응용통계연구 제 16권 2호, 2003년, pp. 351-364

패널회귀모형에서 선형성검정 *

송석헌[1) 최충돈[2)

요 약

본 논문에서는 오차성분을 가지는 패널회귀모형에서 모형의 선형성을 검정 할 수 있는 검정통계량을 제시하고, 유도한 검정통계량의 계산을 위하여 인공회귀방법을 이용하려한다. 모의실험 결과, Double-Length Artificial Resression(DLR)을 이용한 LM 검정통계량은 명목유의수준을 잘 유지하고 있는 것으로 나타났으며 검정력에 있어서도 기존의 검정에 비하여 높게 나타났다.

주요용어: 패널회귀모형, 모형식별, LM검정, 인공회귀

1. 서론

최근 여러 학문분야에서는 패널자료의 다양한 통계적 분석방법을 통하여, 각 분야에 직면한 여러 문제점을 이론적으로 모형화하고, 이를 분석, 적용하려는 연구가 활발히 진행되고 있다(Baltagi(2002), Matyas와 Sevestre(1996)). 특히 패널회귀모형에 대한 최근의 연구사례는 Baltagi, Song과 Jung(2001, 2002a, 2002b), Song과 Trenkler(2001)등이 있다. 그러나 이들의 연구와 더불어 대부분의 패널회귀모형에 대한 기존의 분석방법은 회귀모형의 선형성을 가정하고 랜덤효과들의 존재여부를 검정하는 방법과 랜덤효과들의 존재에 따른 분산성분과 회귀계수의 추정방법 등에 관한 연구가 진행되어 왔다. 그러나 실제 경영, 경제학, 환경학, 의학 등의 제반문제에서는 선형회귀모형을 가정하기에는 자료에 따라 어려운 경우가 많이 존재한다. 예를 들어 경제학에서의 Cobb-Douglas의 생산함수 등은 선형모형으로 다루기에는 무리가 따른다. 따라서 연구자들은 패널자료에 대하여 선형 패널회귀모형으로 가정하여 통계적 추론을 해야할지 비선형 패널회귀모형으로 다루어져야 하는 지에 대한 문제에 봉착하게 된다. 더불어 패널회귀모형에서는 함수의 형태에 관한 식별뿐만 아니라, 랜덤효과들의 존재에 따른 모형식별의 문제를 고려해야한다. 따라서 본 논문에서는 패널회귀모형에서 적절한 회귀모형의 함수의 형태(선형/비선형, 로그선형/비선형 등)와 랜덤효과들의 존재여부를 동시에 검정할 수 있는 검정통계량을 유도하려한다. 또한 동시검정의 결과 귀무가설이 기각되었을 경우, 그 결과가 랜덤효과의 존재 때문인지 또는 비선형의 효과 때문인지 더 이상의 추론이 불가능하다. 이러한 동시검정의 약점을 극복하기 위하여, 본 연구에서는 랜덤효과가 존재한다는 가정하에서 모형의 선형성여부를 검정하는 조건부

* 본 연구는 KRF(2001-041-D00046)연구비 지원에 의하여 수행되었음.
1) (136-701) 서울시 성북구 안암동 5-1 고려대학교 통계학과, 부교수,
E-mail : ssong@korea.ac.kr
2) (100-180) 서울시 중구 249-1 LG화재 CRM팀.

검정(conditional test)에 대한 검정 통계량을 유도하고 이러한 검정통계량들의 소표본에서의 성질을 모의실험을 통해서 보이려한다.

본 논문의 2장에서는 Box-Cox 변환된 패널회귀모형을 다룬다. 3장에서는 동시검정과 조건부 검정을 위하여 Lagrange Multiplier(LM) 검정통계량을 유도한다. 그러나 유도한 검정통계량에 사용되는 정보행렬의 추정에 어려움이 있어 검정통계량의 계산을 위하여 정보행렬의 일치추정량으로 Outer-Product Gradient(OPG)와 Double-Length Artificial Resression(DLR)을 이용한 LM 검정통계량을 제시하였다. 4장에서는 모의실험을 통해서 각 검정통계량의 유의수준과 검정력을 비교하였으며, 마지막으로 본 연구의 결과를 5장에 정리하였다.

2. Box-Cox 변환된 패널회귀모형

다음의 패널회귀모형을 고려해 보자.

$$y_{it} = x'_{it}\beta + u_{it}, \qquad i = 1, \cdots, N, \qquad t = 1, \cdots, T \tag{2.1}$$

여기서 y_{it}는 i번째 개체(개인, 가구 등)의 시점 t에서의 반응값이고, x_{it}는 k개의 변수로 이루어진 설명변수 벡터이며, 오차항 u_{it}는 다음과 같은 오차성분모형을 따른다고 가정한다.

$$u_{it} = \mu_i + \nu_{it} \tag{2.2}$$

여기에서 μ_i는 개체효과(individual specific effect)를 나타내는 확률변수이며, ν_{it}는 나머지 오차항을 나타낸다. μ_i와 ν_{it}는 서로 독립이며 각각 $\mu_i \sim i.i.d.(0, \sigma_\mu^2)$와 $\nu_{it} \sim i.i.d.(0, \sigma_\nu^2)$ 라고 가정하자. 모형 (2.1)을 행렬을 이용하여 나타내면 다음과 같다.

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

y는 $NT \times 1$인 반응변수벡터, X는 $NT \times k$인 독립변수행렬이며, β는 $k \times 1$인 회귀계수벡터이다. 또한 식 (2.2)를 행렬로 표현하면 다음과 같다.

$$u = (I_N \otimes i_T)\mu + \nu \tag{2.4}$$

여기서 I_N는 차원이 N인 단위행렬이고, i_T는 모든 원소가 1로 이루어진 크기가 T인 벡터이며, $\mu = (\mu_1, \cdots, \mu_N)'$, $\nu = (\nu_{11}, \cdots, \nu_{NT})'$이며 $\otimes$는 크로네커곱을 나타낸다. 이와 같은 가정으로부터 오차항의 공분산 행렬은 다음과 같이 구해진다.

$$\Omega = \sigma_\mu^2(I_N \otimes J_T) + \sigma_\nu^2(I_N \otimes I_T) \tag{2.5}$$

J_T은 모든 원소가 1인 $T \times T$인 행렬이다.

