

## 불균형 선형혼합모형에서 추정량 \*

송석현<sup>1)</sup> 정병철<sup>2)</sup>

### 요약

본 논문에서는 오차성분이 계열상관을 갖는 불균형 랜덤모형에서 분산성분의 추정방법에 대하여 연구하였다. 분산성분에 대한 추정량으로 조건부 ANOVA(cANOVA), ML 및 REML추정량등을 유도하였으며, 계열상관값과 불균형의 정도에 따른 추정량의 변동성을 추정량의 분위수를 이용하는 EQDGs 플롯을 이용하여 비교하였다. 모의실험결과 cANOVA추정방법은 불균형의 정도에는 추정량값이 크게 영향을 받지 않는것으로 나타났으나 계열상관값의 증가에 따라서는 변동성을 보이고 있다. 불균형의 정도와 계열상관값을 동시에 고려하는 경우에는 ML추정방법이 cANOVA, REML추정방법보다 변동성이 안정적으로 나타났다.

주요용어: 선형혼합모형, cANOVA, ML, REML, EQDGs

### 1. 서론

랜덤효과와 고정효과가 혼합되어 나타나는 선형혼합모형(Linear mixed Model)에 대한 통계적인 분석은 오랫동안 많은 연구들이 진행되고 왔다 (이준영, 2000; McCulloch와 Searle, 2001). 전통적으로 선형혼합모형에서 분산성분의 추정량에 대한 비교는 추정량의 평균제곱오차(mean squared error, MSE)나 추정량의 분산-공분산행렬의 함수와 같은 단일값 기준에 의해서 이루어지곤 했다 (Ahrens, et al., 1981; Chaloner, 1987; Swallow와 Monahan, 1984). 하지만 단일 값에 의한 비교는 알려지지 않은 분산성분에 대한 사전지식을 요구한다. 이와 같은 단일값 비교의 한계를 극복하고자 Khuri(1997)는 분산성분 추정량의 분위수를 이용할것을 제안하였다. 이는 QDGs(Quantile Dispersion Graphs)라 불리는 분위수 그래프를 이용하는것으로 분산성분에 대한 사전정보를 요구하지 않는 비교방법이다. QDGs는 주어진 모수공간내에서 척도화된 추정량의 분위수의 최소값과 최대값을 그려봄으로써 추정방법들을 비교하는 방법론으로 추정량의 전체적인 분포를 이용하기 때문에 매우 신뢰성 있는 결과를 제공해 줄 수 있다. 만일 분산성분의 추정량에 대한 정확한 분포 및 분위수가 알려지지 않는 경우라면 경험적 QDGs(Empirical QDGs, EQDGs)를 사용한다(Lee와 Khuri, 1999, 2000).

이에 본 연구에서는 선형혼합모형의 한 응용분야인 불균형 랜덤모형을 고려한다. 이와 같은 불균형 반복측정자료를 분석하는데 있어서 랜덤모형은 가장 좋은 방법론을 제공해주

\* 이 논문은 2000년도 한국학술진흥재단의 지원에 의하여 연구되었음(KRF-2000-041-D00055)

1) (136-701) 서울시 성북구 안암동 5-1 고려대학교 통계학과, 부교수

E-mail: ssong@mail.korea.ac.kr

2) (136-701) 서울시 성북구 동선동 249-1 성신여자대학교 통계학과.

E-mail: bcjuwa@hitel.net

기 때문이다(Searle, et al., 1992; Diggle, et al., 1994). 특히 오차성분이 반복측정으로 인하여 자기상관되어 있는 불균형 랜덤모형을 다루려한다(Chi와 Reinsel, 1989). 이러한 모형에서 분산성분에 대한 적절한 추정방법으로 cANOVA, ML 및 REML추정방법을 제안하고자 한다. 또한 각 추정량의 성질을 최근에 제안된 그래프적 분석방법인 EQDGs를 이용하여 비교하고자 한다. 계열상관의 정도와 불균형의 정도에 따라 각각의 분산성분 추정값이 영향을 받을것이므로 계열상관의 값과 불균형의 정도에 따른 추정량의 성질들을 EQDGs 플롯을 이용하여 살펴보고자 한다.

본 논문의 구성은 다음과 같다. 먼저 2장에서는 계열상관을 포함하는 오차성분을 가지는 불균형 랜덤모형을 소개하고, 3장에서는 이러한 모형에서 분산성분들에 대한 조건부 ANOVA, ML 및 REML추정방법을 제안한다. 4장에서는 모의실험을 통하여 각 추정량들의 EQDGs를 이용하여 각 추정량들의 효율성을 비교하며, 마지막으로 5장에서는 결론을 다루었다.

## 2. 모형

다음과 같은 오차성분이 계열상관을 갖는 불균형 선형혼합모형을 고려해보자.

$$\begin{aligned} y_{it} &= \mu + u_{it} \\ &= \mu + \alpha_i + \varepsilon_{it}, \quad i = 1, \dots, N, t = 1, \dots, T_i \end{aligned} \quad (2.1)$$

여기서  $\mu$  는 미지의 모수이고,  $u_{it}$  는 전체오차항을 나타내며  $\alpha_i$  는  $i.i.d. N(0, \sigma_\alpha^2)$  을 따르는  $i$  번째 처리효과를 나타낸다. 또한  $\varepsilon_{it}$  는 정상적인 AR(1) 과정을 따르는 나머지 오차항을 나타낸다. 즉,  $\varepsilon_{it} = \rho\varepsilon_{i,t-1} + e_{it}$ ,  $|\rho| < 1$  이다.  $\rho$  는 자기상관계수를 나타내며  $e_{it}$  는  $i.i.d. N(0, \sigma_e^2)$  을 따른다고 가정한다. 이는 오차성분이 개체내에서는 자기상관되어 있음을 가정한다. 식 (2.1)을 벡터형태로 표현하면 다음과 같다.

$$\vec{y} = \mu i_n + \vec{u} = \mu i_n + \left( \bigoplus_{i=1}^N i_{T_i} \right) \vec{\alpha} + \vec{\varepsilon} \quad (2.2)$$

여기서  $n = \sum_{i=1}^N T_i$ ,  $\vec{y} = (y_{11}, \dots, y_{1T_1}, \dots, y_{N1}, \dots, y_{NT_N})$  과  $\vec{\alpha}' = (\alpha_1, \dots, \alpha_N)$  이며  $\vec{u}$  와  $\vec{\varepsilon}$  은  $\vec{y}$  와 유사하게 정의된다.  $i_n$  과  $i_{T_i}$  는 각각 차수가  $n$  과  $T_i$  인 모든 원소가 1로 구성된 벡터를 나타내며  $\bigoplus$  는 직접합(direct sum)을 나타낸다. 이와 같은 정의하에서  $\vec{y}$  의 평균은  $\mu i_n$  이며 분산-공분산행렬은 다음과 같이 구해진다.

$$\text{cov}(\vec{y}) = \Omega = \sigma_\alpha^2 \left( \bigoplus_{i=1}^N J_{T_i} \right) + \frac{\sigma_e^2}{1 - \rho^2} \left( \bigoplus_{i=1}^N V_i \right) \quad (2.3)$$

여기서  $J_{T_i}$  는 모든 원소가 1로 이루어진  $T_i \times T_i$  행렬을 나타내며  $V_i$  는 차원이  $T_i$  인 AR(1)

상관행렬로 다음과 같다.

$$V_i = \begin{bmatrix} 1 & \rho & \cdots & \rho^{T_i-2} & \rho^{T_i-1} \\ \rho & 1 & \cdots & \rho^{T_i-3} & \rho^{T_i-2} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \rho^{T_i-1} & \rho^{T_i-2} & \cdots & \rho & 1 \end{bmatrix}. \quad (2.4)$$

식 (2.4)의 AR(1) 상관행렬에 대하여 Prais와 Winsten(1954)은 다음과 같은 변환행렬  $C_i$ 를 이용하여 무상관이 되도록 하는 변환(PW변환)을 제안하였다.

$$C_i = \begin{bmatrix} (1-\rho^2)^{1/2} & 0 & 0 & \cdots & 0 & 0 & 0 \\ -\rho & 1 & 0 & \cdots & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 0 & -\rho & 1 \end{bmatrix}. \quad (2.5)$$

