타낸다.

그런데, 모수들의 사전독립의 가정(식 (2.10)참조)에 의해 $p(\beta, \tau) = p(\beta, \tau|\rho_0)$이 성립되고, $p_A(\rho_0) = 1/2$이므로 이를 이용하여 식 (3.3)을 정리하면 베이즈요인 B는 다음과 같이 간단하게 나타낼 수 있다.

$$B = \frac{p_A(\rho_0|\mathbf{y})}{p_A(\rho_0)} = 2 p_A(\rho_0|\mathbf{y}). \tag{3.4}$$

그러나, 식 (3.2)에서 유도된 $p_A(\rho|\mathbf{y})$의 분포가 불완전한 형태(unclosed form)를 가지고 있어 식 (3.4)를 이용하여 베이즈요인 B를 직접 계산하기는 어렵고, 이를 근사적으로 추정하는 방법이 필요하다.

3.2. 베이즈요인의 근사 추정

식 (3.4)로부터 자기상관모수 ρ를 검정하기 위한 베이즈요인을 추정하는 문제는 2절에서 정의된 ρ의 모수공간 $A = \{\rho : -1 \le \rho \le 1\}$에서 유도된 주변사후확률분포함수 $p_A(\rho|\mathbf{y})$의 추정문제와 동일함을 알 수 있다.

본 절에서는 주변사후확률분포의 추정에 널리 사용되고 있는 Gibbs추출법(Gelfand와 Smith(1990) 참조)을 이용하여 식 (3.4)에서 정의된 베이즈요인 B를 추정하는 방법을 설명하고자 한다. 이를 위해서 먼저 오차항이 자기회귀과정을 가지는 단순선형회귀모형하에서 ρ와 β, τ사이의 조건부 사후확률분포(conditional posterior probability distribution)를 유도하였다.

식 (2.13)에 주어진 (ρ, β, τ)의 결합사후확률밀도함수를 β에 대해 적분하면 ρ와 τ의 결합사후확률밀도함수를 얻을 수 있고, 이를 이용하면 ρ와 τ가 주어진 경우 β의 조건부 사후확률밀도함수는 다음과 같이 유도된다. 즉,

$$p(\beta|\rho, \tau, \mathbf{y}) \propto exp\left\{-\frac{\tau}{2}\left(\beta^2 A_1(\rho) - 2\beta A_2(\rho) + \frac{A_2^2(\rho)}{A_1(\rho)}\right)\right\}. \tag{3.5}$$

따라서, ρ와 τ가 주어진 경우 β의 조건부 사후확률밀도함수는 평균이 $A_2(\rho)/A_1(\rho)$이고, 분산이 $1/[\tau A_1(\rho)]$인 정규분포임을 알 수 있다. 유사한 방법에 의해서 ρ와 τ의 조건부 사후확률밀도함수를 각각 구하면, Gibbs 추출법에 필요한 모든 조건부 사후확률분포들은 다음과 같다.

$$\begin{aligned}
&\tau|\beta, \rho, \mathbf{y} \sim G(\alpha^*, \beta^*),\\
&\beta|\tau, \rho, \mathbf{y} \sim N(A_2(\rho)/A_1(\rho), 1/\tau A_1(\rho)),\\
&\rho|\beta, \tau, \mathbf{y} \sim N_A(A_2(\beta)/A_1(\beta), 1/\tau A_1(\beta)),\ A = \{\rho\ ; -1 \le \rho \le 1\},
\end{aligned}$$

여기서, $\alpha^* = \frac{T+2\alpha+1}{2}$, $\beta^*(\rho, \beta) = 2/[A_3(\rho) + \beta^2 A_1(\rho) - 2\beta A_2(\rho)]$이고, $A_1(\beta) = \sum_{t=2}^{T}(y_{t-1} - \beta x_{t-1})^2$, $A_2(\beta) = \sum_{t=2}^{T}(y_{t-1} - \beta x_{t-1})(y_t - \beta x_t)$ 이며, $A_1(\rho)$ 및 $A_2(\rho)$는 식 (3.2)에서와 동일한 함수이다.