모형 (2.1)은 선형패널회귀모형을 나타낸다. 하지만 이와같은 선형성을 가정할 수 없는 경우에는 다음과 같은 Box-Cox 변환(Box and Cox (1964))된 패널회귀모형을 고려해야 한다.

$$B(y_{it}, \lambda) = \sum_{k=1}^{K} \beta_k B(X_{itk}, \lambda) + u_{it} \tag{2.6}$$

여기서 $B(y_{it}, \lambda)$는 다음과 같은 Box-Cox 변환을 나타낸다.

$$B(y_{it}, \lambda) = \begin{cases} (y_{it}^{\lambda} - 1)/\lambda, & \lambda \neq 0 \\ log\ y_{it}, & \lambda = 0 \end{cases}$$

여기서 y_{it}, X_{itk}는 Box-Cox 변환이 사용되므로 언제나 양의 값을 가진다. 모형 (2.6)에서 $\lambda = 1$ 인 경우에는 패널회귀모형이 선형모형이 되며, $\lambda = 0$ 인 경우는 로그선형모형이 된다.

3. DLR 검정통계량과 OPG 검정통계량

모형 (2.6)에서 우도비검정을 이용하여 선형과 로그선형을 구별하는 모형식별이 가능하다. 그러나 우도비검정은 제한된 모형과 제한되지 않은 모형에서의 추정을 필요로 하기 때문에 계산적으로 매우 복잡하다. 때문에 모형의 식별을 위하여 단지 제한된 모형에서의 최대우도추정만을 요구하는 Lagrange Multiplier(LM) 검정을 이용하려한다. 이러한 LM검정통계량도 모형(2.6)의 경우에는 여러 개의 모수가 비선형의 관계를 형성하고 있어서 LM 검정법에 사용되는 정보행렬의 추정은 대단히 어렵다. 따라서 본 장에서는 이러한 문제의 해결을 위해서 정보행렬의 일치추정량으로 Godfrey와 Wickens(1981)가 제안한 Outer-Product Gradiant(OPG)를 이용한 LM 검정과 Davidson과 MacKinnon(1984, 1988, 1990)이 제안한 Double-Length Artificial Resresssion(DLR)에 근거한 LM 검정법들을 유도하려한다.

3.1. Box-Cox 변환된 패널모형에서 DLR 검정통계량

모형 (2.6)에서 오차항의 공분산행렬이 식 (2.5)와 같이 상관되어있다. 이와 같은 상관성을 제거하기 위하여 Fuller와 Battese(1974)의 변환을 이용하여 모형 (2.6)을 다시 표현하면 다음과 같다.

$$B^*(y_{it}, \lambda) = \sum_{k=1}^{K} \beta_k B^*(X_{itk}, \lambda) + u_{it}^* \tag{3.1}$$

여기서
$$B^*(y_{it}, \lambda) = B(y_{it}, \lambda) - \theta \sum_{t=1}^{T} B(y_{it}, \lambda)/T,$$
$$B^*(X_{itk}, \lambda) = B(X_{itk}, \lambda) - \theta \sum_{t=1}^{T} B(X_{itk}, \lambda)/T, \quad u_{it}^* = u_{it} - \theta \sum_{t=1}^{T} u_{it}/T,$$
$$\theta = 1 - (\sigma_\nu/\sigma_1), \quad \sigma_1^2 = T\sigma_\mu^2 + \sigma_\nu^2, \quad u_{it}^* \sim n.i.d.(0, \sigma_\nu^2)$$

DLR의 적용을 위해서(이에 대한 자세한 설명은 Davidson과 MacKinnon(1984, 1988)을 참조), 식 (3.1)을 다음의 형태로 다시 표현한다.

$$f_{it}(y_{it}, \phi) = \varepsilon_{it}, \quad i = 1, 2, \cdots, N,\ t = 1, 2, \cdots, T \tag{3.2}$$

여기서 $\phi = (\beta, \lambda, \theta, \sigma_\nu)$이고 $\varepsilon_{it} \sim n.i.d.(0,1)$이다. 이것은 식(3.1)의 양변을 σ_ν로 나눔으로써 다음과 같이 얻어진다.

$$f_{it}(y_{it}, \phi) = \Big[B^*(y_{it}, \lambda) - \sum_{k=1}^{K} \beta_k B^*(X_{itk}, \lambda)\Big] / \sigma_\nu \tag{3.3}$$

식(3.3)에서 it번째 관측치의 로그우도함수에 대한 기여도는 다음과 같다.

$$l_{it} = constant - \frac{1}{2} f_{it}^2(y_{it}, \phi) + J_{it}(y_{it}, \phi) \tag{3.4}$$

여기서 $J_{it}(y_{it}, \phi) = log|\partial f_{it}(y_{it}, \phi)/\partial y_{it}|$는 y_{it}를 ε_{it}로 변환시킬때 발생하는 자코비안항이다.