이는  $\bigoplus_{i=1}^N C_i$  를 모형 (2.2)의 앞에 곱해주어 PW변환된  $\bar{y}$ 는 다음과 같은 형태를 갖게 된다.

$$\bar{y}^* = \left( \bigoplus_{i=1}^N C_i \right) \bar{y} = \mu \left( \bigoplus_{i=1}^N C_i \right) i_n + (1-\rho) \left( \bigoplus_{i=1}^N i_{T_i}^\eta \right) \bar{\alpha} + \left( \bigoplus_{i=1}^N C_i \right) \bar{\varepsilon} \quad (2.6)$$

여기서  $i_{T_i}^\eta = (\eta, i'_{T_i-1})'$  이며  $\eta = \sqrt{(1+\rho)/(1-\rho)}$  이다.  $\left( \bigoplus_{i=1}^N C_i \right) E(\bar{\varepsilon}\bar{\varepsilon}') \left( \bigoplus_{i=1}^N C_i \right)' = \sigma_e^2 I_n$  이므로,  $\bar{y}^*$ 의 분산-공분산 행렬은 다음과 같이 얻어진다.

$$\Omega^* = E(\bar{y}^* \bar{y}^{*'}) = \sigma_\alpha^2 (1-\rho)^2 \left( \bigoplus_{i=1}^N i_{T_i}^\eta i_{T_i}^{\eta'} \right) + \sigma_e^2 I_n = \bigoplus_{i=1}^N \left[ d_i^2 \sigma_\alpha^2 (1-\rho)^2 \bar{J}_{T_i}^\eta \right] + \sigma_e^2 I_n \quad (2.7)$$

여기서  $d_i^2 = i_{T_i}^\eta i_{T_i}^{\eta'} = \eta^2 + T_i - 1$  이며  $\bar{J}_{T_i}^\eta = i_{T_i}^\eta i_{T_i}^{\eta'} / d_i^2$  이다. 또한  $\bigoplus_{i=1}^N I_{T_i} = I_n$  이므로,  $I_{T_i}$ 를  $E_{T_i}^\eta + \bar{J}_{T_i}^\eta$  라 하면,  $E_{T_i}^\eta = I_{T_i} - \bar{J}_{T_i}^\eta$  임을 이용하여  $\Omega^*$ 를 분광분해하면 다음과 같은 식이 얻어진다.

$$\begin{aligned} \Omega^* &= \bigoplus_{i=1}^N \left\{ \left[ d_i^2 \sigma_\alpha^2 (1-\rho)^2 + \sigma_e^2 \right] \bar{J}_{T_i}^\eta \right\} + \sigma_e^2 \bigoplus_{i=1}^N E_{T_i}^\eta \\ &= \bigoplus_{i=1}^N \lambda_i^2 \bar{J}_{T_i}^\eta + \sigma_e^2 \bigoplus_{i=1}^N E_{T_i}^\eta, \end{aligned} \quad (2.8)$$

여기서  $\lambda_i^2 = d_i^2 \sigma_\alpha^2 (1-\rho)^2 + \sigma_e^2$  이다. 식 (2.8)과 같은 분광분해는, 임의의  $p$ 에 대하여,  $\Omega^{*p}$ 의 계산을 편리하게 할 수 있다는 장점을 제공해준다. 즉,  $\Omega^{*p}$ 는 다음과 같이 표현될 수 있다.

$$\Omega^{*p} = \bigoplus_{i=1}^N \left[ (\lambda_i^2)^p \bar{J}_{T_i}^\eta \right] + (\sigma_e^2)^p \bigoplus_{i=1}^N E_{T_i}^\eta. \quad (2.9)$$

### 3. 추정방법

본 장에서는 모형(2.1)의 분산성분들에 대한 다양한 추정방법들을 유도한다.

#### 3.1. 조건부 ANOVA추정량

만일 계열상관이 존재하지 않는 경우( $\rho = 0$ )에는 모형(2.1)에서 분산성분에 대한 ANOVA 형태의 추정량은 이차형식의 제곱합을 이용하여 얻을 수 있다(Searle, et al., 1992). 반면 나머지 오차항이 서로 상관된 경우라면 ( $\rho \neq 0$ ) ANOVA-형태의 분산성분에 대한 추정량은 오차항들의 상관관계로 인하여 추정에 어려움이 있다. 이를 극복하기 위하여 우선  $\rho$ 에 대한 추정을 한후 추정된  $\rho$ 를 이용하여 분산성분을 추정하는 cANOVA 추정량의 이용을 제안하고자 한다.

먼저,  $\rho$ 는 집단내 제곱합에 근거하여 Baltagi와 Li(1991)에 의해서 제안된 추정을 이용한다.

$$\tilde{\rho} = \frac{\sum_{i=1}^N \sum_{t=2}^{T_i} (y_{it} - \bar{y}_i)(y_{i,t-1} - \bar{y}_i)}{\sum_{i=1}^N \sum_{t=2}^{T_i} (y_{i,t-1} - \bar{y}_i)^2} \quad (3.1)$$

여기서  $\bar{y}_i = \sum_{t=1}^{T_i} y_{it}/T_i$  이다. 다음으로  $\tilde{y}^*$ 에 대한 집단내 제곱합  $\tilde{S}_1$  과 집단간 제곱합  $\tilde{S}_2$  을 다음과 같이 정의하자.

$$\begin{aligned} \tilde{S}_1 &= \tilde{y}^{*'} Q^* \tilde{y}^*, \\ \tilde{S}_2 &= \tilde{y}^{*'} (P^* - \bar{J}_n^*) \tilde{y}^* \end{aligned} \quad (3.2)$$

여기서  $Q^* = \bigoplus_{i=1}^N E_{T_i}^{\tilde{\eta}}$ ,  $P^* = \bigoplus_{i=1}^N \bar{J}_{T_i}^{\tilde{\eta}}$ ,  $\bar{J}_n^* = \tilde{\eta}' \tilde{\eta} / \sum_{i=1}^N \tilde{d}_i^2$ 이며,  $\tilde{\eta} = (\tilde{\eta}_{T_1}, \dots, \tilde{\eta}_{T_N})'$  과  $\tilde{\eta} = \sqrt{(1+\tilde{\rho})/(1-\tilde{\rho})}$  이다. 식 (3.2)에 정의된 2개의 제곱합에 대한 기대값은  $\tilde{\rho}$ 이 주어졌다는 조건하에서 다음과 같이 구해진다.

$$\begin{aligned} E[\tilde{S}_1] &= \sigma_e^2 \text{tr}[Q^*] = \sigma_e^2(n-N), \\ E[\tilde{S}_2] &= \text{tr}[\tilde{\lambda}_i^2 \bar{J}_{T_i}^{\tilde{\eta}}] - \text{tr}\left[\left(\bigoplus_{i=1}^N \tilde{\lambda}_i^2 \bar{J}_{T_i}^{\tilde{\eta}}\right) \bar{J}_n^*\right] \\ &= \sigma_\alpha^2(1-\tilde{\rho})^2 \left( \sum_{i=1}^N \tilde{d}_i^2 - \sum_{i=1}^N \tilde{d}_i^4 / \sum_{i=1}^N \tilde{d}_i^2 \right) + (N-1)\sigma_e^2. \end{aligned} \quad (3.3)$$

그러므로  $\sigma_\alpha^2$ 와  $\sigma_e^2$ 에 대한 조건부 ANOVA-형태의 추정량은  $\rho$ 가 주어졌다는 조건하에서  $\tilde{S}_1$  과  $\tilde{S}_2$  를 그들의 기대값으로 풀게되면 얻을 수 있다.

