The Korean Communications
in Statistics Vol. 11 No. 3, 2004
pp. 475-484

Testing for Grouped Heteroscedasticity in Linear Regression Model[1)]

Seuck Heun Song[2)] and Moon Kyung Choi[3)]

Abstract

This paper consider the testing problem of grouped heteroscedasticity in the linear regression model. We provide the Lagrange Multiplier(LM), Wald, Likelihood Ratio(LR) test statistis for testing of grouped heteroscedasticity. Monte Carlo experiments are conducted to study the performance of these tests.

Keywords : Grouped heteroscedasticity, LM, Wald, LR test

1. 서 론

통상적인 선형회귀모형에서는 오차항이 등분산임을 가정하고 적절한 통계적인 추론을 한다. 그러나 현실적으로 오차항이 이분산인 경우가 많고, 이러한 경우 등분산 가정 하에서의 통계적인 추론은 적절하지 않을 수 있다. 따라서 선형회귀모형에서 오차항의 이분산성에 대한 검정은 Goldfeld와 Quandt(1965), Harvey와 Phillips(1974), Breusch와 Pagan(1979) 등에 의하여 연구되어져 왔다. 그러나 좀 더 포괄적인 형태로써 두 개 이상의 그룹 간에 이분산이 존재하는 경우에 대한 연구는 아직 미흡한 상태이다, 예를 들어 기업의 크기 또는 가계의 소득수준이나 개인의 교육정도 등과 같이 관찰치에 내포되어 있는 일정한 특성에 의하여 관찰치를 구분할 때 발생되어지는 각 그룹의 분산들이 이분산을 갖는 경우를 그룹화 이분산(grouped heteroscedasticity)이라 한다. 실제적으로 그룹화 이분산은 요즘과 같이 체계화되고, 세분화 되어있는 여러 분야의 자료에서 발생될 가능성이 높다. 따라서 본 연구에서는 선형회귀모형에서 오차항에 그룹화 이분산의 존재를 검정하기 위하여 LM검정, LR검정과 Wald검정 통계량을 유도하고 이러한 세 가지 검정통계량의 검정력을 모의실험을 통하여 비교하고자 한다.

2. 회귀모형과 이분산 검정통계량

다음과 같은 선형회귀모형을 고려하자.

---

1) This research was supported by a Korea University Grant 2003.
2) Associate Professor, Department of Statistics, Korea University, Seoul, 136-701, korea
E-mail: ssong@korea.ac.kr
3) Assistant Manager, M.I. Vally, Youngdeungpo Ku, Seoul, 150-874, korea

$$y = X\beta + e \tag{2.1}$$

여기서 y는 $T\times 1$인 종속변수벡터이고, X는 $T\times k$인 독립변수행렬이며, β는 $k\times 1$인 회귀계수벡터이고, e는 $T\times 1$인 오차벡터이다. 또한 $rank(X)=k\leq T$이며, 오차벡터 e는 i.i.d이고 $Cov(e)=\sigma^2 I_T$ 인 정규분포를 따른다고 가정한다. 회귀모형에서 관찰치들이 $E(e)=0$ 두 개 이상의 그룹으로 나뉘어져 있는 경우 오차항의 분산이 그룹 간에 다르다면 이를 그룹화 이분산 이라 한다. 이를 일반화하면 회귀모형에서 T개의 전체 관찰치들이 G개의 그룹으로 나뉘어 진다고 하자. 각 그룹의 개체 수는 $T_1, T_2, \cdots, T_G$로 알려져 있으며, 이때 $\sum_{r=1}^{G} T_r = T$ 이고, 모든 그룹에 대하여 $T_r > k$ 이라고 가정하자. 모형 (2.1)에서 y는 $y' = (y_1', \cdots, y_G')$, X는 $X' = (X_1', \cdots, X_G')$, e는 $e' = (e_1', e_2', \cdots, e_G')$ 으로 표현할 수 있다. 이때 β는 모든 그룹에 대해서 동일하다고 가정한다. 모형 (2.1)에서 r번째 그룹에 대하여 나타내면 다음과 같은 선형회귀모형이 된다.

$$y_r = X_r\beta + e_r, \qquad r = 1, 2, \ldots, G \tag{2.2}$$

여기서 y_r는 $T_r\times 1$인 벡터이고, X_r는 $T_r\times k$인 행렬이고 e_r는 $T_r\times 1$인 벡터로 $E(e_r) = 0$이고, $E(e_r e_r') = \sigma_r^2 I_{T_r}$이다. 따라서 모형 (2.1)에서 분산공분산행렬 Ω는 다음과 같은 행렬이 된다.

$$E(ee') = \Omega = \begin{bmatrix} \sigma_1^2 I_{T_1} & 0 & \cdots & 0 \\ 0 & \sigma_2^2 I_{T_2} & & 0 \\ \vdots & & \ddots & \vdots \\ 0 & 0 & \cdots & \sigma_G^2 I_{T_G} \end{bmatrix} \tag{2.3}$$

여기서, σ_r^2은 r번째 그룹내에서 오차분산이고, I_{T_r}는 $T_r\times T_r$인 단위행렬이다.