다음과 같이 정의 하자

$$F_{itj}(y_{it}, \phi) = \frac{\partial f_{it}(y_{it}, \phi)}{\partial \phi_j} \quad , \quad J_{itj}(y_{it}, \phi) = \frac{\partial J_{it}(y_{it}, \phi)}{\partial \phi_j} \tag{3.5}$$

더불어 $F_{itj}(y_{it}, \phi), J_{itj}(y_{it}, \phi)$를 원소로 하는 $NT \times (K+3)$행렬을 각각 $F(y, \phi)$와 $J(y, \phi)$라 정의하자. 또한 $f(y, \phi)$와 i_{NT}는 각각 원소가 $f_{it}(y_{it}, \phi)$와 1인 $NT \times 1$벡터로 정의하자. 그러면 $2NT$개의 관찰치(이는 최초의 관측치의 두배의 길이)를 갖는 DLR 회귀모형은 다음과 같이 표현할 수 있다(Davidson과 MacKinnon(1984)).

$$\begin{bmatrix} f(y, \phi) \\ i_{NT} \end{bmatrix} = \begin{bmatrix} -F(y, \phi) \\ J(y, \phi) \end{bmatrix} b + Residuals \tag{3.6}$$

즉, 다음과 같이 표현된다.

$$r(\phi) = R(\phi) b + Residuals \tag{3.7}$$

Davidson과 MacKinnon(1984)은 각각의 관측치가 로그우도함수에 대해서 두 개의 기여도를 만든다고 하였다. 그 중 하나는 제곱합항 $-\frac{1}{2} f_{it}^2$ 이고, 나머지 하나는 자코비안항 J_{it}이다.

이 같은 DLR의 속성중에 하나는 귀무가설하에서 모수ϕ의 최대우도 추정치가 $\widehat{\phi}$로 추정될 때, LM 검정통계량이 식(3.6)의 회귀제곱합(SSR)으로 유도된다는 점이다. 식(3.6)에서 귀무가설 하에서의 총제곱합(SST)은 언제나 $2NT$이다. 그러므로 LM 검정통계량은 총제곱합-잔차제곱합(SSE);($2NT$-SSE)로 얻을 수 있으며 이는 다음의 LM 검정통계량과 동일하다.

$$LM_{DLR} = [R(\phi)' r(\phi)]' [R(\phi)' R(\phi)]^{-1} [R(\phi)' r(\phi)] \tag{3.8}$$

식(3.8)에 나타난 DLR에 근거한 LM 검정통계량은 귀무가설 하에서 근사적으로 χ^2 분포를 따르며 자유도는 제한된 모수의 수와 동일하다.

식 (3.3)으로부터, it번째 관측치를 위한 자코비안항은 다음과 같이 주어진다.

$$J_{it}(y_{it},\phi) = log|\frac{\partial f_{it}(y_{it},\phi)}{\partial y_{it}}| = \frac{1}{T}log(1-\theta) + (\lambda-1)log(y_{it}) - log(\sigma_\nu)$$

위의 결과로 자코비안항의 Regressors를 아래와 같이 얻을 수 있다.

$$\frac{\partial J_{it}(y_{it},\phi)}{\partial \beta_k} = 0, \quad \frac{\partial J_{it}(y_{it},\phi)}{\partial \sigma_\nu} = -\frac{1}{\sigma_\nu},$$

$$\frac{\partial J_{it}(y_{it},\phi)}{\partial \theta} = -\frac{1}{T(1-\theta)}, \quad \frac{\partial J_{it}(y_{it},\phi)}{\partial \lambda} = log(y_{it})$$

또한, 식 (3.3)으로 부터 제곱합의 Regressors를 얻을 수 있다.

$$\frac{\partial f_{it}(y_{it},\phi)}{\partial \beta_k} = -\frac{B^*(X_{itk},\lambda)}{\sigma_\nu}, \quad \frac{\partial f_{it}(y_{it},\phi)}{\partial \sigma_\nu} = -\frac{u_{it}^*}{\sigma_\nu^2}, \quad \frac{\partial f_{it}(y_{it},\phi)}{\partial \theta} = -\frac{-\sum_{t=1}^T u_{it}/T}{\sigma_\nu},$$

$$\begin{aligned}\frac{\partial f_{it}(y_{it},\phi)}{\partial \lambda} &= [\partial B^*(y_{it},\lambda)/\partial\lambda - \sum_{k=1}^K \beta_k \partial B^*(X_{itk},\lambda)/\partial\lambda]/\sigma_\nu \\ &= \frac{1}{\sigma_\nu}\{[C(y_{it},\lambda) - \theta\sum_{t=1}^T C(y_{it},\lambda)/T] - \sum_{k=1}^K \beta_k[C(X_{itk},\lambda) - \theta\sum_{t=1}^T C(X_{itk},\lambda)/T]\}\end{aligned}$$

여기서 $C(y_{it},\lambda) = [\lambda y_{it}^\lambda log(y_{it}) - (y_{it}^\lambda - 1)]/\lambda^2$, $C(X_{itk},\lambda) = [\lambda X_{itk}^\lambda log(X_{itk}) - (X_{itk}^\lambda - 1)]/\lambda^2$,

u_{it}^*는 식(3.3)의 오차항과 동일하다.

식 (3.1)에서 패널선형회귀모형($\lambda = 1$)이며 개체효과가 존재하지 않는 경우($\theta = 0$)와 패널로그선형모형($\lambda = 0$)이며 개체효과가 존재하지 않는 경우($\theta = 1$)에 대하여 각각 DLR을 이용한 동시 LM 검정통계량(LM_{DLR})들을 유도하기 위하여 식(3.8)의 $R(\theta)$의 원소를 구하면 다음 표와 같다.

표 3.1. 동시검정을 위한 $F(y,\phi)$와 $J(y,\phi)$

	모수	$\theta = 0$ and $r = 1$	$\theta = 0$ and $r = 0$
$F(y,\phi)$	β_k	$(X_{itk}-1)/\widehat{\sigma}_\nu$	$\log(X_{itk})/\widehat{\sigma}_\nu$
	σ_ν	$\widehat{\nu}_{it}/\widehat{\sigma}_\nu^2$	$\widehat{\nu}_{it}/\widehat{\sigma}_\nu^2$
	θ	$\frac{\widehat{\bar{\nu_{i.}}} - \sum_{k=1}^K (1-\widehat{\beta}_k)}{\widehat{\sigma}_\nu}$	$\widehat{\bar{\nu_{i.}}}/\widehat{\sigma}_\nu$
	λ	$\frac{C(y_{it},1) - \sum_{k=1}^K C(X_{itk},1)\widehat{\beta}_k}{\widehat{\sigma}_\nu}$	$\frac{C(y_{it},0) - \sum_{k=1}^K C(X_{itk},0)\widehat{\beta}_k}{\widehat{\sigma}_\nu}$
$J(y,\phi)$	β_k	0	0
	σ_ν	$-1/\widehat{\sigma}_\nu$	$-1/\widehat{\sigma}_\nu$
	θ	$1/T$	$1/T$
	λ	$\log(y_{it})$	$\log(y_{it})$

여기서 $\widehat{\bar{\nu_{i.}}} = \sum_{t=1}^T \widehat{\nu}_{it}/T$이고 $C(y_{it},1) = y_{it}log(y_{it}) - (y_{it}-1)$ 이며 $C(y_{it},0) = log(y_{it})^2$이다. 각각의 LM_{DLR} 검정통계량은 귀무가설 하에서 근사적으로 $\chi^2(2)$를 따른다.