3.2. 최대가능도 추정량(ML추정량)

최대가능도 추정방법은 분산성분과 계열상관계수를 동시에 추정한다. 먼저  $\phi = \sigma_\alpha^2/\sigma_e^2$  이라 정의하면  $\vec{y}^*$ 에 대한 척도화된 공분산행렬은 다음과 같이 구해지게 된다.

$$\Sigma^* = \Omega^*/\sigma_e^2 = \bigoplus_{i=1}^N \theta_i^2 \bar{J}_{T_i}^\alpha + \bigoplus_{i=1}^N E_{T_i}^\alpha \tag{3.4}$$

여기서  $\theta_i^2 = \lambda_i^2/\sigma_e^2 = d_i^2(1-\rho^2)\phi + 1$  이다.  $|\Sigma^*| = \prod_{i=1}^N \theta_i^2$ 이고  $|\bigoplus_{i=1}^N C_i| = (1-\rho^2)^{N/2}$  이므로  $\vec{y}$ 에 대한 척도화된 공분산행렬의 행렬식은 다음과 같다.

$$|\Sigma| = \left| \bigoplus_{i=1}^N C_i^{-1} \right| \cdot |\Sigma^*| \cdot \left| \bigoplus_{i=1}^N C_i \right| = \prod_{i=1}^N \theta_i^2 (1-\rho^2)^{-N}. \tag{3.5}$$

그러므로  $\vec{y}$ 에 대한 로그우도함수는 식 (3.5)를 이용하면 다음과 같이 얻어진다.

$$L(\mu, \phi, \rho, \sigma_e^2) = Const. - \frac{n}{2} \log \sigma_e^2 + \frac{N}{2} \log(1-\rho^2) - \frac{1}{2} \sum_{i=1}^N \log \theta_i^2 - \frac{1}{2\sigma_e^2} \vec{u}' \Sigma^{-1} \vec{u}. \tag{3.6}$$

식 (3.6)의 로그우도함수에 대하여  $\rho$  와  $\phi$ 가 주어졌을 때  $\mu$  와  $\sigma_e^2$ 에 대한 일차편미분은 폐쇄형의 해를 제공해준다. 하지만  $\phi$  와  $\rho$ 에 대한 일차편미분은  $\mu$  와  $\sigma_e^2$  이 알려졌다고 하더라도 비선형의 형태를 보이게 된다. 그러므로  $\phi$  와  $\rho$ 에 대해서는 반복에 의한 수치적인 해를 구해야 한다. 이때 필요한  $\phi$  와  $\rho$  에 대한 일차편미분 값은 다음과 같이 얻어진다.

$$\begin{aligned} \frac{\partial L}{\partial \phi} &= -\frac{1}{2} \left[ \sum_{i=1}^N d_i^2(1-\rho)^2 - \sum_{i=1}^N \frac{d_i^4(1-\rho)^4 \phi}{\lambda_i^2} \right] + \frac{1}{2\sigma_e^2} \vec{u}' \left[ \Sigma^{-1} \left( \bigoplus_{i=1}^N J_{T_i} \right) \Sigma^{-1} \right] \vec{u}, \\ \frac{\partial L}{\partial \rho} &= -\frac{1}{2(1-\rho^2)} \left[ 2\rho n + \sum_{i=1}^N \text{tr}(\Sigma_i^{-1} F_i) - 2\rho(1-\rho)^2 \phi \sum_{i=1}^N \frac{d_i^2}{\theta_i^2} - \phi \sum_{i=1}^N i'_{T_i} \Sigma_i^{-1} F_i \Sigma_i^{-1} i_{T_i} \right] \\ &\quad + \frac{1}{2(1-\rho^2)\sigma_e^2} \vec{u}' \left[ \Sigma^{-1} \left( 2\rho \bigoplus_{i=1}^N \Sigma_i + \bigoplus_{i=1}^N F_i \right) \Sigma^{-1} \right] \vec{u} \end{aligned} \tag{3.7}$$

여기서

$$F_i = \frac{\partial V_i}{\partial \rho} = \begin{bmatrix} 0 & 1 & 2\rho & \cdots & (T_i-1)\rho^{T_i-2} \\ 1 & 0 & 1 & \cdots & (T_i-2)\rho^{T_i-3} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ (T_i-1)\rho^{T_i-2} & (T_i-2)\rho^{T_i-3} & (T_i-3)\rho^{T_i-4} & \cdots & 0 \end{bmatrix} \text{ 이다.}$$

또한  $\phi$  에  $\rho$  대한 정보행렬은 다음과 같다.

$$\begin{aligned}
 E\left[-\frac{\partial^2 L}{\partial \phi^2}\right] &= \frac{(1-\rho)^4}{2} \sum_{i=1}^N \frac{d_i^4}{\theta_i^4}, \\
 E\left[-\frac{\partial^2 L}{\partial \rho^2}\right] &= \frac{1}{2(1-\rho^2)^2} \left[ 4\rho^2 n + \sum_{i=1}^N \text{tr}(\Sigma_i^{-1} F_i \Sigma_i^{-1} F_i) + 4\rho^2 \phi^2 \sum_{i=1}^N \frac{d_i^4 (1-\rho)^4}{\theta_i^4} \right. \\
 &\quad + \phi^2 \sum_{i=1}^N \frac{1}{\theta_i^4} (i'_{T_i} \Sigma_i^{-1} F_i \Sigma_i^{-1} i_{T_i})^2 + 4\rho \sum_{i=1}^N \text{tr}(\Sigma_i^{-1} F_i) - 8\rho^2 \phi \sum_{i=1}^N \frac{d_i^2 (1-\rho)^2}{\theta_i^2} \\
 &\quad - 8\rho \phi \sum_{i=1}^N \frac{1}{\theta_i^2} (i'_{T_i} \Sigma_i^{-1} F_i \Sigma_i^{-1} i_{T_i}) - 2\phi \sum_{i=1}^N \frac{1}{\theta_i^2} (i'_{T_i} \Sigma_i^{-1} F_i \Sigma_i^{-1} F_i \Sigma_i^{-1} i_{T_i}) \\
 &\quad \left. + 4\rho \phi^2 \sum_{i=1}^N \frac{d_i^2 (1-\rho)^2}{\theta_i^4} (i'_{T_i} \Sigma_i^{-1} F_i \Sigma_i^{-1} i_{T_i}) \right], \\
 E\left[-\frac{\partial^2 L}{\partial \phi \partial \rho}\right] &= \frac{1}{2(1-\rho^2)} \left[ 2\rho(1-\rho)^2 \sum_{i=1}^N \frac{d_i^2}{\theta_i^4} + \sum_{i=1}^N \frac{1}{\theta_i^4} (i'_{T_i} \Sigma_i^{-1} F_i \Sigma_i^{-1} i_{T_i}) \right]. \quad (3.8)
 \end{aligned}$$

$\phi$ 와  $\rho$ 에 대한 ML추정량은 식 (3.7)과 (3.8)을 이용하는 Fisher의 스코어링 방법을 사용한다. 즉, 초기값에서 출발하여  $(r+1)$ 번째 갱신되는  $\hat{\phi}$  과  $\hat{\rho}$  는 다음과 같은 식에 의해서 얻어진다.

$$\begin{bmatrix} \hat{\phi} \\ \hat{\rho} \end{bmatrix}_{r+1} = \begin{bmatrix} \hat{\phi} \\ \hat{\rho} \end{bmatrix}_r + \begin{bmatrix} E[-\frac{\partial^2 L}{\partial \phi^2}] & E[-\frac{\partial^2 L}{\partial \phi \partial \rho}] \\ E[-\frac{\partial^2 L}{\partial \phi \partial \rho}] & E[-\frac{\partial^2 L}{\partial \rho^2}] \end{bmatrix}_r^{-1} \begin{bmatrix} \frac{\partial L}{\partial \phi} \\ \frac{\partial L}{\partial \rho} \end{bmatrix}_r. \quad (3.9)$$

### 3.3. 제한적 최대가능도 추정량(REML추정량)

모형(2.1)에서 분산성분에 대한 ML추정량은 고정된 모수의 추정으로 인하여 발생하는 자유도의 손실을 보정하지 않는다는 관점에서 비판을 받아왔다. 이에 대한 대안으로 Patterson과 Thompson(1971)은 비정칙변환(singular transformation)을 이용한 제한적 최대가능도 추정법(restricted ML추정법, REML추정법)을 제안하였다. 이는 가능도함수를 고정된 모수에 의존하는 부분과 의존하지 않는 부분 등 두 부분으로 나누어주는 변환을 이용한 추정방법이다. 즉, REML은 원자료에 대하여  $y'[A : \Sigma^{-1} i_n / \sigma_e^2]$  이라는 변환에 근거한 추정법으로, 여기서 행렬  $A$ 는  $A'A = I_n - i_n(i_n' i_n)^{-1} i_n' = I_n - \bar{J}_n$  과  $AA' = I_n$  을 만족하는 계수  $(n-1) \times n$  인 행렬이다. 이 때  $A'y$  과  $i_n' \Sigma^{-1} y / \sigma_e^2$  는 서로 독립이 되며 다음과 같은 분포를 따르게 된다.

