응용통계연구 제 17권 3호, 2004년, pp. 547-556

장애모수가 존재하는 불균형 패널회귀모형에서의 검정법*

이재원[1)] 정병철[2)] 송석헌[3)]

요 약

본 논문에서는 불균형 패널회귀모형에서 장애모수가 존재하는 경우 오차성분의 존재유무를 검정하기 위하여 Lagrange Multiplier 검정통계량을 제안하였다. 이러한 LM 통계량은 그룹변환에 대한 불변성의 성질을 이용하여 유도된 통계량으로 모의실험 결과, 제안된 LM검정은 명목유의수준을 잘 유지하고 있는 것으로 나타났으며 LR검정에 비하여 검정력에 있어서도 높게 나타났다.

주요용어: 불균형 패널회귀모형, 장애모수, LM검정

1. 서론

그 동안 분산성분모형에서 대한 추정과 오차성분의 존재유무에 대한 가설검정에 대해서는 많은 연구들이 진행되어 오고 있다(Khuri and Sinha(1998), Mathew and Sinha(1988), Moulton and Randolph(1989)). 이와 같은 오차성분모형은 최근 들어 패널자료를 분석하는 모형으로 널리 사용되고 있다(Baltagi(2001), Baltagi, et al.(2001), Jung et al. (2001), Song et al. (2003)). 패널모형을 다루는 주요한 이득중 한가지는 개체 또는 시간의 흐름 속에서 발생하는 이질성을 오차성분을 이용하여 조절할 수 있다는 것이다(Baltagi, 2001). 이러한 이질성은 개체효과를 나타내는 오차성분 및 시간효과를 나타내는 오차성분 등에 의하여 모형화된다. 각 효과를 나타내는 오차성분이 존재하는가를 검정하기 위해서 Breusch and Pagan (1980)은 라그랑지 승수(Lagrange Multiplier, LM) 검정법을 유도하였다. LM 검정은 귀무가설하에서의 추정에 근거하기 때문에 대부분의 경우 계산이 간편하다는 장점이 있다. 반면 LM검정을 분산성분모형에서 사용하였을 경우 발생하는 어려운 문제중의 하나는 장애모수(nuisance parameter)가 존재하는 경우이다. 즉, 예를 들어, 랜덤 개체효과가 존재하는가를 검정하는 문제에 있어서 시간효과를 무시하게되면 잘못된 추론을 이끌게 된다. 그러므로 이와 같은 경우에는 시간효과가 존재한다는 가정하에서 개체효과가 존재하는가를 검정하는 조건부 검정이 사용되어야 한다. 이때 시간효과를 나타내는 분산성분은 장애모수로서 고려된다.

---

* 본 연구는 KRF(2002-042-C00008)연구비 지원에 의하여 수행되었음.
1) (136-701) 서울시 성북구 안암동 5가 1번지 고려대학교 통계학과, 교수,
E-mail: jelee@korea.ac.kr
2) (136-742) 서울시 성북구 동선동 249-1 성신여자대학교 통계학과.
E-mail: bcjung@korea.ac.kr
3) (136-701) 서울시 성북구 안암동 5가 1번지 고려대학교 통계학과, 부교수(교신저자)
E-mail: ssong@korea.ac.kr

본 논문에서는 장애모수가 존재하는 불균형 불균형 패널회귀모형에서 오차성분의 존재 유무를 검정하자 한다. 이를 위해 불변성(Invariance)의 성질하에서 주변우도함수(marginal likelihood)를 이용한 단측 LM검정 통계량을 유도하고자 한다. 또한 유도된 LM검정통계량의 성질을 모의실험을 통하여 파악하고자 한다.

2. 불균형 패널회귀모형

다음과 같은 불균형 패널회귀모형을 고려하여 보자.

$$y_{it} = x_{it}'\beta + u_{it}, \qquad i = 1, \cdots, N_t, \quad t = i, \cdots, T \tag{2.1}$$

여기서 y_{it}는 i번째 개체(개인, 가구 등)에 대한 시점 t에서의 관측치를 나타내는 반응변수이고, x_{it}는 k개의 변수로 이루어진 설명변수벡터이며 u_{it}는 나머지 오차항을 나타낸다. N_t는 시점 t에서 관찰된 개체의 수를 나타낸다. 모형 (2.1)에서 오차항 u_{it}는 다음과 같은 이원오차성분모형을 따른다고 가정하자.

$$u_{it} = \mu_i + \lambda_t + \nu_{it}, \qquad i = 1, \cdots, N_t, \quad t = i, \cdots, T \tag{2.2}$$

여기서 μ_i는 관측될 수 없는 개체효과(individual specific effect)를 나타내는 확률변수이고 λ_t는 시간효과(time specific effect)를 나타내는 확률변수이며, ν_{it}는 나머지 오차항을 나타낸다. μ_i, λ_t와 ν_{it}은 서로 독립이며 각각 $\mu_i \overset{i.i.d.}{\sim} N(0, \sigma_\mu^2)$, $\lambda_t \overset{i.i.d.}{\sim} N(0, \sigma_\lambda^2)$, $\nu_{it} \overset{i.i.d.}{\sim} N(0, \sigma_\nu^2)$라고 가정하자. 모형 (2.1)을 행렬을 이용하여 나타내면 다음과 같다.