식 (3.1)이 패널선형회귀모형($\lambda = 1$)인 경우와 패널로그선형모형($\lambda = 0$)인 경우에 대하여 각각 DLR을 이용한 조건부 LM 검정통계량(LM_{DLR})들을 유도하여보자. 첫번째로 식 (3.1)에서 ($\lambda = 1$)인 경우, 식(3.3)은 다음과 같다.

$$y_{it}^* = \sum_{k=1}^{K} \beta_k X_{itk}^* + u_{it}^* \tag{3.9}$$

여기서 $y_{it}^* = (y_{it} - \theta \sum_{t=1}^{T} y_{it}/T)$, $X_{itk}^* = X_{itk} - \theta \sum_{t=1}^{T} X_{itk}/T$

이 경우 DLR 회귀모형의 Regressors는 처음에 NT개의 인공인 관측치를 가지며, 그 원소는 $\widehat{u_{it}^*}/\widehat{\sigma_\nu}$이다. 그리고 나머지 NT개의 관측치는 1이다. $\widehat{u_{it}^*}$는 Fuller-Battese 변환된 모형(3.9)에서의 최대우도추정치로 부터 얻어진 it 번째 잔차이다. 그리고 $\widehat{\sigma_\nu}$또한 귀무가설에서 얻어진 최대우도추정치이다. Regressor의 첫 번째 관측치 NT개와 두 번째 관측치 NT개는 다음과 같이 얻을 수 있다.

$$\begin{aligned}
&\text{for } \beta_k \;:\; \frac{1}{\widehat{\sigma}_\nu}[(X_{itk} - \widehat{\theta}\sum_{t=1}^{T} X_{itk}/T) - (1-\widehat{\theta})] \quad \text{and} \quad 0\ , \\
&\text{for } \sigma_\nu \;:\; \frac{\widehat{u}_{it}^*}{\widehat{\sigma}_\nu^2} \quad \text{and} \quad -\frac{1}{\widehat{\sigma}_\nu}, \\
&\text{for } \theta \;:\; \frac{1}{\widehat{\sigma}_\nu}\Big[(\sum_{t=1}^{T} \widehat{u}_{it}/T)\Big] \quad \text{and} \quad -\frac{1}{T(1-\widehat{\theta})}, \\
&\text{for } \lambda \;:\; \frac{1}{\widehat{\sigma}_\nu}\Big\{\sum_{k=1}^{K}\Big[\beta_k C(X_{itk},1) - \widehat{\theta}\sum_{t=1}^{T} C(X_{itk},1)/T\Big] - \Big[C(y_{it},1) - \widehat{\theta}\sum_{t=1}^{T} C(y_{it},1)/T\Big]\Big\} \\
&\qquad \text{and } log(y_{it})
\end{aligned}$$

여기서 $C(X_{itk},1) = X_{itk}log(X_{itk}) - (X_{itk}-1)$, $C(y_{it},1) = y_{it}log(y_{it}) - (y_{it}-1)$.

검정통계량 LM_{DLR}은 $2NT - SSE$의 값과 동일하며, 이것은 귀무가설 하에서 근사적으로 $\chi^2(1)$을 따른다.

또한 식 (3.1)이 로그선형모형인 경우($\lambda = 0$), 식 (3.3)는 다음과 같다.

$$log^*(y_{it}) = \sum_{k=1}^{K} \beta_k log^*(X_{itk}) + u_{it}^* \tag{3.10}$$

여기서 $log^*(y_{it}) = log(y_{it}) - \theta \sum_{t=1}^{T} log(y_{it})/T$, $\quad log^*(X_{itk}) = log(X_{itk}) - \theta \sum_{t=1}^{T} log(X_{itk})/T$

이 경우, 위에서와 마찬가지로 DLR의 Regressand는 처음에 NT개의 인공인 관측치를 가지며, 그 원소는 $\widehat{u}_{it}^*/\widehat{\sigma}_\nu$이다. 그리고 나머지 NT개의 관측치는 1이다. $\widehat{u}_{it}^*$는 Fuller-Battese의 변환된 로그선형모형에서의 최대우도추정치로 부터 얻어진 it 번째 잔차이다. 그리고 $\widehat{\sigma_\nu}$는 귀무가설에서 얻어진 σ_ν의 최대우도추정치이다. Regressors의 첫 번째 관측치 NT개와 두 번째 관측치 NT개는 다음과 같이 얻을 수 있다.

$$\text{for } \beta_k \ : \ \frac{1}{\hat{\sigma}_\nu} log(X_{itk}) - \hat{\theta}\sum_{t=1}^{T} log(X_{itk})/T \quad \text{and} \quad 0,$$
$$\text{for } \sigma_\nu \ : \ \frac{\hat{u}^*_{it}}{\hat{\sigma}^2_\nu} \quad \text{and} \quad -\frac{1}{\hat{\sigma}_\nu},$$
$$\text{for } \theta \ : \ \frac{1}{\hat{\sigma}_\nu}\Big[(\sum_{t=1}^{T}\hat{u}_{it}/T)\Big] \quad \text{and} \quad -\frac{1}{T(1-\hat{\theta})},$$
$$\text{for } \lambda \ : \ \frac{1}{\hat{\sigma}_\nu}\sum_{k=1}^{K}\hat{\beta}_k C(X_{itk},0) - \hat{\theta}\sum_{t=1}^{T} C(X_{itk},0)/T - C(y_{it},0) - \hat{\theta}\sum_{t=1}^{T} C(y_{it},0)/T \text{ and } log(y_{it}),$$