$$\begin{bmatrix} A' \\ i_n' \Sigma^{-1} / \sigma_e^2 \end{bmatrix} \bar{y} \sim N \left( \begin{bmatrix} 0 \\ i_n' \Sigma^{-1} i_n \mu / \sigma_e^2 \end{bmatrix}, \begin{bmatrix} \sigma_e^2 A \Sigma A' & 0 \\ 0 & i_n' \Sigma^{-1} i_n / \sigma_e^2 \end{bmatrix} \right). \quad (3.10)$$

식(3.10)에서  $A'y$ 와  $i'_n \Sigma^{-1} y / \sigma_e^2$ 의 로그우도함수를 각각  $\log L_1$ 과  $\log L_2$ 라 하면  $\log L_1$ 은 고정된 모수에 의존하지 않게 된다. 그러므로 이 부분을 최대화하면 분산성분과 계열상관계수에 대한 REML추정량을 얻게 된다. Hocking(1985)에 의하면  $A'(A\Sigma A')^{-1}A = \Sigma^{-1}[I - i_n(i'_n \Sigma^{-1} i_n)^{-1} i'_n \Sigma^{-1}] = \Sigma^{-1}(I - M)$ , 여기서  $M = i_n(i'_n \Sigma^{-1} i_n)^{-1} i'_n \Sigma^{-1}$  라는 사실을 얻을 수 있으므로  $\log L_1$ 에 대한  $\sigma_e^2$ ,  $\phi$ 와  $\rho$ 에 대한 일차미분값은 다음과 같이 구해진다.

$$\begin{aligned} \frac{\partial A'y}{\partial \sigma_e^2} &= -\frac{n-1}{2\sigma_e^2} + \frac{1}{2\sigma_e^4} \bar{y}' \Sigma^{-1} (I - M) \bar{y}, \\ \frac{\partial A'y}{\partial \phi} &= -\frac{1}{2} \text{tr} \left[ \left( \bigoplus_{i=1}^N i_{T_i} \right)' A' (A\Sigma A')^{-1} A \left( \bigoplus_{i=1}^N i_{T_i} \right) \right] \\ &\quad + \frac{1}{2\sigma_e^2} \bar{y}' \left[ A' (A\Sigma A')^{-1} A \left( \bigoplus_{i=1}^N J_{T_i} \right) A' (A\Sigma A')^{-1} A \right] \bar{y}, \\ \frac{\partial A'y}{\partial \rho} &= -\frac{1}{2} \text{tr} \left[ A' (A\Sigma A')^{-1} A \left( \frac{\partial \Sigma}{\partial \rho} \right) \right] \\ &\quad + \frac{1}{2\sigma_e^2} \bar{y}' \left[ A' (A\Sigma A')^{-1} A \left( \frac{\partial \Sigma}{\partial \rho} \right) A' (A\Sigma A')^{-1} A \right] \bar{y} \end{aligned} \quad (3.11)$$

식 (3.11)를 이용하면  $\sigma_e^2$ 에 대한 REML추정량을 구할 수 있다. 반면  $\phi$ 와  $\rho$ 에 대한 REML추정량을 구하기 위해서는 수치적인 방법으로 Fisher의 스코어링방법을 사용하였으며 이때 필요한 정보행렬은 다음과 같이 얻어진다.

$$\begin{aligned} E \left[ -\frac{\partial^2 A'y}{\partial \phi^2} \right] &= \frac{1}{2} \text{tr} \left[ \left( \bigoplus_{i=1}^N J_{T_i} \right) \Sigma^{-1} (I - M) \left( \bigoplus_{i=1}^N J_{T_i} \right) \Sigma^{-1} (I - M) \right], \\ E \left[ -\frac{\partial^2 A'y}{\partial \rho^2} \right] &= \frac{1}{2(1-\rho^2)^2} \text{tr} \left[ \left( 2\rho \bigoplus_{i=1}^N \Sigma_i + \bigoplus_{i=1}^N F_i \right) \Sigma^{-1} (I - M) \right. \\ &\quad \left. \cdot \left( 2\rho \bigoplus_{i=1}^N \Sigma_i + \bigoplus_{i=1}^N F_i \right) \Sigma^{-1} (I - M) \right], \\ E \left[ -\frac{\partial^2 A'y}{\partial \phi \partial \rho} \right] &= \frac{1}{2(1-\rho^2)} \cdot \\ &\quad \text{tr} \left[ \left( \bigoplus_{i=1}^N J_{T_i} \right) \Sigma^{-1} (I - M) \left( 2\rho \bigoplus_{i=1}^N \Sigma_i + \bigoplus_{i=1}^N F_i \right) \Sigma^{-1} (I - M) \right]. \end{aligned} \quad (3.12)$$

#### 4. 모의실험에 의한 추정량 비교

##### 4.1. 모의실험 방법

3장에서 유도한 분산성분에 대한 추정방법들의 효율성을 비교하기 위하여 모의실험을 실시하였다. 모형 (2.1)에서 먼저  $\mu = 5$ 라고 고정하고,  $\phi = \sigma_\alpha^2 / \sigma_e^2$ 라 놓는다면 모의실험모

형은 다음과 같다.

$$y_{it} = 5 + \alpha_i + \varepsilon_{it}, \quad i = 1, \dots, N, \quad t = 1, \dots, T_i \quad (4.1)$$

여기서  $\alpha_i \sim iid N(0, \phi\sigma_e^2)$ 이고  $\varepsilon_{it} = \rho\varepsilon_{i,t-1} + e_{it}$ 이며  $|\rho| < 1$ 이고  $e_{it} \sim iid N(0, \sigma_e^2)$ 이다.

모형 (2.1)에 대한 분산성분의 추정량은 상관계수의 값과 불균형의 정도에 의존하게 된다. 추정방법들의 효율성을 비교하기 위하여 균형된 디자인(balanced design, bal), 약간 불균형된 디자인(moderately unbalanced design, mu), 아주 심하게 불균형된 디자인(severely unbalanced design, su) 등 3가지 형태의 디자인이 사용되었으며 이들은 다음과 같은 형태를 갖는다.

$$\begin{aligned} D_{bal} &= \{10, 10, 10, 10, 10\}, \\ D_{mu} &= \{6, 6, 6, 16, 16\}, \\ D_{su} &= \{3, 3, 3, 3, 38\}. \end{aligned} \quad (4.2)$$

식 (4.2)에서 고려된 3가지 디자인은 모두 그룹의 수  $N = 5$ 이며 총 표본수  $n = 50$ 으로 고정되었으므로 각 디자인과 추정방법에 대한 비교가 가능할 수 있다.

모형 (4.1)에서  $\phi$ ,  $\rho$ 와  $\sigma_e^2$ 에 대한 모수공간  $U$ 는 다음과 같이 선택되었다.

$$U = \{(\phi, \rho, \sigma_e^2) \mid 0.5 \leq \phi \leq 10, \quad 0 < \rho < 1, \quad 0.2 \leq \sigma_e^2 \leq 2\}. \quad (4.3)$$

여기에서  $\rho$ 는 현실적으로 대부분이 양의 자기상관을 나타내므로 양의 값으로 제한하였다. 또한 이와 같은 모수공간에서 모의실험을 위하여  $(\phi, \rho, \sigma_e^2)$ 의 값들은 다음과 같이 선택하였으며 선택된 공간은 모두 600개의 점으로 이루어졌다.

$$T = \{(\phi, \rho, \sigma_e^2) \mid \phi = 0.5(0.5)10, \quad \rho = 0.2(0.3)0.8, \quad \sigma_e^2 = 0.2(0.2)2.0\}. \quad (4.4)$$

$\phi$ ,  $\rho$ 와  $\sigma_e^2$ 에 대한 cANOVA, ML 및 REML 추정량을 구하기 위하여 모형 (4.1)에서 주어진 공간  $T$ 의 각  $(\phi, \rho, \sigma_e^2)$ 값과 주어진 각 디자인  $D_i$ ,  $i = bal, mu, su$ 에서 랜덤벡터  $\vec{y}$ 가 발생되었다. 모수  $\phi$ 와  $\rho$ 는 척도에 의존하므로 비교를 위하여 척도화된 모수의 추정값  $\widehat{W}_\phi = \widehat{\phi}/\phi$ 이 주어진 디자인  $D_i$ ,  $i = bal, mu, su$ 와 각 추정방법  $m$ ,  $m = cANOVA, ML, REML$ 에서 계산되었다.