$$y \;=\; X\beta \;+\; u \tag{2.3}$$

여기서 $\sum_{t=1}^{T} N_t = n$이라 할 때 y 는 $n \times 1$인 반응변수벡터, X는 $n \times k$인 설명변수행렬이며, β는 $k \times 1$인 회귀계수벡터이다. 또한 식 (2.2)의 오차항은 다음과 같다.

$$u \;=\; \Delta_1\mu \;+\; \Delta_2\lambda \;+\; \nu \tag{2.4}$$

여기서 $\Delta_1 = (D_1', D_2', \cdots, D_T')$, $\Delta_2 = \bigoplus_{t=1}^{T}(D_t\iota_n) = \bigoplus_{t=1}^{T}\iota_{N_t}$이며, D_t는 항등행렬 I_N에서 시점 t에 관측되지 않은 개체에 해당하는 행을 제거한 $N_t \times N$행렬이며, ι_n과 ι_{N_t}는 각각 모든 원소가 1로 이루어진 $n \times 1$과 $N_t \times 1$ 벡터를 나타낸다. μ는 $(\mu_1, \cdots, \mu_{N_t})'$이고, λ는 $(\lambda_1, \cdots, \lambda_T)'$인 벡터이며 ν는 $(\nu_{11}, \cdots, \nu_{N_1 1}, \cdots, \mu_{N_t 1}, \cdots, \nu_{N_t T})'$인 벡터이다. 이와 같은 가정으로부터 오차항의 분산 공분산 행렬은 다음과 같이 주어진다.

$$Cov(u) \;=\; \Omega \;=\; \sigma_\mu^2\Delta_1\Delta_1' + \sigma_\lambda^2\Delta_2\Delta_2' + \sigma_\nu^2 I_n \tag{2.5}$$

3. 검정통계량

본 절에서는 장애모수가 존재하는 불균형 이원오차성분모형에서 분산성분의 존재유무에 대한 가설검정을 위한 적절한 검정통계량을 제안하고자 한다. 식 (2.5)와 같은 공분산 행렬을 갖는 불균형 이원오차성분모형에서 귀무가설 $H_0: \sigma_\mu^2 = 0$ (또는 $H_0: \sigma_\lambda^2 = 0$)에 대한 검정법을 고려해보자. 이는 개체효과(또는 시간효과)의 존재유무에 대한 검정으로 σ_λ^2(또는 σ_μ^2)은 장애모수(nuisance parameter)로 고려될 수 있다. 이는 랜덤 개체효과가 존재하는가를 검정하는 문제에 있어서 시간효과를 무시하게되면 잘못된 추론을 이끌게 된다. 그러므로 이 경우에는 시간효과가 존재한다는 가정하에서 개체효과가 존재하는가를 검정하여야 한다. 이때 시간효과를 나타내는 오차성분은 장애모수로서 고려된다. 본 절에서 검정하고자하는 가설은 다음과 같다.

$$H_0 \ : \ \sigma_\mu^2 = 0 \ (\sigma_\lambda^2 > 0) \quad vs \quad H_1 \ : \ \sigma_\mu^2 > 0 \ (\sigma_\lambda^2 > 0) \tag{3.1}$$

정리 3.1: 식 (3.1)의 가설을 검정하기 위한 LM검정통계량은 다음과 같이 얻어진다.

$$LM_\mu = \frac{\hat{S}_\mu}{(\hat{I}_{11} - \hat{I}_{12}\hat{I}_{22}^{-1}\hat{I}_{21})^{1/2}} \tag{3.2}$$

여기서 $\hat{S}_\mu$과 $\hat{I}_{11}, \hat{I}_{12}, \hat{I}_{21}, \hat{I}_{22}$ 는 귀무가설하에서 얻어진 최대우도추정량을 이용한 스코어 벡터와 정보행렬의 각 원소를 나타낸다.

증명: 만일 $\theta_1 = \sigma_\mu^2/\sigma_\nu^2$ 및 $\theta_2 = \sigma_\lambda^2/\sigma_\nu^2$ 라 했을 때 식 (2.5)의 공분산행렬은 다음과 같이 표현할 수 있다.

$$\Omega = \sigma_\nu^2(\theta_1\Delta_1\Delta_1' + \theta_2\Delta_2\Delta_2' + I_n) = \sigma_\nu^2\Sigma(\theta) \tag{3.3}$$

여기서 $\theta' = (\theta_1, \theta_2)$이다. 또한 식 (3.1)에 주어진 가설은 다음과 같은 가설과 동일한 검정문제가 된다.