이와 같이 도출된 조건부 사후확률분포를 이용하면 Gibbs 추출법으로 τ, β 및 ρ의 사후표본을 추출할 수 있다. 특이한 사항은 Gibbs 알고리즘의 적용에 있어서 ρ의 주변사후확률분포는 절단된 정규분포(truncated normal distribution)의 형태로서 모수공간이 $[-1, 1]$의 구간으로 제한되어 있다. 그러므로, ρ의 사후표본 추출에 Devroy(1986)가 제안한 one-for-one 추출방법을 사용하였다.

위에서 설명한 사후표본 추출절차를 Gelfand와 Smith(1990)가 제안한 Gibbs 알고리즘에 맞게 적용하면 식 (3.4)에서 정의된 베이즈요인을 다음과 같이 추정할 수 있다.

주어진 세 모수에 대한 조건부확률분포 $\tau^{(i)} \sim \rho|\beta^{(i-1)}, \rho^{(i-1)}, \mathbf{y}$와 $\beta^{(i)} \sim \beta|\rho^{(i-1)}, \tau^{(i)}, \mathbf{y}$, $\rho^{(i)} \sim \rho|\beta^{(i)}, \tau^{(i)}, \mathbf{y}$ $(i = 1, 2, \cdots, m)$를 이용하여 m번째 iteration이 끝난 후의 첫번째 표본 $(\rho_1^{(m)}, \beta_1^{(m)}, \tau_1^{(m)})$을 얻는다. 여기서, 초기값 $\beta^{(0)}$와 $\rho^{(0)}$는 식 (2.5)로부터 구한 β와 ρ의 최소

제곱추정값(Pindyck와 Rubinfeld(1981) 참조)을 각각 사용한다. 이러한 과정을 N번 반복함으로써

$$(\rho_1^{(m)}, \beta_1^{(m)}, \tau_1^{(m)}), (\rho_2^{(m)}, \beta_2^{(m)}, \tau_2^{(m)}), \cdots, (\rho_N^{(m)}, \beta_N^{(m)}, \tau_N^{(m)})$$

와 같은 N개 사후표본을 얻을 수 있고, 이를 이용하여 식 (3.4)에서 정의된 베이즈요인은 다음과 같이 추정된다.

$$\hat{B} = 2\hat{p}_A(\rho_0|\mathbf{y}) = 2 \cdot \frac{1}{N} \cdot \sum_{j=1}^{N} p_A(\rho_0|\beta_j^{(m)}, \tau_j^{(m)}, \mathbf{y}). \tag{3.6}$$

추정량 $\hat{B}$의 성질 및 m과 N의 결정 문제는 이미 연구되어 있는 관계로 본 논문에서는 다루지 않았다(Gelfand와 Smith(1990) 참조).

4. 모의실험 및 경험적 자료분석

4.1. 모의실험

3절에서는 오차항이 1차 자기회귀과정을 가지는 단순선형회귀모형하에서 ρ와 β, τ사이의 조건부 사후확률분포를 도출하였고, 일반화 Savage-Dickey 밀도비와 Gibbs 추출법을 이용하여 자기상관모수에 대한 가설 (2.8)의 베이즈 검정에 필요한 베이즈요인의 근사 추정법(식 (3.6) 참조)을 제시하였다. 이 장에서는 제안된 추정법에 의해 추정된 베이즈요인으로 자기상관모수를 검정하는 경우, 그 검정의 유효성을 검정력의 기준에 의해 평가하기 위하여 귀무가설 $H_0 : \rho = \rho_0$하에서 다음 절차에 의해 Monte Carlo 모의실험을 실시하였다.

i) 모형 (2.4)에서 정의된 바와 같이 오차항이 1차 자기회귀과정을 따르는 단순회귀모형하에서 자기상관모수 ρ의 값에 따라 종속변수의 값 $\{y_t; t = 1, \cdots, T\}$을 발생시킨다. 여기서, T, β, τ의 값은 임의로 고정시킨다.