여기서, 로피탈(L.Hospital)정리를 이용하면

$$C(y_{it},0) = \lim_{\lambda\to 0} C(y_{it},\lambda) = \lim_{\lambda\to 0}[\lambda y_{it}^{\lambda} log(y_{it}) - (y_{it}^{\lambda} - 1)]/\lambda^2 = [log(y_{it})]^2/2,$$
$$C(X_{itk},0) = \lim_{\lambda\to 0} C(X_{itk},\lambda) = \lim_{\lambda\to 0}[\lambda X_{itk}^{\lambda} log(X_{itk}) - (X_{itk}^{\lambda} - 1)]/\lambda^2 = [log(x_{itk})]^2/2.$$

또한 LM_{DLR}은 $2NT - SSE$의 값과 동일하며, 이는 귀무가설 하에서 근사적으로 $\chi^2(1)$을 따른다.

3.2. Box-Cox 변환된 패널모형에서 OPG 검정통계량

위에서 언급한 DLR과 마찬가지로 Outer-Product Gradient(OPG)는 인공회귀모형의 한 종류이며 Godfrey와 Wickens(1981)에 의해 검정통계량을 계산하기 위해서 처음 사용되었다. 이 역시 인공회귀의 성질을 만족하므로 DLR과 마찬가지로 정보행렬의 추정량으로 이용이 가능하다. OPG Regression의 구성은 다음과 같다.

$$\iota \ = \ G(\phi)c \ + \ Residuals \tag{3.11}$$

여기서, $G(\phi)$는 식 (3.4)의 우도함수를 각각의 모수에 대해서 일차미분한 값들로 구성된 행렬로 다음과 같이 구해진다.

$$G_{itj}(\phi) = \frac{\partial l_{it}(\phi)}{\partial \phi_j}, \ \ i = 1,...,N, \ \ t = 1,...,T, \ \ j = 1,..,k+3 \tag{3.12}$$

ι_{NT} 는 각각의 원소가 1인 $NT \times 1$ 벡터이고, $g(\theta)$를 모든 원소가 G_{itj}로 이루어진 $NT \times (K+3)$인 행렬이라면, Godfrey와 Wickens(1981)의 결과에 의해 다음과 같은 인공회귀를 만들 수 있다. OPG에 근거한 LM 검정통계량을 계산할 수 있다.

$$\iota_{NT} \ = \ G(\phi)c \ + \ Residuals \tag{3.13}$$

그러므로 OPG에 근거한 LM 검정통계량을 식 (3.13)의 인공회귀에서 회귀제곱합(SSR)을 이용하여 다음과 같이 계산할 수 있다.

$$LM_{OPG} \ = \ [G(\hat{\phi})'\iota_{NT}]'[G(\hat{\phi})'G(\hat{\phi})]^{-1}[C(\phi)'\iota_{NT}] \tag{3.14}$$

여기서 $G(\hat{\phi})$는 귀무가설에서 ϕ에 대한 ML 추정치 $\hat{\phi}$를 식 (3.12)에 대입하여 얻어진 행렬을 나타낸다. 식 (3.14)에서 얻어진 OPG에 근거한 LM 검정통계량은 귀무가설 하에서 근사적으로 $\chi^2(1)$을 따른다.

4. 모의실험

본 장에서는 3장에서 유도된 DLR에 근거한 LM 검정통계량(LM_{DLR})과 OPG에 근거한 LM 검정통계량(LM_{OPG})의 효율성을 비교하고자 한다. 모의실험에 사용된 모형은 다음과 같다.

$$y_{it}^{(\lambda)} = \gamma Z_{it} + \beta x_{it}^{(\lambda)} + u_{it}, \quad i = 1, \cdots, N, \; t = 1, \cdots, T. \tag{4.1}$$

$$u_{it} = \mu_i + \nu_{it} \tag{4.2}$$

여기서 설명변수 x_{it}는 균일분포(5, 15)에서 발생시켰으며 Z는 상수항을 나타낸다. 회귀계수와 상수항은 각각 $\beta = 2$와 $\lambda = 1$로 고정하였으며 μ_i는 $N(0, \sigma_\mu^2)$에서 ν_{it}는 $N(0, \sigma_\nu^2)$에서 각각 발생시켰다. 이때 전체분산을 $\sigma^2 = \sigma_\mu^2 + \sigma_\nu^2$이라 했을때 $\sigma^2 = 1$로 고정시켰으며 $w_1 = \sigma_\mu^2/\sigma^2$ 의 값을 0.2에서 0.8까지 0.2단위로 변화시켜가며 실험하였다. 또한 비선형 모수를 나타내는 λ의 값은 0에서 1까지 0.2단위로 변화시켜가며 실험하였으며, 각 실험에서 y_{it}는 다음과 같은 식에 의하여 얻어졌다.

$$\begin{aligned} y_{it} &= \left(1 + r(\gamma Z_{it} + \beta x_{it}^{(\lambda)} + u_{it})\right)^{1/\lambda} && \text{if } \lambda \neq 0 \\ y_{it} &= \exp\left(\gamma Z_{it} + \beta \log(x_{it}) + u_{it}\right) && \text{if } \lambda = 0. \end{aligned} \tag{4.3}$$

모든 실험은 1000번 독립적으로 반복 실시하였으며, SAS/IML 프로시져를 이용하여 수행하였다. 본 실험에 사용된 표본은 (N,T)의 값을 각각 (10,10) (20,10) (20,20)으로 변화시켜가며 실시되었으며 모든 w_1과 r의 조합에서 다음과 같은 4가지 가설에 대하여 DLR와 OPG에 근거한 LM 검정통계량을 계산하고 명목유의수준 0.05하에서 1000번의 반복을 통하여 추정된 유의수준과 검정력을 계산하였다.