$$H_0 \ : \ \theta_1 = 0 \ (\theta_2 > 0) \quad vs \quad H_1 \ : \ \theta_1 > 0 \ (\theta_2 > 0) \tag{3.4}$$

식 (3.4)와 같은 가설에 대한 검정은 $y \to \eta_0 y + X\eta$와 같은 변환에 대하여 불변(invariant)한다는 사실이 잘 알려져 있다(King, 1980). 여기서 η_0 는 0보다 큰 상수이고 η는 $k \times 1$벡터이다. Ara and King(1993)은 이와 같은 가설검정 문제에 대하여 $\nu = Pz/(z'P'Pz)^{1/2}$ 라는 변환을 고려하여 접근하였다. 여기서 P는 $PP' = I_n$, $P'P = I - X(X'X)^{-1}X'$을 만족하는 $m \times n$행렬이며 $m = \sum N_t - k$이다. 이 경우 ν 는 위에서 정의된 $y \to \eta_0 y + X\eta$ 라는 변환에서 불변이 되는 벡터가 되며, ν에 대한 확률분포는 다음과 같다.

$$f(v,\theta)dv = \frac{1}{2}\Gamma(m/2)\pi^{-m/2}|P\Sigma(\theta)P'|^{-1/2}a^{-m/2}dv \tag{3.5}$$

여기서 $a = v'(P\Sigma(\theta)P')^{-1}v = \hat{u}'\Sigma^{-1}(\hat{\theta})\hat{u}/z'z$이며 $\hat{u}$는 GLS잔차를 나타내고, z는 OLS잔차를 나타낸다. 그러므로 식 (3.5)에 대한 로그우도함수는 다음과 같이 구해진다.

$$L(\theta) = const - \frac{1}{2}|P\Sigma(\theta)P'| - \frac{m}{2}\log[v'(P\Sigma(\theta)P')^{-1}v] \tag{3.6}$$

식 (3.6)의 로그우도함수를 이용하여 가설 (3.4)에 대한 검정통계량을 유도하기 위해서는 귀무가설하에서 구해지는 θ에 대한 일차미분값과 정보행렬이 필요하다. $\hat{\theta}_2$을 귀무가설하에서(즉, $\theta_1 = 0$) 식 (3.6)의 주변우도함수를 최대화시키는 값이라 하면 귀무가설하에서 θ에 대한 최대우도추정량은 $\hat{\theta}' = (0, \hat{\theta}_2)$이 된다.
식 (3.4)의 귀무가설하에서 얻어진 최대우도추정량 $\hat{\theta}$값을 이용한 스코어벡터는 다음과 같이 구해진다.

$$S(\theta_i) = \frac{\partial L(\theta)}{\partial \theta_i} = \frac{m}{2}\Big(\frac{u'\Sigma^{-1}A_i(\hat{\theta})\Sigma^{-1}u}{u'\Sigma^{-1}u}\Big) - \frac{1}{2}tr(\Delta(\hat{\theta})A_i(\hat{\theta})) \tag{3.7}$$

여기서 $A_i(\hat{\theta}) = \dfrac{\partial \Sigma(\theta)}{\partial \theta_i}$이며 $\Delta(\theta) = \Sigma^{-1}(\theta) - \Sigma^{-1}(\theta)X(X'\Sigma^{-1}(\theta)X)^{-1}X'\Sigma^{-1}(\theta)$이다. 또한 정보행렬의 (i,j)번째 원소는 다음과 같이 구해진다.

$$\begin{aligned} &E\Big[-\frac{\partial^2 L(\theta)}{\partial\theta_i\partial\theta_j}\Big] \\ &= \frac{m}{2m+4}tr\big[\Delta(\theta)A_i(\theta)\Delta(\theta)A_j(\theta)\big] - \frac{1}{2m+4}\Big\{tr\big[\Delta(\theta)A_i(\theta)\big]tr\big[\Delta(\theta)A_j(\theta)\big]\Big\} \end{aligned} \tag{3.8}$$

식 (3.7)의 스코어벡터와 식 (3.8)의 정보행렬을 구하기 위해서는, 귀무가설하에서의 공분산행렬과 그의 역행렬을 구해면 다음과 같다.

$$\Sigma(\theta) = \phi_2\Delta_2\Delta_2' + I_n = \phi_2\bigoplus_{t=1}^{T} J_{N_t} + I_n,$$

$$\Sigma^{-1}(\theta) = \bigoplus_{t=1}^{T}\frac{1}{N_t\theta_2+1}\bar{J}_{N_t} + \bigoplus_{t=1}^{T}E_{N_t} \tag{3.9}$$

J_{N_t}는 모든 원소가 1로 이루어진 $N_t \times N_t$행렬이다. 또한$\bar{J}_{N_t} = J_{N_t}/N_t$이고 $E_{N_t} = I_{N_t} - \bar{J}_{N_t}$이며 $\bar{J}_{N_t}\cdot E_{N_t} = 0$이다. 이를 이용하여 계산하면 다음과 같은 결과를 얻을 수 있다.