총 1000번의 반복에 의해서  $\widehat{W}_\phi$  값이 실험공간  $T$ 에서 만들어졌으며 이들을 통해 경험적인 분위수가 얻어졌다. 먼저 주어진 디자인  $D$ 와 추정방법  $m$ 에서  $\widehat{W}_\phi$ 의  $p^{th}$  경험적 분위수를  $\widehat{Q}_\phi^m(p, D, \phi, \rho, \sigma_e^2)$ 라 정의하면 실험공간  $T$ 에서  $p^{th}$  경험적 분위수의 최대값과 최소값은 다음과 같이 정의된다.

$$\begin{aligned} \widehat{Q}_\phi^m(p, D) &= \max_{(\phi, \rho, \sigma_e^2) \in T} \widehat{q}_\phi(p, D, \phi, \rho, \sigma_e^2), \\ \widehat{R}_\phi^m(p, D) &= \min_{(\phi, \rho, \sigma_e^2) \in T} \widehat{q}_\phi(p, D, \phi, \rho, \sigma_e^2) \end{aligned}$$

여기서  $D = D_{bal}, D_{mu}, D_{su}$  이고  $m = cANOVA, ML, REML$  이다. 척도화된 추정량  $\widehat{W}_\phi$ 에 대한 EQDGs는  $\widehat{Q}_\phi^m(p, D)$ 과  $\widehat{R}_\phi^m(p, D)$ 를  $p$ 에 대하여 플롯함으로써 얻어진다.



#### 4.2. 모의실험의 결과

모형 (2.1)에서  $\rho$ 는 장애모수로 취급되며 본 연구에서는  $\rho$ 의 각 수준에서  $\widehat{W}_\phi$ 가 어떻게 변화하는가를 알아보는 것이 첫 번째 목적이므로  $\rho$ 에 대한 EQDGs는 고려하지 않고  $\rho$ 의 값에 따라 각각의 디자인 및 추정방법들에 의하여 얻어진  $\widehat{W}_\phi$ 의 최대값과 최소값인  $\widehat{Q}_\phi^m(p, D)$ 과  $\widehat{R}_\phi^m(p, D)$ 은 표 4.1-4.3에 주어졌으며,  $\rho$ 의 각 수준 변화에 따른 세 가지 추정방법의 효율성을 비교하기 위한 EQDGs는 그림 4.1-4.3에 주어졌다.

표 4.1-4.3과 그림 4.1-4.3을 살펴보면 다음과 같은 결과들을 알 수 있다. 먼저 고려된 계열상관계수의 값 ( $\rho = 0.2$ : 그림 4.1,  $\rho = 0.5$ : 그림 4.2,  $\rho = 0.8$ : 그림 4.3)과 세 가지 불균형 디자인((a)  $D_{bal}$ , (b)  $D_{mu}$ , (c)  $D_{su}$ )에서  $\widehat{W}_\phi$ 에 대한 ML추정량이 추정량의 분위수에서 가장 작은 변동을 보여주었고, cANOVA추정량의 변동이 가장 크게 나타남을 알 수 있다. 반면 REML추정량은 분위수의 변동에 있어서 ML추정량보다는 변동이 크고 cANOVA추정량보다는 변동이 작은 결과를 보여주었다. 이와 같은 결과는 계열상관계수의 값이 0.8인 경우에는 변동에 대한 차이가 좀 더 크게 나타나고 있다. 또한 그림 4.3에서의 분위수 플롯은 그림 4.1과 그림 4.2에 나타난 분위수 플롯보다 최소값 및 최대값의 범위가 몹시 커지고 있음을 알 수 있다. 이는 계열상관계수의 값에 분산성분의 추정이 영향을 받는 것으로 계열상관의 값이 커질수록 분산성분 추정값에 대한 변동이 커지게 됨을 알 수 있다. 즉 계열상관값 및 디자인의 불균형 정도를 동시에 고려하는 경우에는 ML추정량이 cANOVA, REML추정량 보다 안정적인 변동성을 제공해주는 추정방법으로 나타났다.

이와 같은 결과와 더불어  $\sigma_\alpha^2$ 의 추정에 불균형 디자인이 어떠한 영향을 주는가를 3가지 추정방법의  $\widehat{W}_\phi$ 에 대한 EQDGs를 통하여 알아보았다. 그림 4.4에서 그림 4.6은 이와 같은 결과들을 나타내 주고 있다. 먼저 그림 4.4(a), 그림 4.5(a) 및 그림 4.6(a)는 cANOVA추정량에 대한 EQDGs를 나타내 준다. 이 결과를 보면 cANOVA추정량의 분위수 변동에 있어서는 주어진 계열상관값에서는 디자인의 불균형 정도에 대하여 변동성에 대하여 큰 차이를 보이지 않고 있다. 즉, cANOVA추정량은 적어도 디자인의 불균형 정도에 대하여 추정결과가 큰 영향을 받지 않는 것으로 판단된다. 반면 ML추정방법과 REML추정방법에 있어서는  $\rho$ 의 값에 상관없이 균형된 디자인과 약간 불균형된 디자인에 있어서는 큰 차이가 발견할 수 없었다. 그러나 아주 심하게 불균형된 디자인( $D_{su}$ )의 경우에는 분위수 값의 변동의 정도가 다른 두 가지 형태의 디자인에 비하여 크게 나타나고 있음을 알 수 있다. 이와 같은 사실로부터 디자인이 아주 심하게 불균형 되지 않는 이상 ML과 REML추정방법은 계열상관의 값과 디자인의 불균형에 크게 영향을 받지 않는다는 것을 알 수 있다(그림 4.4(b,c), 그림 4.5(b,c) 및 그림 4.6(b,c) 참조).

표 4.1:  $D_{bal}$ 인 경우  $\rho$ 의 수준에 대한 각 분위수에서  $\widehat{W}_\rho$ 의 최대값과 최소값

$\rho$	$p$	cANOVA		ML		REML	
		minimum	maximum	minimum	maximum	minimum	maximum
0.2	0.01	-0.2132	0.0880	0.0000	0.0638	0.0000	0.0832
	0.05	-0.0483	0.2014	0.0000	0.1520	0.0000	0.1948
	0.10	0.0629	0.2928	0.0000	0.2292	0.0000	0.2879
	0.20	0.2420	0.4408	0.0493	0.3449	0.1444	0.4357
	0.30	0.4409	0.5973	0.1912	0.4712	0.3181	0.5925
	0.40	0.5948	0.7395	0.3610	0.5871	0.5301	0.7423
	0.50	0.7758	0.9160	0.5249	0.7126	0.7301	0.9030
	0.60	0.9320	1.1779	0.7013	0.8659	0.9295	1.0854
	0.70	1.1382	1.4277	0.9053	1.0496	1.1369	1.3393
	0.80	1.3812	1.8599	1.0983	1.3599	1.3771	1.7679
0.5	0.90	1.7840	2.3627	1.4146	1.7685	1.7731	2.2884
	0.95	2.1688	2.9946	1.6956	2.2622	2.1298	2.8974
	0.99	2.8813	4.6298	2.2763	3.5397	2.8544	4.4871
	0.01	-0.5102	0.0760	0.0000	0.0213	0.0000	0.0371
	0.05	-0.1989	0.1976	0.0000	0.1103	0.0000	0.1549
	0.10	0.0031	0.3096	0.0000	0.1911	0.0000	0.2567
	0.20	0.2508	0.4405	0.0000	0.3174	0.0000	0.4057
	0.30	0.4835	0.5989	0.0000	0.4429	0.0503	0.5632
	0.40	0.6346	0.7969	0.1323	0.5710	0.3312	0.7236
	0.50	0.7914	1.1117	0.3554	0.7067	0.6224	0.8947
0.8	0.60	0.9584	1.3912	0.5930	0.8466	0.8909	1.0707
	0.70	1.1590	1.7058	0.8388	1.0400	1.1400	1.3568
	0.80	1.4306	2.1587	1.1101	1.3628	1.4081	1.8832
	0.90	1.8459	2.9931	1.4120	1.9467	1.7956	2.6302
	0.95	2.2635	3.7739	1.7838	2.6091	2.2424	3.4281
	0.99	2.9790	5.5696	2.3095	4.1651	2.8971	5.3666
	0.01	-1.4656	0.0673	0.0000	0.0000	0.0000	0.0000
	0.05	-0.3831	0.2016	0.0000	0.0000	0.0000	0.0000
	0.10	0.0876	0.3124	0.0000	0.0000	0.0000	0.0000
	0.20	0.3917	0.9001	0.0000	0.0907	0.0000	0.1748
0.30	0.5491	1.4160	0.0000	0.2695	0.0000	0.3752	
0.40	0.6903	2.0482	0.0000	0.4140	0.0000	0.5751	
0.50	0.8554	2.7633	0.0000	0.5640	0.0916	0.7578	
0.60	1.0634	3.4232	0.3358	0.7584	0.7702	1.3133	
0.70	1.3097	4.2755	0.8032	1.3682	1.0488	2.2128	
0.80	1.5829	5.2985	1.0570	2.3914	1.3510	3.4651	
0.90	2.0613	7.1604	1.4474	3.9904	1.8291	5.2928	
0.95	2.4916	8.9317	1.7821	5.2290	2.2807	6.8776	
0.99	3.2171	13.5505	2.4364	8.6800	3.0839	10.9178	