$$\begin{aligned} H_0^a &: \lambda = 1 \text{ and } \sigma_\mu^2 = 0 \\ H_0^b &: \lambda = 0 \text{ and } \sigma_\mu^2 = 0 \\ H_0^c &: \lambda = 1 \text{ (given } \sigma_\mu^2 > 0) \\ H_0^d &: \lambda = 0 \text{ (given } \sigma_\mu^2 > 0) \end{aligned} \tag{4.4}$$

H_0^a는 개체효과의 존재유무와 선형성을 동시에 검정하는 가설을 나타내고, H_0^c는 개체효과가 존재한다는 가정하에서 선형성을 검정하고자하는 조건부 검정을 나타낸다. 또한 H_0^b는 개체효과의 존재유무와 로그선형성을 동시에 검정하는 가설을 나타내며 H_0^d는 개체효과가 존재한다는 가정하에서 로그선형성을 검정하고자하는 조건부 검정을 나타낸다.

4.1. 모의실험 결과

표 4.1은 표본수에 따른 선형성 검정인 H_0^a(동시검정)와 H_0^c(조건부 검정)에 대하여 명목유의수준이 0.05일때 DLR과 OPG의 추정된 유의수준과 검정력을 나타낸다. 먼저 선형성과 개체효과에 대한 동시검정을 나타내는 H_0^a의 경우 DLR의 추정된 유의수준은 $(N,T) = (10,10)$인 경우 0.0555 및 $(N,T) = (20,20)$인 경우 0.0650 을 나타내어 명목유의수준을 제

대로 유지하고 있는 것을 알 수 있다. 반면 OPG의 경우는 추정된 유의수준이 $(N,T)=(10,10)$인 경우 0.1035 및 $(N,T)=(20,20)$인 경우 0.0790 을 나타내어 명목유의수준을 과대추정하는 것을 보여주고 있다. 동시 검정의 검정력 면에서는 λ가 1에서 멀어질수록 w_1이 0보다 커질수록 두 검정법 모두 높은 수준의 검정력을 보여주고 있다. 표 4.1의 H_0^c는 개체효과가 존재한다는 가정하에서 선형성에 대한 검정결과를 나타낸다. 그러므로 첫 번째 블록은 두 검정방법의 추정된 유의수준을 나타낸다. 이의 결과를 살펴보면, $(N,T)=(10,10)$ 또는 $(N,T)=(20,20)$인 경우 w_1의 값이 어떤 값을 갖는다 할지라도 DLR의 추정된 유의수준은 명목유의수준을 제대로 유지하는것으로 나타났다. 반면 OPG의 경우 $(N,T)=(10,10)$인 모의실험에서 명목유의수준을 과대추정하는 결과를 보여주고 있다(0.0775-0.0965). 하지만 표본수가 $(N,T)=(20,20)$인 경우가 되면 OPG도 명목유의수준을 제대로 유지하는것으로 나타나고 있다. 이와 같은 조건부 검정의 추정된 검정력은 두 검정법 모두 λ의 값이 1에서 멀어질수록(즉, 선형모형에서 멀어질수록) 증가하였다. 특히 $\lambda=0.4$보다 작은 경우, 두 검정법 모두 검정력이 1로 나타나 모형의 선형성을 기각하는 것으로 나타났다.

표 4.2는 표본수에 따른 로그선형성 검정인 H_0^b(동시검정)와 H_0^d(조건부 검정)에 대하여 명목유의수준이 0.05일때 DLR과 OPG의 추정된 유의수준과 검정력을 나타낸다. 로그선형성과 개체효과에 대한 동시검정을 나타내는 H_0^b의 경우 DLR의 추정된 유의수준은 $(N,T)=(10,10)$인 경우 0.0570 및 $(N,T)=(20,20)$인 경우 0.0515를 나타내어 명목유의수준을 제대로 유지하고 있는 것을 알 수 있다. 반면 OPG의 경우는 추정된 유의수준이 $(N,T)=(10,10)$인 경우 0.0995 및 $(N,T)=(20,20)$인 경우 0.0790 을 나타내어 명목유의수준을 과대추정하는 것을 보여주고 있다. 동시 검정의 검정력 면에서는 λ가 1에서 멀어질수록 w_1이 0보다 커질수록 두 검정법 모두 높은 수준의 검정력을 보여주고 있다. 표 4.2의 H_0^d는 개체효과가 존재한다는 가정하에서 모형의 로그선형성에 대한 검정결과를 나타낸다. 그러므로 맨 마지막 블록은 두 검정방법의 추정된 유의수준을 나타낸다. 이의 결과를 살펴보면, $(N,T)=(10,10)$ 또는 $(N,T)=(20,20)$인 경우 w_1의 값이 어떤 값을 갖는다 할지라도 DLR의 추정된 유의수준은 명목유의수준을 제대로 유지하는것으로 나타났다. 반면 OPG의 경우 $(N,T)=(10,10)$인 모의실험에서 명목유의수준을 과대추정하는 결과를 보여주고 있다(0.0685-0.1085). 하지만 표본수가 $(N,T)=(20,20)$인 경우가 되면 OPG도 w_1이 0과 0.2인 경우를 제외하곤 명목유의수준을 제대로 유지하는것으로 나타나고 있다. 이와 같은 조건부 검정의 추정된 검정력은 두 검정법 모두 λ의 값이 0에서 멀어질수록(즉, 로그선형모형에서 멀어질수록) 증가하였다. 특히 $\lambda=0.2$보다 큰 경우, 두 검정법 모두 검정력이 0.8보다 큰 것으로 나타나 아주 높은 수준의 검정력을 보여주고 있다. 이와 같은 모의실험을 통하여 OPG의 경우 명목유의수준을 과대추정하는 문제점이 있으나, DLR의 경우에는 명목유의수준을 제대로 유지하고 검정력도 높게 나타난다.