$$\frac{\partial\Sigma(\theta)}{\partial\theta_1} = A_1(\theta) = \Delta_1\Delta_1'$$

$$\begin{aligned} \hat{\Sigma}^{-1}(\theta)A_1(\theta)\hat{\Sigma}^{-1} &= \Big\{I_n - \Big(\bigoplus_{t=1}^{T}\frac{N_t\hat{\theta}_2}{N_t\hat{\theta}_2+1}\bar{J}_{N_t}\Big)\Big\}\Delta_1\Delta_1'\Big\{I_n - \Big(\bigoplus_{t=1}^{T}\frac{N_t\hat{\theta}_2}{N_t\hat{\theta}_2+1}\bar{J}_{N_t}\Big)\Big\} \\ &= \Delta_1\Delta_1' - \Big(\bigoplus_{t=1}^{T}\frac{N_t\hat{\theta}_2}{N_t\hat{\theta}_2+1}\bar{J}_{N_t}\Big)\Delta_1\Delta_1' - \Delta_1\Delta_1'\Big(\bigoplus_{t=1}^{T}\frac{N_t\hat{\theta}_2}{N_t\hat{\theta}_2+1}\bar{J}_{N_t}\Big) \\ &+ \Big(\bigoplus_{t=1}^{T}\frac{N_t\hat{\theta}_2}{N_t\hat{\theta}_2+1}\bar{J}_{N_t}\Big)\Delta_1\Delta_1'\Big(\bigoplus_{t=1}^{T}\frac{N_t\hat{\theta}_2}{N_t\hat{\theta}_2+1}\bar{J}_{N_t}\Big) \end{aligned} \tag{3.10}$$

또한 $\Delta(\hat{\theta})A_1 = \big(\hat{\Sigma}^{-1} - \hat{\Sigma}^{-1}X(X'\hat{\Sigma}^{-1}X)^{-1}X'\hat{\Sigma}^{-1}\big)\Delta_1\Delta_1'$이므로 일차미분값과 정보행렬을

구하기위해서 필요한 $tr[\Delta(\hat{\theta})A_1]$은 다음과 같이 구해진다.

$$\begin{aligned}
tr[\Delta(\hat{\theta})A_1] &= tr\Big[\Big\{I_n - \Big(\bigoplus_{t=1}^{T} \frac{N_t\hat{\theta}_2}{N_t\hat{\theta}_2+1}\bar{J}_{N_t}\Big)\Big\}\Delta_1\Delta_1'\Big] \\
&\quad - tr\Big[X'\hat{\Sigma}^{-1}\Delta_1\Delta_1'\hat{\Sigma}^{-1}X(X'\hat{\Sigma}^{-1}X)^{-1}\Big] \\
&= n - \sum_{t=1}^{T}\frac{N_t\hat{\theta}_2}{N_t\hat{\theta}_2+1} - tr\Big[X'\hat{\Sigma}^{-1}\Delta_1\Delta_1'\hat{\Sigma}^{-1}X(X'\hat{\Sigma}^{-1}X)^{-1}\Big] \\
&= (n-T) - \sum_{t=1}^{T}\frac{1}{N_t\hat{\theta}_2+1} - tr\Big[X'\hat{\Sigma}^{-1}\Delta_1\Delta_1'\hat{\Sigma}^{-1}X(X'\hat{\Sigma}^{-1}X)^{-1}\Big] \quad (3.11)
\end{aligned}$$

여기서 $\hat{\Sigma}^{-1}\Delta_1\Delta_1'\hat{\Sigma}^{-1}$은 식 (3.11)에 주어져 있다. 그러므로, $\hat{\theta} = (0, \hat{\theta}_2)$하에서 계산된 일차 미분값은 다음과 같이 구해진다.

$$\hat{S}_\mu = \frac{m}{2}\Big[\frac{\hat{u}'\hat{\Sigma}^{-1}\Delta_1\Delta_1'\hat{\Sigma}^{-1}\hat{u}}{\hat{u}'\hat{\Sigma}^{-1}\hat{u}}\Big] - \frac{1}{2}tr\big[\Delta(\hat{\theta})\Delta_1\Delta_1'\big] \quad (3.12)$$

여기서 $\hat{u}$은 귀무가설하에서 계산된 GLS 잔차를 나타낸다. 또한, 귀무가설하에서 θ에 대한 정보행렬을 유도하기 위하여 다음과 같은 결과들을 얻을 수 있다.

$$\frac{\partial\Sigma}{\partial\theta_2} = A_2 = \Delta_2\Delta_2' = \bigoplus_{t=1}^{T} J_{N_t}$$

$$\begin{aligned}
\Delta(\hat{\theta})A_2 &= \Big(\hat{\Sigma}^{-1} - \hat{\Sigma}^{-1}X(X'\hat{\Sigma}^{-1}X)^{-1}X'\hat{\Sigma}^{-1}\Big)\bigoplus_{t=1}^{T} J_{N_t} \\
&= \Big(\bigoplus_{t=1}^{T}\frac{N_t}{N_t\hat{\theta}_2+1}\bar{J}_{N_t}\Big) - \hat{\Sigma}^{-1}X(X'\hat{\Sigma}^{-1}X)^{-1}X'\Big(\bigoplus_{t=1}^{T}\frac{N_t}{N_t\hat{\theta}_2+1}\bar{J}_{N_t}\Big)
\end{aligned}$$