표 4.2:  $D_{mu}$ 인 경우  $\rho$ 의 수준에 대한 각 분위수에서  $\widehat{W}_\phi$ 의 최대값과 최소값

$\rho$	$p$	cANOVA		ML		REML	
		minimum	maximum	minimum	maximum	minimum	maximum
0.2	0.01	-0.2187	0.0859	0.0000	0.0669	0.0000	0.0866
	0.05	-0.0585	0.1868	0.0000	0.1531	0.0000	0.1942
	0.10	0.0506	0.2715	0.0000	0.2310	0.0000	0.2919
	0.20	0.2391	0.4443	0.0000	0.3604	0.1276	0.4523
	0.30	0.4227	0.5770	0.1572	0.4839	0.3300	0.6123
	0.40	0.5940	0.7216	0.3126	0.5912	0.5137	0.7454
	0.50	0.7397	0.9272	0.4996	0.7068	0.7339	0.9175
	0.60	0.9088	1.0957	0.6584	0.8616	0.9093	1.0981
	0.70	1.0920	1.3869	0.8790	1.0391	1.1252	1.3432
	0.80	1.3658	1.8170	1.0974	1.2939	1.3791	1.7099
	0.90	1.8830	2.4775	1.4608	1.8131	1.8310	2.3702
0.95	2.3339	3.1982	1.7542	2.2892	2.1963	2.9301	
0.99	3.0846	4.5601	2.3359	3.5772	2.9226	4.5855	
0.5	0.01	-0.5926	0.0756	0.0000	0.0000	0.0000	0.0359
	0.05	-0.2239	0.1772	0.0000	0.1115	0.0000	0.1555
	0.10	-0.0456	0.2605	0.0000	0.1892	0.0000	0.2472
	0.20	0.2002	0.4247	0.0000	0.3150	0.0000	0.4143
	0.30	0.4173	0.5992	0.0000	0.4411	0.0299	0.5662
	0.40	0.5975	0.7561	0.0396	0.5641	0.3019	0.7139
	0.50	0.7565	1.0608	0.2978	0.6986	0.6129	0.9107
	0.60	0.9387	1.3855	0.5449	0.8616	0.9157	1.1095
	0.70	1.1343	1.7643	0.8688	1.0616	1.1472	1.4266
	0.80	1.4460	2.2703	1.1123	1.4004	1.4078	1.9372
	0.90	1.9113	3.0135	1.4658	2.0730	1.8460	2.7552
0.95	2.3278	3.8851	1.7676	2.6720	2.2487	3.5274	
0.99	3.1860	5.8353	2.3621	4.0481	2.9675	5.2239	
0.8	0.01	-2.5656	0.0556	0.0000	0.0000	0.0000	0.0000
	0.05	-0.7505	0.1661	0.0000	0.0000	0.0000	0.0000
	0.10	-0.1869	0.2854	0.0000	0.0000	0.0000	0.0000
	0.20	0.3460	0.6262	0.0000	0.1024	0.0000	0.2051
	0.30	0.5181	1.2050	0.0000	0.2792	0.0000	0.4097
	0.40	0.6664	1.7506	0.0000	0.4301	0.0000	0.6026
	0.50	0.8248	2.3638	0.0000	0.5816	0.0750	0.7748
	0.60	1.0225	2.9997	0.2429	0.7566	0.8032	1.5123
	0.70	1.2359	3.8771	0.8199	1.4292	1.1002	2.2342
	0.80	1.5498	5.1321	1.0812	2.3228	1.3875	3.5502
	0.90	1.9967	6.9929	1.4475	4.0522	1.8516	5.6626
0.95	2.4796	9.1095	1.7921	5.5360	2.2611	7.3768	
0.99	3.3094	13.3061	2.4905	8.6413	3.1452	11.6219	

표 4.3:  $D_{su}$  인 경우  $\rho$ 의 수준에 대한 각 분위수에서  $\widehat{W}_\phi$ 의 최대값과 최소값

$\rho$	$p$	cANOVA		ML		REML	
		minimum	maximum	minimum	maximum	minimum	maximum
0.2	0.01	-0.5821	0.0611	0.0000	0.0000	0.0000	0.0598
	0.05	-0.2811	0.1530	0.0000	0.1100	0.0000	0.1727
	0.10	-0.1357	0.2358	0.0000	0.1915	0.0000	0.2642
	0.20	0.1074	0.3785	0.0000	0.3240	0.0000	0.4271
	0.30	0.3154	0.5314	0.0000	0.4402	0.1589	0.5689
	0.40	0.5076	0.6597	0.0000	0.5670	0.4085	0.7199
	0.50	0.6753	0.8477	0.0000	0.6996	0.6020	0.8971
	0.60	0.8536	1.1104	0.2391	0.8497	0.8658	1.0733
	0.70	1.0839	1.4443	0.5406	1.0399	1.1255	1.4727
	0.80	1.3713	1.9235	0.9313	1.3130	1.4159	1.9331
	0.90	1.8715	2.8048	1.4403	1.8875	1.8047	2.7325
0.95	2.3716	3.6366	1.7616	2.5748	2.2313	3.4623	
0.99	3.6009	6.1017	2.2724	4.2872	2.8495	5.6192	
0.5	0.01	-1.1974	0.0532	0.0000	0.0000	0.0000	0.0243
	0.05	-0.6364	0.1570	0.0000	0.0545	0.0000	0.1495
	0.10	-0.3817	0.2305	0.0000	0.1655	0.0000	0.2430
	0.20	-0.0594	0.3857	0.0000	0.3054	0.0000	0.4034
	0.30	0.1846	0.5329	0.0000	0.4288	0.0000	0.5576
	0.40	0.4902	0.6866	0.0000	0.5549	0.1756	0.7319
	0.50	0.6953	0.9840	0.0000	0.6944	0.4971	0.8876
	0.60	0.8877	1.2991	0.0000	0.8709	0.9016	1.1323
	0.70	1.0947	1.7099	0.4058	1.0766	1.1358	1.5977
	0.80	1.4187	2.3457	0.9959	1.3288	1.4378	2.2659
	0.90	1.8784	3.4241	1.4444	2.1575	1.8387	3.2980
0.95	2.3624	4.2660	1.7978	3.1419	2.2633	4.4991	
0.99	3.6005	6.7623	2.3994	5.2389	3.0282	6.7418	
0.8	0.01	-5.2724	-0.0086	0.0000	0.0000	0.0000	0.0000
	0.05	-2.1803	0.0989	0.0000	0.0000	0.0000	0.0000
	0.10	-1.3971	0.2126	0.0000	0.0000	0.0000	0.0582
	0.20	-0.4585	0.3768	0.0000	0.1483	0.0000	0.2869
	0.30	0.2273	0.5419	0.0000	0.3058	0.0000	0.4650
	0.40	0.5831	1.0858	0.0000	0.4573	0.0000	0.6418
	0.50	0.7381	1.7762	0.0000	0.6351	0.0836	0.8384
	0.60	0.9504	2.4662	0.0000	0.8025	0.8227	1.4445
	0.70	1.2065	3.4677	0.1439	1.0307	1.1316	2.4383
	0.80	1.5330	4.6351	1.1225	2.0680	1.4794	3.8804
	0.90	2.0350	6.8994	1.5247	4.8413	1.9321	6.6897
0.95	2.4518	8.7326	1.9550	6.6124	2.4816	8.9543	
0.99	3.5326	14.5791	2.5920	10.2892	3.2637	14.3856	