ii) 3.2절에서 제안된 Gibbs 추출법에 의해 식 (3.5)로부터 베이즈요인을 근사 추정한다(여기서 사용된 Gibbs sequence는 $m = 10$이며, 반복회수(또는 Gibbs 표본크기) $N = 1000$이다. 그림 4.1과 그림 4.2에 의하면 $m = 10$과 $N = 1000$을 사용하면 3.2절에서 제안된 Gibbs 추출법이 ρ 및 β의 주변사후확률분포에 충분히 수렴함을 보여주고 있다).

iii) 추정된 베이즈요인 값이 1 이상이면 귀무가설 $H_0 : \rho = \rho_0$를 채택하고, 그 값이 1 미만이면 이를 기각한다(귀무가설 및 대립가설의 사전확률인 π_0와 π_1을 모두 0.5로 가정하였슴).

위에서 설명한 절차 i), ii) 및 iii)을 $T = 10, 30, 50$, $\beta = 5$인 경우에 대해 각각 주어진 자기상관모수 값에서 100번씩 반복실험하여(단, 사전분포의 모수는 $\alpha = 2$, $\gamma = 2$, $\mu = 5$, $\xi = 1$로 설정함), 평균 베이즈요인 값과 각 가설에 대한 검정력을 구했다. 표 4.1은 $\beta = 5$와

표 4.1: $H_0 : \rho = \rho_0$ 하에서 실제모수 ρ의 변화에 따른 평균 베이즈요인과 검정력 ($\beta = 5$)

귀무가설 (H_0)의 형 태	실제 모수	$T = 10$ 평 균 베이즈요인	$T = 10$ 검정력 (%)	$T = 30$ 평 균 베이즈요인	$T = 30$ 검정력 (%)	$T = 50$ 평 균 베이즈요인	$T = 50$ 검정력 (%)
$\rho = -0.5$	-0.9	1.07185	51	0.34469	91	0.16962	9
	-0.7	1.54726	20	2.24294	28	2.27597	43
	-0.5	1.71454	13	3.39083	4	4.40151	2
	-0.3	1.41796	25	2.50996	22	2.49449	25
	-0.1	1.13581	45	1.15941	57	0.58702	82
	0.0	0.99856	46	0.70181	76	0.23312	95
	0.1	0.87244	61	0.40263	88	0.09445	99
	0.3	0.66250	77	0.12422	99	0.02397	99
	0.5	0.51321	87	0.04030	100	0.00680	100
	0.7	0.41368	93	0.01302	100	0.00124	100
	0.9	0.35505	96	0.00397	100	0.00003	100
$\rho = 0$	-0.9	0.06720	97	0.00681	100	0.00001	100
	-0.7	0.29840	89	0.08656	98	0.00422	100
	-0.5	0.68723	74	0.48556	82	0.19690	91
	-0.3	1.31014	52	1.42370	48	1.29021	62
	-0.1	1.90363	24	2.50894	14	3.31746	12
	0.0	2.17757	17	2.79040	4	3.90232	3
	0.1	2.13718	14	2.75671	8	3.58566	9
	0.3	1.93131	25	1.90181	30	1.42752	53
	0.5	1.49032	41	0.85442	66	0.25073	94
	0.7	1.09720	56	0.30406	90	0.08637	98
	0.9	0.86375	64	0.11683	97	0.03500	99
$\rho = 0.5$	-0.9	0.0902	98	0.00017	100	0.00000	100
	-0.7	0.2005	95	0.00193	100	0.00000	100
	-0.5	0.3365	91	0.02133	100	0.00007	100
	-0.3	0.5369	84	0.12870	99	0.00290	100
	-0.1	0.7811	64	0.26517	92	0.06782	99
	0.0	0.9087	53	0.49125	83	0.20777	94
	0.1	1.0355	47	1.29115	68	0.52315	86
	0.3	1.2697	33	2.42636	29	2.18413	31
	0.5	1.5839	14	2.93768	7	4.20396	6
	0.7	1.5675	15	2.33807	10	2.57854	33
	0.9	1.5608	17	1.84315	35	0.63159	82