그러므로, 적절한 연산을 통하면 $tr\big[\Delta(\hat{\theta})A_2\big]$은 다음과 같다.

$$\begin{aligned}
tr\big[\Delta(\hat{\theta})A_2\big] &= tr\Big[\Big(\bigoplus_{t=1}^{T}\frac{N_t}{N_t\hat{\theta}_2+1}\bar{J}_{N_t}\Big) - \hat{\Sigma}^{-1}X(X'\hat{\Sigma}^{-1}X)^{-1}X'\Big(\bigoplus_{t=1}^{T}\frac{N_t}{N_t\hat{\theta}_2+1}\bar{J}_{N_t}\Big)\Big] \\
&= \sum_{t=1}^{T}\frac{N_t}{N_t\hat{\theta}_2+1} - tr\Big[X'\Big(\bigoplus_{t=1}^{T}\frac{N_t}{(N_t\hat{\theta}_2+1)^2}\bar{J}_{N_t}\Big)X(X'\hat{\Sigma}^{-1}X)^{-1}\Big] \quad (3.13)
\end{aligned}$$

마지막으로 정보행렬을 구하기 위하여 $B_{ij} = tr[\Delta(\hat{\theta})A_i\Delta(\hat{\theta})A_j]$, $i, j = 1, 2$ 라 하여 계산하면 다음과 같이 된다.

$$\begin{aligned}
B_{11} &= tr[\Delta(\hat{\theta})A_1\Delta(\hat{\theta})A_1] \quad (3.14) \\
&= tr[\hat{\Sigma}^{-1}\Delta_1\Delta_1'\hat{\Sigma}^{-1}\Delta_1\Delta_1'] - 2tr[X'\hat{\Sigma}^{-1}\Delta_1\Delta_1'\hat{\Sigma}^{-1}\Delta_1\Delta_1'\hat{\Sigma}^{-1}X(X'\hat{\Sigma}^{-1}X)^{-1}] \\
&\quad + tr[\{X'\hat{\Sigma}^{-1}\Delta_1\Delta_1'\hat{\Sigma}^{-1}X(X'\hat{\Sigma}^{-1}X)^{-1}\}^2]
\end{aligned}$$

$$\begin{aligned}
&= \sum_{i=1}^{N} T_i^2 - 2\sum_{s=1}^{T}\sum_{t=1}^{T} \frac{\hat\theta_2 C_{ts}}{N_t\hat\theta_2+1} + \sum_{s=1}^{T}\sum_{t=1}^{T} \frac{\hat\theta_2^2}{(N_t\hat\theta_2+1)(N_s\hat\theta_2+1)} C_{ts}^2 \\
&\quad -2\, tr[X'\hat\Sigma^{-1}\Delta_1\Delta_1'\hat\Sigma^{-1}\Delta_1\Delta_1'\hat\Sigma^{-1}X(X'\hat\Sigma^{-1}X)^{-1}] \\
&\quad +tr[\{X'\hat\Sigma^{-1}\Delta_1\Delta_1'\hat\Sigma^{-1}X(X'\hat\Sigma^{-1}X)^{-1}\}^2]
\end{aligned}$$

$$\begin{aligned}
B_{22} &= tr[\Delta(\hat\theta)A_2\Delta(\hat\theta)A_2] \qquad (3.15)\\
&= tr\left[\left\{\left(\bigoplus_{t=1}^{T}\frac{N_t}{N_t\hat\theta_2+1}\bar J_{N_t}\right) - \hat\Sigma^{-1}X(X'\hat\Sigma^{-1}X)^{-1}X'\left(\bigoplus_{t=1}^{T}\frac{N_t}{N_t\hat\theta_2+1}\bar J_{N_t}\right)\right\}^2\right] \\
&= tr\left[\bigoplus_{t=1}^{T}\frac{N_t^2}{(N_t^2\hat\theta_2+1)^2}\bar J_{N_t}\right] - 2\, tr\left[X'\left(\bigoplus_{t=1}^{T}\frac{{N_t}^2}{(N_t\hat\theta_2+1)^3}\bar J_{N_t}\right)X(X'\hat\Sigma^{-1}X)^{-1}\right] \\
&\quad +tr\left[\left\{X'\left(\bigoplus_{t=1}^{T}\frac{N_t}{(N_t\hat\theta_2+1)^2}\bar J_{N_t}\right)X(X'\hat\Sigma^{-1}X)^{-1}\right\}^2\right] \\
&= \sum_{t=1}^{T}\frac{N_t^2}{(N_t\hat\theta_2+1)^2} - 2tr\left[X'\bigoplus_{t=1}^{T}\left(\frac{N_t^2}{(N_t\hat\theta_2+1)^3}\bar J_{N_t}\right)X(X'\hat\Sigma^{-1}X)^{-1}\right] \\
&\quad +tr\left[\left\{X'\bigoplus_{t=1}^{T}\left(\frac{N_t}{(N_t\hat\theta_2+1)^2}\bar J_{N_t}\right)X(X'\hat\Sigma^{-1}X)^{-1}\right\}^2\right]
\end{aligned}$$