그림 4.1: cANOVA, ML 및 REML추정량의  $\widehat{W}_\phi$ 에 대한 EQDGs ( $\rho = 0.2$ )

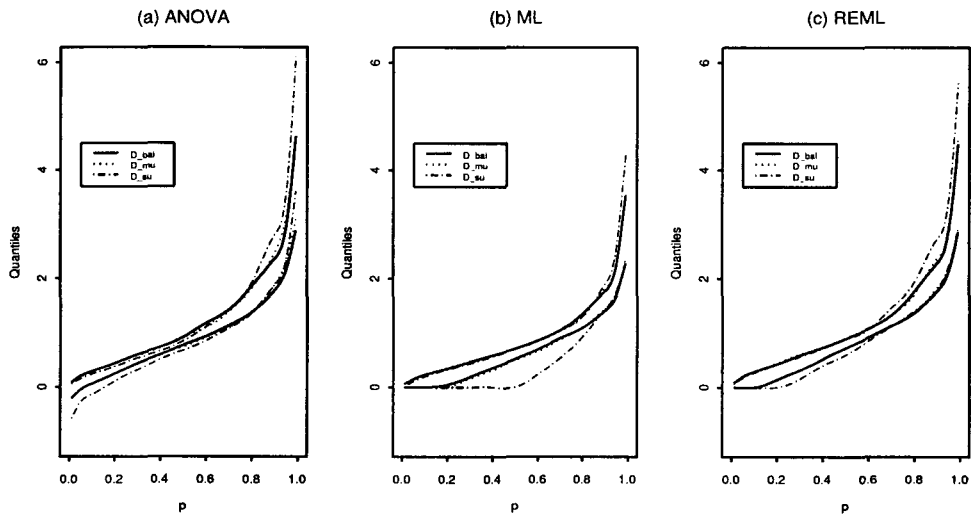


그림 4.2: cANOVA, ML 및 REML추정량의  $\widehat{W}_\phi$ 에 대한 EQDGs ( $\rho = 0.5$ )

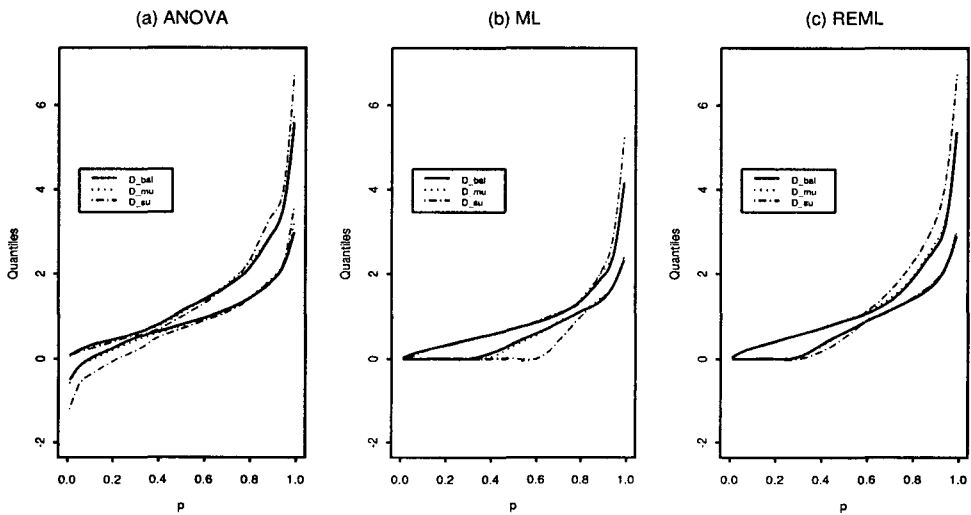


그림 4.3: cANOVA, ML 및 REML 추정량의  $\widehat{W}_\phi$ 에 대한 EQDGs ( $\rho = 0.8$ )

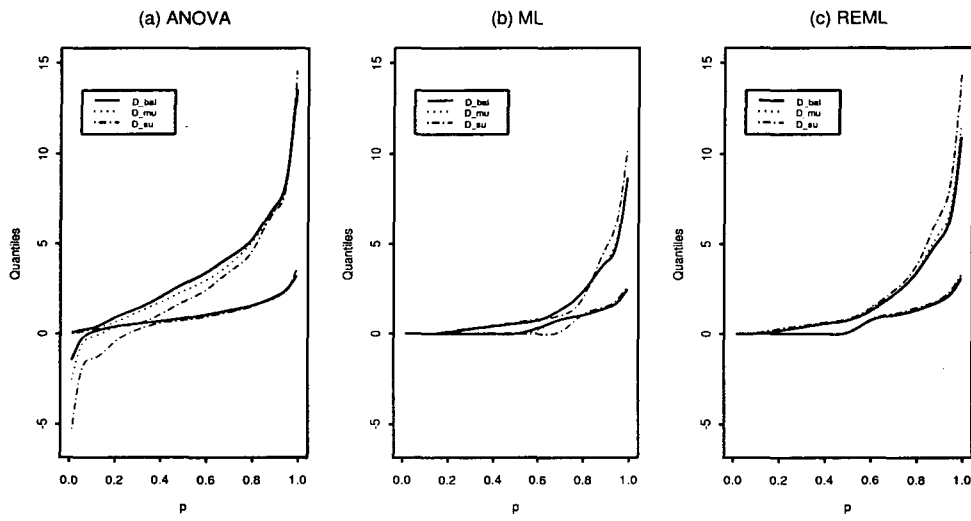


그림 4.4: 디자인  $D_{bal}, D_{mu}, D_{su}$ 에서  $\widehat{W}_\phi$ 에 대한 EQDGs( $\rho = 0.2$ )

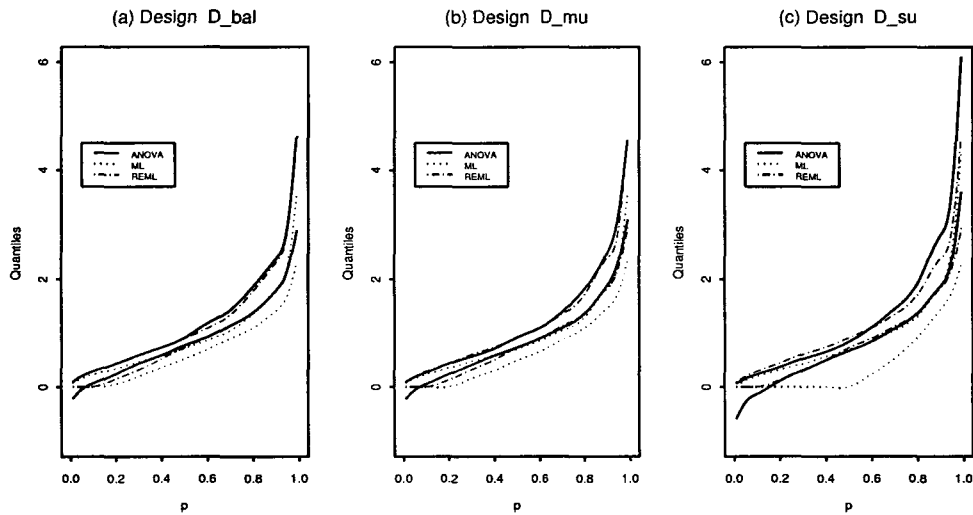


그림 4.5: 디자인  $D_{bal}$ ,  $D_{mu}$ ,  $D_{su}$ 에서  $\widehat{W}_\phi$ 에 대한 EQDGs( $\rho = 0.5$ )