$\rho_0 = -.5, 0, .5$인 경우의 모의실험 결과로서 가설 (2.8)에 대한 평균 베이즈요인 값과 검정력을 나타낸다. 이 표에 나타난 베이즈요인 평균값과 검정력을 대비해 볼 때, 3.2절에서 제시된 베이즈요인 근사추정법에 의하면 일반적인 귀무가설 $H_0 : \rho = \rho_0$를 정확하게 검정할 수 있음을 알 수 있다.

또한, 그림 4.1은 자기상관모수 $\rho(\rho = 0, \pm.5)$의 주변사후확률분포를 나타내며, 이 그림에 의하면 3.2절에서 설명된 Gibbs 알고리즘이 ρ의 참값에서 단봉의 형태로 정확하게 추정함을 알 수 있다. 여기서, $\beta = 5$, $T = 30$, $\tau = 1$을 가정하였다.

이와 더불어 $\rho = 0.5$이고 회귀계수(β) 값이 각각 $4.5, 5.0, 5.5$인 모형 (2.4)에서 $T = 30$인 종속변수의 값 $\{y_t; t = 1, \cdots, T\}$들을 발생시켜서 제안된 Gibbs 알고리즘으로 β의 사후확률분포를 추정한 결과를 그림 4.2에 나타내었다. 이 그림에 의하면 3.2절에서 제안된 Gibbs 알고리즘이 β의 사후확률분포를 정확히 추정하고 있음을 나타내고 있다. 그러므로, 그림 4.1과 그림 4.2는 3.2절에서 설명한 Gibbs 알고리즘이 베이즈요인 (3.4)를 정확히 근사추정하고 있음을 나타내고 있다.

4.2. 경험적 자료분석

이 절에서는 Marr와 Quesenberry(1991)에서 사용한 자료를 가지고 3절에서 제안한 자기상관검정이 SPC분야에서 보다 나은 공정관리를 위해 유용하게 사용될 수 있음을 보이고자 한다. 이 자료는 공작기계의 마모에 따라 그것에 대한 자동조절기능을 가진 변속기 생산공정에서 사용된 어느 절단기구(cutting tool)가 폐기될 때까지 생산한 346개 변속기 부품의 생산순서 및 지름을 측정한 것이다. 이들 중에서 자동조절장치(setting) (87, 107, 112)에서 각각 생산된 부품(77개, 20개, 16개)의 자료만 이용하여서 생산된 부품의 지름과 생산순서간에 회귀모형을 다음과 같이 설정하여 분석하였다.

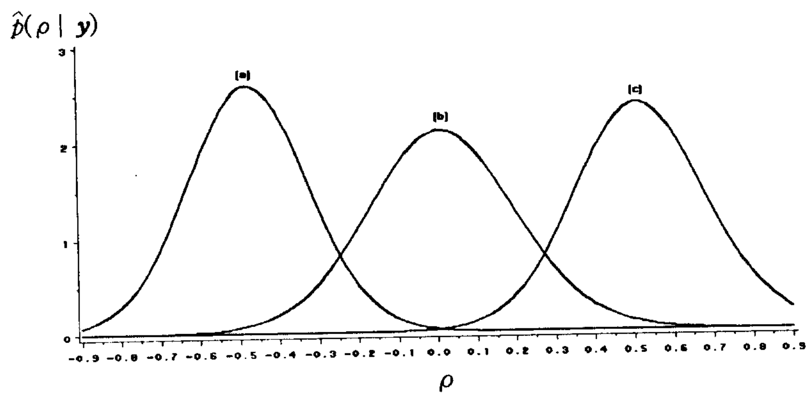

그림 4.1: Gibbs sampling 방법에 의해 추정된 자기상관모수
(a) $\rho = -0.5$ (b) $\rho = 0$ (c) $\rho - 0.5$ 에 대한 ρ의 주변사후확률분포