$$\begin{aligned}
B_{12} &= tr[\Delta(\hat\theta)A_1\Delta(\hat\theta)A_2] \qquad (3.16)\\
&= tr[\hat\Sigma^{-1}\Delta_1\Delta_1'\hat\Sigma^{-1}\Delta_2\Delta_2'] - 2tr[X'\hat\Sigma^{-1}\Delta_1\Delta_1'\hat\Sigma^{-1}\Delta_2\Delta_2'\hat\Sigma^{-1}X(X'\hat\Sigma^{-1}X)^{-1}] \\
&\quad +tr[\{X'\hat\Sigma^{-1}\Delta_1\Delta_1'\hat\Sigma^{-1}X(X'\hat\Sigma^{-1}X)^{-1}\}\{X'\hat\Sigma^{-1}\Delta_1\Delta_1'\hat\Sigma^{-1}X(X'\hat\Sigma^{-1}X)^{-1}\}] \\
&= \sum\frac{N_t}{(N_t\hat\theta_2+1)^2} - 2tr[X'\hat\Sigma^{-1}\Delta_1\Delta_1'\hat\Sigma^{-1}\Delta_2\Delta_2'\hat\Sigma^{-1}X(X'\hat\Sigma^{-1}X)^{-1}] \\
&\quad +tr[\{X'\hat\Sigma^{-1}\Delta_1\Delta_1'\hat\Sigma^{-1}X(X'\hat\Sigma^{-1}X)^{-1}\}\{X'\hat\Sigma^{-1}\Delta_2\Delta_2'\hat\Sigma^{-1}X(X'\hat\Sigma^{-1}X)^{-1}\}]
\end{aligned}$$

여기서 C_{st}는 시점 t와 시점 S에서 동시에 관찰된 개체의 수를 나타낸다. 그러므로, $\hat\theta = (0,\hat\theta_2)$ 하에서 계산된 정보행렬은 다음과 같다.

$$\hat I_{ij} = \frac{m}{2m+4}B_{ij} - \frac{1}{2m+4}tr[\Delta(\hat\theta)A_i]tr[\Delta(\hat\theta)A_j], \quad i,j=1,2 \qquad (3.17)$$

그러므로, $H_0 : \theta_1 = 0$을 검정하기 위한 단측 LM 검정통계량은 다음과 같이 유도된다.

$$LM_\mu = \frac{\hat S_\mu}{(\hat I_{11} - \hat I_{12}\hat I_{22}^{-1}\hat I_{21})^{1/2}} \qquad (3.18)$$

여기서 $\hat S_\mu$은 $\hat\theta = (0,\hat\theta_2)$ 하에서 계산된 θ_1에 대한 일차미분값이고 $\hat I_{11}, \hat I_{12}, \hat I_{21}, \hat I_{22}$ 는 $\hat\theta = (0,\hat\theta_2)$ 하에서 계산된 θ_1과 θ_2에 대응하는 정보행렬의 각 원소이다. 식 (3.18)의 검정통계량은 귀무가설하에서 점근적으로 표준정규분포를 따른다.

4. 모의실험

4.1. 모의실험 방법

다음과 같은 패널회귀모형을 고려해보자.

$$y_{it} = \alpha + x'_{it}\beta + u_{it}, \qquad t = 1, \cdots, T, \ \ i = 1, \cdots, N_t \tag{4.1}$$

여기서 $u_{it} = \mu_i + \lambda_t + \nu_{it}$이다. 설명변수 x_{it} 벡터는 Nerlove(1971)에 의해서 제안된 방법으로 다음 식에 의해 생성되었다.