5. 결 론

본 연구에서는 Box-Cox 변환된 패널회귀모형에서 모형식별을 위하여 우도비 검정통계량을 이용하는 경우 이를 구하는 과정은 대단히 복잡하고 어렵기 때문에 귀무가설 하에

표 4.1: 표본수에 따른 선형성 검정(H_0^a, H_0^b)의 추정된 유의수준 및 검정력

		(N, T)=(10, 10)				(N, T)=(20, 20)			
		H_0^a(Joint)		H_0^c(Conditional)		H_0^a(Joint)		H_0^c(Conditional)	
λ	w_1	DLR	OPG	DLR	OPG	DLR	OPG	DLR	OPG
1.0	0.0	0.0555	0.1035	0.0550	0.0890	0.0650	0.0790	0.0510	0.0615
1.0	0.2	0.7820	0.8225	0.0520	0.0965	0.9990	0.9985	0.0520	0.0625
1.0	0.4	0.9825	0.9870	0.0535	0.0920	1.0000	1.0000	0.0460	0.0530
1.0	0.6	1.0000	1.0000	0.0595	0.0855	1.0000	1.0000	0.0415	0.0435
1.0	0.8	1.0000	1.0000	0.0540	0.0775	1.0000	1.0000	0.0530	0.0500
0.8	0.0	0.0945	0.1830	0.3425	0.4260	0.2955	0.3430	0.5050	0.5485
0.8	0.2	0.7945	0.8330	0.1375	0.2235	1.0000	1.0000	0.3875	0.4275
0.8	0.4	0.9815	0.9845	0.1400	0.2030	1.0000	1.0000	0.4490	0.4810
0.8	0.6	0.9985	0.9995	0.1580	0.2445	1.0000	1.0000	0.5515	0.5630
0.8	0.8	1.0000	1.0000	0.2430	0.2870	1.0000	1.0000	0.7205	0.7140
0.6	0.0	0.2985	0.4650	0.5335	0.6595	0.8825	0.9255	0.9465	0.9660
0.6	0.2	0.8580	0.8970	0.3640	0.5210	1.0000	1.0000	0.9315	0.9530
0.6	0.4	0.9815	0.9850	0.4255	0.5555	1.0000	1.0000	0.9445	0.9570
0.6	0.6	0.9980	0.9985	0.5215	0.5965	1.0000	1.0000	0.9855	0.9875
0.6	0.8	1.0000	1.0000	0.6385	0.7280	1.0000	1.0000	0.9985	0.9985
0.4	0.0	0.7040	0.8630	0.8480	0.9260	1.0000	1.0000	1.0000	1.0000
0.4	0.2	0.9475	0.9730	0.8075	0.9135	1.0000	1.0000	0.9995	1.0000
0.4	0.4	0.9915	0.9955	0.8095	0.8960	1.0000	1.0000	1.0000	1.0000
0.4	0.6	0.9995	1.0000	0.8825	0.9335	1.0000	1.0000	1.0000	1.0000
0.4	0.8	1.0000	1.0000	0.9560	0.9745	1.0000	1.0000	1.0000	1.0000
0.2	0.0	0.9955	0.9990	0.9990	0.9995	1.0000	1.0000	1.0000	1.0000
0.2	0.2	1.0000	1.0000	0.9985	1.0000	1.0000	1.0000	1.0000	1.0000
0.2	0.4	1.0000	1.0000	0.9980	1.0000	1.0000	1.0000	1.0000	1.0000
0.2	0.6	1.0000	1.0000	0.9975	1.0000	1.0000	1.0000	1.0000	1.0000
0.2	0.8	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
0.0	0.0	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
0.0	0.2	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
0.0	0.4	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
0.0	0.6	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
0.0	0.8	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000

표 4.2: 표본수에 따른 로그선형성 검정(H_0^c, H_0^d)의 추정된 유의수준 및 검정력

		(N, T)=(10, 10)				(N, T)=(20, 20)			
		H_0^b(Joint)		H_0^d(Conditional)		H_0^b(Joint)		H_0^d(Conditional)	
λ	w_1	DLR	OPG	DLR	OPG	DLR	OPG	DLR	OPG
1.0	0.0	0.9200	0.9395	0.9710	0.9775	1.0000	1.0000	1.0000	1.0000
1.0	0.2	0.9870	0.9915	0.9480	0.9610	1.0000	1.0000	1.0000	1.0000
1.0	0.4	0.9990	0.9995	0.9860	0.9915	1.0000	1.0000	1.0000	1.0000
1.0	0.6	1.0000	1.0000	0.9945	0.9955	1.0000	1.0000	1.0000	1.0000
1.0	0.8	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
0.8	0.0	0.7875	0.8270	0.8995	0.9150	1.0000	1.0000	1.0000	1.0000
0.8	0.2	0.9805	0.9880	0.9315	0.9505	1.0000	1.0000	1.0000	1.0000
0.8	0.4	0.9990	0.9995	0.9455	0.9595	1.0000	1.0000	1.0000	1.0000
0.8	0.6	1.0000	1.0000	0.9450	0.9485	1.0000	1.0000	1.0000	1.0000
0.8	0.8	1.0000	1.0000	0.9940	0.9940	1.0000	1.0000	1.0000	1.0000
0.6	0.0	0.5940	0.6680	0.7860	0.8420	0.9975	0.9975	0.9990	0.9995
0.6	0.2	0.9330	0.9500	0.7280	0.7870	1.0000	1.0000	1.0000	1.0000
0.6	0.4	0.9935	0.9970	0.7630	0.8085	1.0000	1.0000	0.9995	0.9995
0.6	0.6	0.9995	0.9995	0.8405	0.8730	1.0000	1.0000	1.0000	1.0000
0.6	0.8	1.0000	1.0000	0.9185	0.9160	1.0000	1.0000	1.0000	1.0000
0.4	0.0	0.4175	0.5215	0.6255	0.6965	0.9505	0.9540	0.9790	0.9845
0.4	0.2	0.8745	0.9025	0.4695	0.533	1.0000	1.0000	0.9690	0.9715
0.4	0.4	0.9820	0.9875	0.5155	0.5685	1.0000	1.0000	0.9815	0.9805
0.4	0.6	0.9985	1.0000	0.6055	0.6380	1.0000	1.0000	0.9920	0.9915
0.4	0.8	1.0000	1.0000	0.7415	0.7615	1.0000	1.0000	0.9995	0.9995
0.2	0.0	0.2070	0.3120	0.4395	0.5235	0.6695	0.6955	0.8080	0.8260
0.2	0.2	0.8170	0.8560	0.2770	0.3720	1.0000	1.0000	0.8065	0.8110
0.2	0.4	0.9810	0.9855	0.2720	0.3330	1.0000	1.0000	0.8135	0.8070
0.2	0.6	0.9985	0.9990	0.3160	0.3685	1.0000	1.0000	0.8945	0.8790
0.2	0.8	1.0000	1.0000	0.4835	0.4895	1.0000	1.0000	0.9515	0.9405
0.0	0.0	0.0570	0.0995	0.0515	0.1015	0.0565	0.0790	0.0525	0.0740
0.0	0.2	0.7765	0.8175	0.0475	0.1085	0.9995	0.9990	0.0530	0.0715
0.0	0.4	0.9765	0.9800	0.0485	0.0975	1.0000	1.0000	0.0455	0.0540
0.0	0.6	0.9990	0.9990	0.0440	0.0725	1.0000	1.0000	0.0440	0.0500
0.0	0.8	1.0000	1.0000	0.0450	0.0685	1.0000	1.0000	0.0485	0.0495