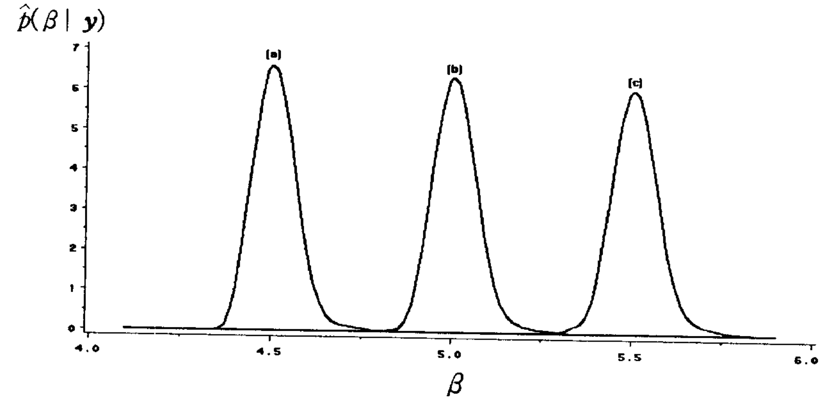

그림 4.2: Gibbs sampling 방법에 의해 추정된 회귀계수
(a) $\beta = 4.5$ (b) $\beta = 5$ (c) $\beta = 5.5$ 에 대한 β의 주변사후확률분포 ($\rho = 0.5$)

$$(\text{부품의지름}) = \alpha + \beta(\text{생산순서}) + \text{오차항}. \tag{4.1}$$

위 모형의 회귀계수 β는 절단기구의 마모도를 나타내고, 이 값은 SPC에서 절단기구의 수 명 및 자동조절장치 값을 계산하는데 사용된다.

Kass와 Raftery(1995)가 언급한 것과 같이 일반적으로 베이즈요인에 의한 검정법과 표본이론에 의한 검정법을 서로 비교하는 것은 의미가 없다. 따라서, 여기서는 자기상관검정을 위해 3절에서 제안한 베이즈요인 검정법의 유용성을 보이기 위해 사전분포 식 (2.11)의 모수를 4.1절에서 사용한 것과 동일하게 설정하여 3.2절에서 설명된 Gibbs 추출법 및 베이즈요인 근사추정에 의해 얻은 결과를 표 4.2와 표 4.3에 나타내었다. 여기서, 모수 μ의 값은 β의 최소제곱추정값을 사용하였으며, Gibbs sequence는 $m = 10$, 반복회수는 $N = 1000$으로 설정하였다.

표 4.2: 경험적 자료에 대한 Gibbs sampling 방법에 의해 추정된 베이즈요인과 Durbin-Watson 검정통계값

자동조절 setting	87	107	122
B	0.21453	0.22621	0.24218
$D-W$	1.441	1.220	1.093

B : Gibbs sampling 방법에 의해 추정된 베이즈요인
$D-W$: Durbin-Watson 검정통계값

표 4.2는 $H_0 : \rho = 0$의 검정을 위해 각 자동조절장치로부터 얻은 자료를 가지고 추정한 베이즈요인 값 및 Durbin-Watson 검정통계값들을 나타낸다. 이를 살펴보면 각 자료에 대해서 모두 베이즈요인의 값이 1보다 작게 나타나며, 따라서 자기상관계수의 값이 0이라는 귀무가설을 기각하게 됨을 알 수 있다. 이에 반하여 유의수준 $\alpha = 0.05$하에서 실시된 Durbin-Watson 검정에 의하면 장치 107과 112에 대해서 설정된 회귀모형의 자기상관은 검정불능(test fail)이었고, 장치 87인 경우에만 제안된 검정 결과와 같이 양의 자기상관이 있는 것으로 나타났다.