$$x_{it} = 0.1t + 0.5x_{it} + \omega_{it} \tag{4.2}$$

여기서는 ω_{it}는 균일 분포$(-0.5, 0.5)$를 따르는 난수이며, 이때 초기값 x_{i0}는 $5 + 10\omega_{i0}$이다. 모의실험 전체를 통하여 $\alpha = 5, \beta = 2$를 사용하였다. 오차항에 대해서는 $\mu_i \sim i.i.N.(0, \sigma_\mu^2$, $\lambda_t \sim i.i.N.(0, \sigma_\lambda^2)$와 $\nu_{it} \sim i.i.N.(0, \sigma_\nu^2)$을 사용하였으며 총 분산을 $\sigma^2 = \sigma_\mu^2 + \sigma_\lambda^2 + \sigma_\nu^2$이라 했을 때, $\sigma^2 = 20$으로 고정하였다. 각 실험에서는 $\rho_1 = \sigma_\mu^2/\sigma^2$과 $\rho_2 = \sigma_\lambda^2/\sigma^2$ 의 값을 $(0.0, 0.05, 0.1, 0.2, 0.4, 0.6, 0.8)$ 사이에서 변화시켜가며 실험하였으며 항상 $1-\rho_1-\rho_2$의 값이 0보다 크도록 하였다. 또한 개체의 수는 $N = 30$으로 고정하였으며 불균형 정도에 따라 검정통계량이 어떤 영향을 받는지를 알아보기 위하여 Swallow와 Searle(1978)이 사용한 불균형 패턴을 사용하여 모의실험을 실시하였다. 만일 5(15)를 15개의 개체가 5년동안 측정되는 형태의 T패턴이라 정의하면, 본 실험에서는 $P_1 = 5(15), 9(15)$, $P_2 = 5(10), 7(10), 9(10)$, $P_3 = 3(6), 5(6), 7(6), 9(6), 11(6)$, $P_4 = 3(9), 5(6), 9(6), 11(9)$, $P_5 = 3(24), 23(6)$,$P_6 = 2(15), 12(15)$ 등 6개의 T 패턴들이 사용되었다. 이와 같은 T패턴들은 직관적으로 낮은 불균형의 정도에서 높은 불균형의 정도를 갖는 것을 알 수 있으며, 고려된 모든 T 패턴에서 총 표본의 수는 210으로 고정하였다. 각 실험은 1000번의 반복이 실시되었으며, 식 (3.1)의 귀무가설에 대한 LM통계량과 비교를 위하여 우도비(LR) 검정통계량의 기각횟수를 계산하였다.

4.2. 모의실험 결과

표 4.1은 (3.1)의 귀무가설에 대하여 고려된 6개의 불균형 T패턴에 대한 모의실험 결과를 나타낸다. 표 4.1의 결과를 살펴보면, LM검정은 모든 불균형 패턴에서 명목유의수준 $\alpha = 0.05$를 제대로 유지하는 것으로 나타났다. 반면 LR검정통계량은 명목유의수준을 약간 과소추정하고 있음을 알 수 있다. 비율에 대한 신뢰구간의 방법을 적용하면 1000번의 반복을 통하여 기각횟수가 37에서 63 사이에 존재하는 경우, 신뢰수준 0.95에서 이는 기각횟수 50회와 큰 차이가 없음을 의미한다. 각 검정법의 검정력은 ρ_1이 증가할 때 높아졌다. 사실, $\rho_1 \geq 0.2$인 경우 모든 고려된 불균형 T 패턴에서 각 검정방법은 95나타났다. 또한 $\rho_2 = 0$인 경우라면, 즉 실제모형은 일원오차성분모형인 경우에 이원오차성분을 가정한 모형에서 얻어진 검정통계량에 의한 검정인 과대검정(over specified test)인 경우에도 검정력에 큰 영향을 주지 않는 것으로 나타났다. 더불어 ρ_1이 고정된 경우, ρ_2가 증가하면 σ_λ^2은 커지고 동시에 오차분산 σ_ν^2값은 작아지게 된다. 즉 ρ_2가 커지면 동일한 σ_μ^2 수준에서 σ_ν^2값이 작아지게 되므로, ρ_2가 커질때 σ_μ^2에 대한 검정에서 검정력이 높아지게 된다고 판단된다. 또한 두 검

정방법의 검정력을 비교해보면, 고려된 모든 불균형 패턴에서 LM검정법의 검정력이 LR검정에 비하여 높게 나타나는 경향을 보여주었다.

표 4.1: 1000번 반복 실험에서 검정통계량의 기각횟수

		P_1		P_2		P_3		P_4		P_5		P_6	
ρ_1	ρ_2	LM	LR	LM	LR	LM	LR	LM	LR	LM	LR	LM	LR
0.00	0.00	42	36	48	32	45	28	49	21	52	35	53	30
0.00	0.05	42	26	54	33	57	41	52	24	59	37	55	38
0.00	0.10	44	33	43	35	59	31	45	40	63	42	52	39
0.00	0.20	68	41	67	29	72	46	74	37	77	42	52	39
0.00	0.40	59	46	49	42	57	36	46	37	58	32	62	33
0.00	0.60	48	44	49	39	49	40	50	34	57	46	58	44
0.00	0.80	58	47	53	40	52	37	62	34	49	47	70	35
0.05	0.00	341	253	323	234	318	254	323	253	323	258	353	262
0.05	0.05	338	259	345	266	353	274	367	272	363	282	349	278
0.05	0.10	358	275	382	306	374	298	363	281	372	285	399	305
0.05	0.20	423	349	402	330	433	360	443	359	426	360	422	339
0.05	0.40	568	498	567	505	595	508	578	507	522	463	554	487
0.05	0.60	794	748	792	744	778	719	776	713	788	729	779	730
0.05	0.80	990	985	991	988	981	977	984	981	987	979	985	982
0.10	0.00	655	600	662	613	708	609	676	580	686	598	642	606
0.10	0.05	702	618	703	623	686	608	701	630	712	648	712	629
0.10	0.10	740	689	732	678	767	657	758	654	733	646	740	676
0.10	0.20	792	754	770	754	800	718	816	741	769	702	787	730
0.10	0.40	919	878	917	897	908	832	928	885	893	879	922	895
0.10	0.60	993	990	993	985	986	978	990	989	982	975	991	986
0.10	0.80	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000
0.20	0.00	961	937	960	948	966	956	961	952	961	955	981	949
0.20	0.05	970	960	968	952	981	964	971	961	976	964	973	961
0.20	0.10	983	958	988	975	980	972	985	972	973	960	972	970
0.20	0.20	997	987	991	991	986	984	992	986	989	983	993	982
0.20	0.40	998	998	997	999	999	999	1000	998	1000	999	999	997
0.20	0.60	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000
0.40	0.00	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000
0.40	0.05	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000
0.60	0.00	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000
0.60	0.05	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000
0.80	0.00	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000