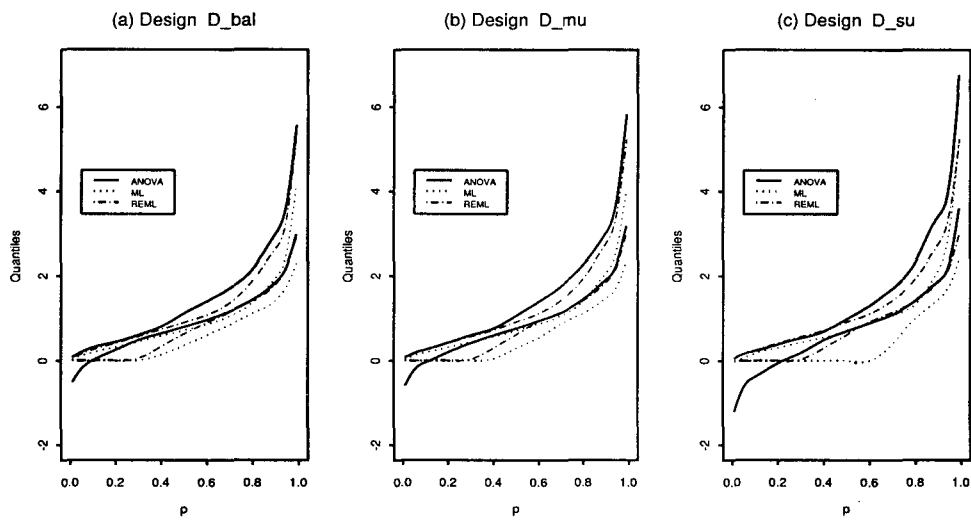
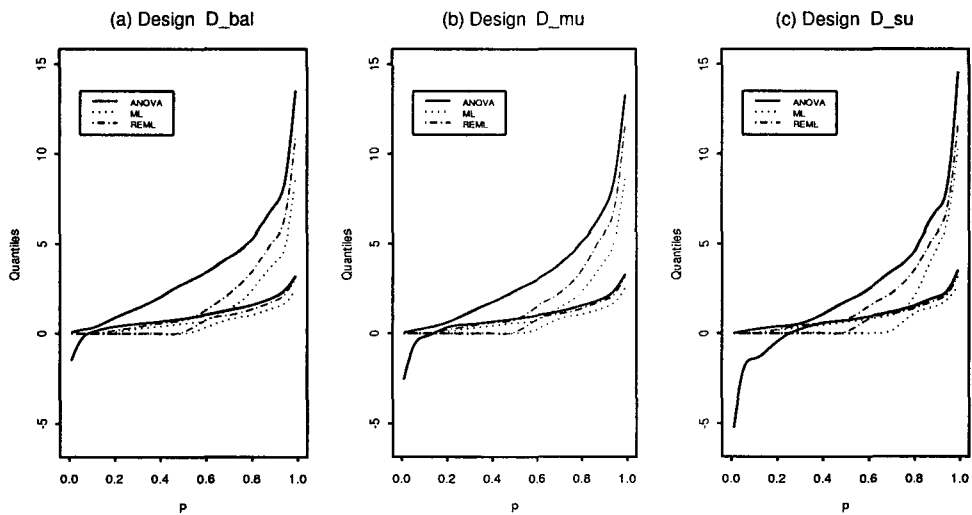


그림 4.6: 디자인  $D_{bal}$ ,  $D_{mu}$ ,  $D_{su}$ 에서  $\widehat{W}_\phi$ 에 대한 EQDGs( $\rho = 0.8$ )



## 5. 결론

본 연구에서는 오차성분이 계열상관을 갖는 불균형 랜덤모형에서 분산성분의 추정방법에 대하여 연구하였다. 특히 본 연구에서 유도된 분산성분에 대한 cANOVA 추정방법은 2단계 추정방법으로 복잡한 수치적인 방법을 이용하는 ML이나 REML추정방법에 비하여 계산이 편리하다는 장점을 갖고 있다. EQDGs를 이용한 모의실험에서는 cANOVA추정량은 디자인의 불균형정도에 있어서는 크게 영향을 받지 않는것으로 나타났다. 그러나 계열상관값에 대해서는 변동성을 보이고있다. 더불어 계열상관값 및 디자인의 불균형 정도를 동시에 고려하는 경우에는 ML추정이 cANOVA, REML추정보다 안정적인 추정량을 제공해주어 선호되는 추정방법으로 나타났다.

## 감사의 글

본 논문에 대하여 많은 조언을 해주신 심사위원님들께 감사를 드립니다.

## 참고문헌

- [1] 이준영, (2000). 일반화된 선형혼합모형에 관한 최근 연구 동향, 응용통계연구, 제13권 2호, 541-562.
- [2] Ahrens, H., Kleffe, J. and Tensler, R. (1981). Mean square error comparison for MINQUE, ANOVA and two alternative estimators under the unbalanced one way random model, *Biometrical Journal*, vol. 23, 323-342.
- [3] Baltagi, B. and Li, Q. (1991). A transformation that will circumvent the problem of autocorrelation in an error component model, *Journal of Econometrics*, vol. 48, 385-393.
- [4] Chaloner, K. (1987). A Bayesian approach in the estimation of variance components for the unbalanced one-way random model, *Technometrics*, vol. 29, 323-337.
- [5] Chi, E.M. and Reinsel, G.C. (1989). Models for longitudinal data with random effects and AR(1) errors, *Journal of the American Statistical Association*, vol. 84, 452-459.
- [6] Diggle, P.J., Liang, K.Y., Zeger, S.L. (1994). *Analysis of Longitudinal Data*, Clarendon Press, Oxford.
- [7] Hocking, R.R. (1985). *The Analysis of Linear Model*, Brooks/Cole company, Monterey, California.
- [8] Khuri, A.I. (1997). Quantile dispersion graphs for analysis of variance estimates of variance components, *Journal of Applied Statistics*, vol. 24, 711-722.



- [9] Lee, J. and Khuri, A.I. (1999). Graphical technique for comparing designs for random models, *Journal of Applied Statistics*, vol. 26, 933-947.
- [10] Lee, J. and Khuri, A.I. (2000). Quantile dispersion graphs for the comparison of designs for a random two-way model, *Journal of Statistical Planning and Inference*, vol. 91, 123-137.
- [11] McCulloch, C.E. and Searle, S.R. (2001). *Generalized, Linear and Mixed Models*, Wiley, New York.
- [12] Patterson, H.D. and Thompson, R. (1971). Recovery of inter-block information when block sizes are unequal, *Biometrika*, vol. 58, 545-554.
- [13] Prais, S.J. and Winsten, C.B. (1954). Trend estimators and serial correlation, *Cowles Commission Discussion Paper, No. 373*, University of Chicago.
- [14] Searle, S.R., Casella G. and McCulloch, C.E. (1992). *Variance Components*, Wiley, New York.
- [15] Swallow, W.H., Monahan, J.F. (1984). Monte Carlo comparison of ANOVA, MIVQUE, REML and ML estimators of variance components, *Technometrics*, vol. 26, 47-57.

[ 2002년 2월 접수, 2002년 7월 채택 ]

## A Comparison of Estimation in an Unbalanced Linear Mixed Model\*

Seuck Heun Song<sup>1)</sup> Byoung Cheol Jung<sup>2)</sup>

### ABSTRACT

This paper derives three estimation methods for the between group variance component for serially correlated random model. To compare their estimation capability, three designs having different degree of unbalancedness are considered. The so-called empirical quantile dispersion graphs(EQDGs) used to compare estimation methods as well as designs. The proposed conditional ANOVA estimation is robust for design unbalancedness, however, ML estimation is preferred to the conditional ANOVA and REML estimation regardless of design unbalancedness and correlation coefficient.

*Keywords:* linear mixed model, cANOVA, ML, REML, EQDGs

---

\* This work was supported by Korea Research Foundation Grant(KRF-2000-041-D00055)

1) Associate Professor, Dept of Statistics, Korea University, Seoul 136-701, Korea

2) Dept of Statistics, Sungshin Women's University, Seoul 136-742, Korea