표 4.3은 각 자동조절장치 자료로부터 추정된 β의 추정치로써 오차항간에 자기상관이 있을 경우, 이를 감안한 추정치와 일반적인 최소자승추정치간에 유의한 차가 있음을 보여준다. 이것은 변속기 생산공정의 불량률 관리 및 공작기계 수명 계산을 위해서는 공작기계 마모도(β)의 추정을 위한 모형 (4.1)의 자기상관계수 검정이 매우 중요함을 나타낸다.

표 4.3: 경험적 자료에 대한 회귀계수(β)의 추정치 및 표준편차

자동조절 setting	87	107	122
최소자승법에 의한 추정치	0.0002958	0.0003091	0.0007471
(표준편차)	(0.00003)	(0.00018)	(0.00027)
Gibbs 추출법에 의한 추정치	0.0002783	0.0001899	0.0005204
(표준편차)	(0.00014)	(0.00033)	(0.00079)

또한, 다음에 제시된 그림 4.3에서 그림 4.5는 Gibbs 알고리즘을 통해서 추정된 ρ 및 β의 주변사후확률밀도함수(marginal posterior density function)를 각각 그림으로 나타낸 것이다. 이 그림들은 자기상관검정에 사용된 Gibbs 알고리즘이 모의실험에서와 같이 ρ와 β의 주변사후확률분포를 단봉 형태로 추정하고 있음을 보여주고 있어 ρ와 β의 점추정 뿐만 아니라 구간추정에도 이들을 유용하게 사용할 수 있음을 보여준다.

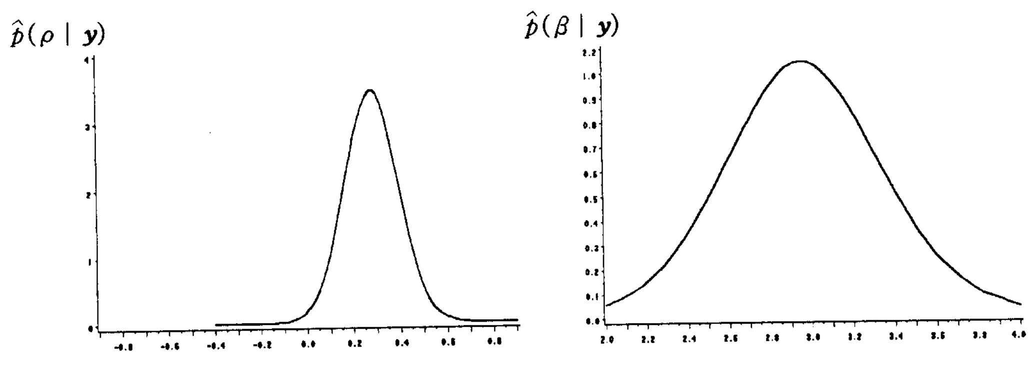

그림 4.3: ρ와 $\beta \times 10^4$의 주변사후확률분포 (장치 87)

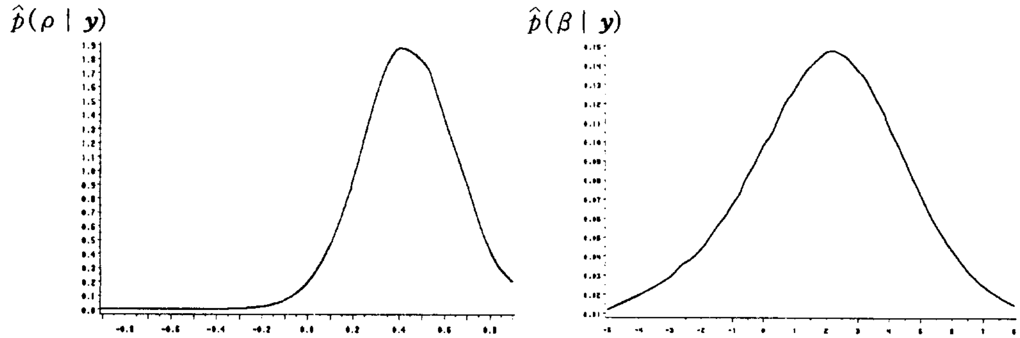

그림 4.4: ρ와 $\beta \times 10^4$의 주변사후확률분포 (장치 107)