5. 결론

본 논문에서는 불균형 이원오차성분모형에서 분산성분의 존재유무에 대한 가설검정을 다루었다. 본 모형에서 장애모수가 존재하는 경우에 적용할 수 있는 LM검정통계량을 유도하고 그의 효율성을 모의실험을 통하여 파악하였다. 모의실험 결과 LR검정법은 명목유의수준을 과소 추정하는 경향을 보여주었고, 제안된 LM검정은 명목유의수준을 잘 유지하고 있는 것으로 나타났다. 또한 검정력에 있어서도 점근적 LM검정법이 LR검정에 비하여 높

게 나타나는 경향을 보여주었다. 이와 같은 결과를 토대로 본 모형에 대해서는 LM검정법이 LR검정에 비하여 좀 더 나은 검정방법임을 알 수 있었다.

참고문헌

송석헌, 최충돈 (2003). 패널회귀모형에서 선형성검정, <응용통계연구>, **16**, 351-364.

정병철, 조민화, 송석헌 (2001). 패널회귀모형에서의 예측량의 효율에 관한 비교, <응용통계연구>, **14**, 121-135.

Ara, I. and King, M.L.(1993). Marginal likelihood based tests of the regression disturbances, *Working papers*, Monash University.

Baltagi, B.H. (2001). *Econometric Analysis of Panel Data*, Wiley, Chichester.

Baltagi, B.H. Song, S.H. and Jung, B.C.(2001). The unbalanced nested error component regression model, *Journal of Econometrics*, **101**, 357-381.

Breusch, T.S. and Pagan, A.R. (1980). The Lagrange multiplier test and its application to model specification in econometrics, *Review of Economic Studies*, **47**, 239-254.

Khuri, A.I., Mathew, T. and Sinha, B.K.(1998). *Statistical Tests for Mixed Linear Models*, Wiley, New York.

King, M. L.(1980). Robust tests for spherical symmetry and their application to least squares regression, *Annals of Statistics*, **8**, 1265-1271.

Mathew, T. and Sinha, B.K.(1988). Optimum tests for fixed effects and variance components in balanced models, *Journal of the American Statistical Association*, **83**, 133-135.

Moulton, B.R. and Randolph, W.C. (1989). Alternative tests of the error components model, *Econometrica*, **57**, 685-693.

Nerlove, M. (1971). Further evidence on the estimation of dynamic economic relations from a time-series of cross-sections, *Econometrica*, **39**, 359-382.

Swallow, W.H. and Searle, S.R.(1978). Minimum variance quadratic unbiased estimation(MIVQUE) of variance components, *Technometrics*, **20**, 265-272.

[2004년 1월 접수, 2004년 7월 채택]

Test in Unbalanced Panel Regression Model with Nuisance Parameter*

JaeWon Lee[1)] Byoung Cheol Jung[2)] Seuck Heun Song[3)]

Abstract

This paper consider the testing problem of variance component for the unbalanced two-way error component model with nuisance parameter. We derive the one-sided LM test statistic for testing zero individual(time) effects assuming that the other time-specific(individual) effects are present. Using the Monte Carlo experiments, the computational more demanding LR test slightly underestimates the nominal size and has the low powers relative to LM test statistic.

Keywords: Unbalanced panel regression, Nuisance parameter, LM test

* This work was supported by Korea Research Foundation Grant(KRF-2002-042-C00008)

1) Professor, Dept of Statistics, Korea University, Annam-Dong 5-1, Seoul 136-701, Korea
E-mail: jelee@korea.ac.kr
2) Dept of Statistics, Sungshin Women's University, Dongsun-Dong 249-1, Seoul 136-742, Korea
E-mail: bcjung@korea.ac.kr
3) Associate Professor, Dept of Statistics, Korea University, Annam-Dong 5-1, Seoul 136-701, Korea
E-mail: ssong@korea.ac.kr