서의 최대우도추정량을 이용하는 LM 검정법을 선택하였다. 그러나 이러한 LM 검정방법 또한 정보행렬을 구하는 과정에서 여러 개의 모수가 서로 비선형의 형태를 가지고 있으므로 이 역시 간단하지 않다. 이러한 문제를 해결하기 위해서 여러 가지 정보행렬의 일치추정량들 중에서 OPG와 DLR에 근거한 LM 검정통계량 유도하여 패널회귀모형에 있어서 선형/비선형 혹은 로그선형/비선형의 모형식별과 랜덤효과의 존재유무에 관한 검정에 대해서 동시검정과 조건부검정을 수행할 수 있도록 하였다. 모의실험의 결과 소표본의 경우에는 OPG에 근거한 LM 검정통계량은 과대기각(over rejection)하는 경향이 있는 반면 DLR의 경우에는 명목유의수준을 제대로 유지하고 검정력도 높게 나타난다. 따라서 본 연구에서는 DLR에 근거한 LM 검정통계량을 적극 추천하고자 한다. 더불어 이러한 결과는 보다 복잡한 패널 회귀모형으로의 확장이 가능하리라 생각되며, 이를 통하여 적절한 패널회귀모형의 선택은 물론이고 효율적인 모수 추정방법과 예측방법의 개발이 가능하게 된다.

참고문헌

[1] Baltagi, B.H. (2002). *Econometric Analysis of Panel Data*, John Wiley, New York.

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

[3] Baltagi, B.H., Song. S.H. and Jung. B.C. (2002a). Simple LM tests for unbalanced nested error component regression model, *Econometric Reviews*, Vol. 21, 167-187.

[4] Baltagi, B.H., Song. S.H. and Jung. B.C. (2002b). A comparative study of alterative estimators for the unbalanced two-way error component regression model, *The Econometrics Journal*, Vol.5, 480-493.

[5] Box, G.E.P. and Cox D.R. (1964). An analysis of transformations, *Journal of The Royal Statistical Society Series*, B 26, 211-252.

[6] Davidson, R. and MacKinnon, J.G. (1984). Model specification test based on artificial regression, *International Economic Review*, Vol. 25, 485-502.

[7] Davidson, R. and MacKinnon, J.G. (1985). Testing linear and loglinear regressions against Box-Cox alternatives, *Canadian Journal of Economics*, Vol. 18, 499-517.

[8] Davidson, R. and MacKinnon, J.G. (1988). Double-Length artificial regression, *Oxford Bulletin of Economics and Statistics*, Vol. 50, 203-217.

[9] Davidson, R. and MacKinnon, J.G. (1990). Specification tests based on artificial regressions, *Journal of the American Statistical Association*, Vol. 85, 220-227.

[10] Fuller, W. A. and Battese, G.E. (1974). Estimation of linear models with cross-error structure, *Journal of Econometrics*, Vol. 2, 67-68.

[11] Godfrey, L. G. and Wickens, M.R. (1981). Testing linear and log-linear regressions for functional form, *Review of Economic Studies*, Vol. 48, 487-496.

[12] Màtyàs, L. and Sevestre, P. (1996). *The Econometrics of Panel Data: handbook of theory and applications,* Kluwer Academic Publishers, Dordrecht.

[13] Song, S.H. and Trenkler, G. (2001). On the efficiency of the Cochrane-Orcutt estimator in the serially correlated error components regression model for panel data, *Communications in Statistics, Theory and Methods*, Vol. 30, 195-208.

[2003년 3월 접수, 2003년 6월 채택]

Test of Linearity in Panel Regression Model*

Seuck Heun Song[1] Chung Don Choi[2]

ABSTRACT

This paper derives Lagrange multiplier tests based on Double-Length Artificial Regression and Outer-Product Gradient for testing linear and log-linear panel regressions against Box-Cox alternatives. The proposed DLR based LM tests are easy to implement in an error component model. From the Monte Carlo study, the DLR based LM tests are recommended for testing functiona forms.

Keywords: Panel regression model; model specification; LM tests, artificial regression

* This work was supported by Korea Research Foundation Grant(KRF(2001-041-D00046))

1) Associate Professor, Dept of Statistics, Korea University, Seoul
E-mail : ssong@korea.ac.kr
2) CRM Team, LG Insurancd Co., Seoul 100-180