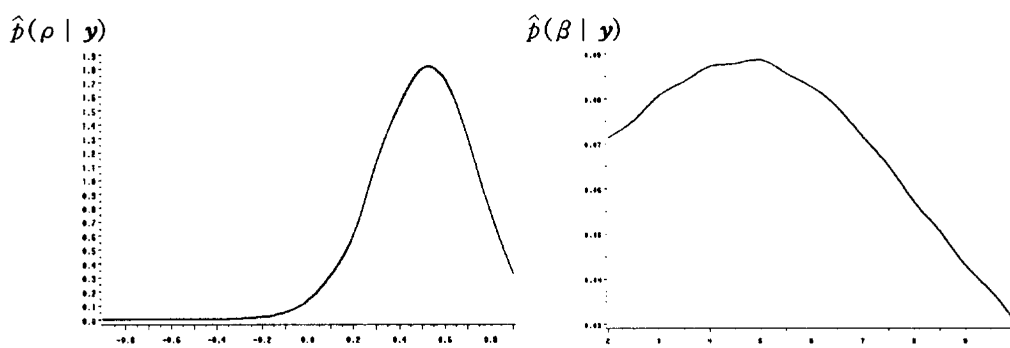

그림 4.5: ρ와 $\beta \times 10^4$의 주변사후확률분포 (장치 122)

5. 결론

본 논문은 단순회귀모형 (2.5)에서 오차항들 사이에 존재하는 자기상관모수의 존재 여부(또는 자기상관모수의 값)를 검정하는 방법으로 베이지안 추론에서 널리 사용되는 베이즈요인을 정보적 사전분포의 가정하에서 유도하였고, 이를 자기상관모수 검정법에 적용하는 방법을 제시하였다. 또한, 이와 같이 얻어진 베이즈요인을 일반화 Savage-Dickey 밀도비와 Gibbs 추출법을 합성시켜서 근사적으로 추정하기 위해 모수들 사이의 조건부 사후확률

분포들을 도출하였고, 이를 이용한 베이즈요인 근사추정 및 자기상관모수 검정법을 제안하였다.

제안된 베이지안 자기상관모수 검정법의 유용성을 검토하기 위해 시행된 Monte Carlo 모의실험을 통해 제안된 검정법의 유용성을 검토한 결과 귀무가설 $H_0 : \rho = \rho_0$의 검정에서 표 4.1에 나타나 있는 바와 같이 양호한 검정력을 보이고 있음을 알 수 있다. 또한, 검정을 위하여 제안된 Gibbs 알고리즘은 회귀계수 β 및 ρ의 주변사후확률분포를 정확하게 추정하여 줌을 보았다. 이와 더불어, 귀무가설 $H_0 : \rho = 0$하에서 제안된 검정법을 경험적 자료에 적용해 본 결과 현재 널리 사용되고 있는 Durbin-Watson 검정법에서는 검정불능인 경우를 쉽게 검정해 낼 수 있는 방법임을 알았다.

본 논문은 단순회귀모형하에서의 자기상관검정법을 제안하였으나, 도출된 결과는 쉽게 중회귀모형에 적용될 수 있도록 일반화시킬 수 있다. 한편, 본 논문에서 다룬 오차항간의 1차 자기상관검정법을 일반화시켜서 2차, 3차 및 고차의 자기상관인 경우에 대한 검정법과 1차 자기상관을 가진 경우 대립가설 H_1이 $H_1 : \rho > \rho_0$ 또는 $H_1 : \rho < \rho_0$와 같은 형태일 때의 검정법의 개발도 흥미있는 연구과제이다. 이에 대한 연구는 현재 진행 중에 있다.

참고문헌

[1] Abrahamse, A. P. J. and Koerts, J.(1971). New Estimate Disturbances in Regression Analysis, *Journal of the American Statistical Association*, vol. 66, 71-74.

[2] Alwan, L. C. and Robert, H. V.(1989). Time Series Modeling for Statistical Process Control, in *Statistical Process Control in Automated Manufacturing*, eds. J. B. Keats and N. F. Hubele, New York : Marcel Dekker, 45-65.

[3] Bromeling, L. D.(1985). *Bayesian Analysis of Linear Models*, Marcel, Dekker.

[4] Casella, G. and George, E. I.(1992). Explaining the Gibbs Sampler, *American Statistical Association*, vol. 46, No. 3, 167-174.

[5] Devroy, L.(1986). *Non-uniform Random Variate Generation*, Springer-Verlag, New York.

[6] Durbin, J. and Watson, G. S.(1951). Testing for Serial Correlation in Least Square Regression, *Biometrika*, vol. 38, 159-177.

[7] Gelfand, A. E. and Smith, A. F. M.(1990). Sampling-based Approaches to Calculating Marginal Densities, *Journal of the American Statistical Association*, vol. 85, 398-409.

[8] Gelfand, A. and Dey, D.(1994). Bayesian Model Choice : Asymptotics and Exact Calculations, *Journal of Royal Statistical Society, Ser. B*, vol. 56, 501-514.

[9] Kass, R. E. and Raftery, A. E.(1995). Bayes Factors, *Journal of the American Statistical Assocoation*, vol. 90, 773-795.

[10] Lee, P. E.(1988). *Bayesian Statistics : An Introduction*, University of York, England.

[11] Marr, R. L. and Quesenberry, C. P.(1991). A NU Test of Serial Correlation of Residuals from One or More Regression Regimes, *Technometrics*, vol. 33, 441-457.

[12] Newton, M. A. and Raftery, A. E.(1994). Approximate Bayesian inference with the Weighted Likelihood Bootstrap, *Journal of Royal Statistical Society, Ser. B*, vol. 56, 3-48.

[13] Palmer, J. L. and Broemeling, L. D.(1993). Regression Models with Autocorrelated Errors : A Gibbs Sampling Approach, *1993 ASA Proceeding of the Section on Bayesian Statistical Science*, 91-95.

[14] Pindyck, R. S. and Rubinfeld, D. L.(1981). *Econometrics Models and Economic Forecasts*, 2nd. ed., McGraw-Hill, New York.

[15] Philips, G. D. A. and Harvey, A. C.(1974). A Simple Test for Serial Correlation in Regression Analysis, *Journal of the American Statistical Association*, vol. 69, 935-939.

[16] Theil, H.(1968). A Simplication of the BLUS Procedures for Analyzing Regression Disturbances, *Journal of the American Statistical Association*, vol. 60, 1067-1079.

[17] Verdinelli, I. and Wasserman, L.(1995). Computing Bayes Factors Using a Generalization of the Savage-Dickey Density Ratio, *Journal of the American Statistical Association*, vol. 90, 614-618.

[18] Wonnacott, R. J. and Wonnacott, T. H.(1979). *Econometrics*, John & Wiley Pub. Co., New York.

[19] Zellner, A.(1971). *Bayesian Inference in Econometrics*, John Wiley & Sons, Inc, New York.

[1997년 4월 접수, 1997년 11월 최종수정]

A Bayesian Test for the First-order Autocorrelations in Regression Analysis

Hea-Jung Kim[†] Sung-Sil Han[‡]

ABSTRACT

This paper suggests a Bayesian method for testing first-order markov correlation among linear regression disturbances. As a Bayesian test criterion, Bayes factor is derived in the form of generalized Savage-Dickey density ratio that is easily estimated by means of posterior simulation via Gibbs sampling scheme. Performance of the Bayesian test is evaluated and examined based upon a Monte Carlo experiment and an empirical data analysis. Efficiency of the posterior simulation is also examined.

† Professor, Department of Statistics, Dongguk University, Seoul 100-715, Korea.
‡ Lecturer, Department of Statistics, Dongguk University, Seoul 100-715, Korea.