그룹화 이분산을 검정하기 위한 귀무가설과 대립가설은 다음과 같다.

$$H_0 : \Omega = \sigma^2 I, \qquad H_1 : \Omega \neq \sigma^2 I \tag{2.4}$$

귀무가설에서는 $\sigma_1^2 = \sigma_2^2 = \cdots = \sigma_G^2 = \sigma^2$이라는 등분산을 가정하고, 대립가설에서는 적어도 하나 이상의 그룹화 이분산 $\sigma_r^2 \neq \sigma^2$이 존재한다고 가정한다.

정리 2.1: 모형 (2.1)에서 그룹 이분산을 검정하기 위한 LM 검정통계량은 다음과 같이 구해진다.

$$LM = \frac{1}{2}\sum_{r=1}^{G} T_r \left(1 - \frac{\widehat{\sigma_r}^2}{\widehat{\sigma}^2}\right)^2 \tag{2.5}$$

여기서, $\widehat{\sigma_r}^2 = \widehat{e_r}'\widehat{e_r}/T_r$는 r번째 그룹의 분산추정량이며, $\widehat{\sigma}^2$은 귀무가설 하에서 얻어진 최대우도추정량을 나타낸다.

증명: 모형 (2.1)에 대한 로그우도함수는 다음과 같이 주어진다.

$$\log L_1(y|X, \beta, \sigma_1^2, \cdots, \sigma_G^2) = -\frac{T}{2}\log(2\pi) - \frac{1}{2}\log|\Omega| - \frac{1}{2}(y - X\beta)'\Omega^{-1}(y - X\beta) \tag{2.6}$$

여기서 Ω는 식 (2.3)과 같다. 식 (2.6)의 로그우도함수를 이용하여 가설 (2.4)에 대한 LM검정통계량을 유도하기 위해서는 귀무가설 하에서 구해지는 β와 σ^2에 대한 일차미분값과 정보행렬이 필요하다(Buse, 1982). 이를 위해 Hartley와 Rao(1967), Hemmerle와 Hartley (1973)는 다음과 같은 행렬 연산을 제시하였다.

Lemma 2.2: 오차항이 정규분포로 가정된 로그우도함수에서 모수에 대한 일차미분벡터와 정보행렬은 다음과 같은 식을 이용한다. 로그우도함수는

$$\log L(\theta) = constant - \frac{1}{2}\log|\Omega| - \frac{1}{2}e'\Omega^{-1}e$$

이며, θ는 분산공분산행렬 Ω를 구성하는 p개의 성분으로 이루어진 모수벡터이다.

θ벡터의 i번째 원소 θ_i에 대한 로그우도함수의 일차미분벡터는 다음과 같다.

$$\partial \log L/\partial\theta_i = -\frac{1}{2}tr[\Omega^{-1}(\partial\Omega/\partial\theta_i)] + \frac{1}{2}e'\Omega^{-1}(\partial\Omega/\partial\theta_i)\Omega^{-1}e \tag{2.7}$$

또한 θ에 대한 정보행렬의 (i, j)번째 원소는 다음과 같다.

$$J_{ij} = E[-\partial^2 \log L/\partial\theta_i\partial\theta_j] = \frac{1}{2}tr[\Omega^{-1}(\partial\Omega/\partial\theta_i)\,\Omega^{-1}(\partial\Omega/\partial\theta_j)],\quad i, j = 1, 2, \cdots, p \tag{2.8}$$

증명: Hartley와 Rao(1967)와 Hemmerle와 Hartley(1973)를 참조하라.

식 (2.7)을 이용하여 얻어지는 일차미분벡터 식에 $\hat{\theta} = (\hat{\beta}, \hat{\sigma}^2)$을 대입하면

$$\frac{\partial \log L_1}{\partial\beta}\Big|_{\hat{\theta}} = X'\Omega^{-1}y - X'\Omega^{-1}X\beta\Big|_{\hat{\theta}} = 0$$

이 되며,

$$\frac{\partial \log L_1}{\partial\sigma_r^2}\Big|_{\hat{\theta}} = -\frac{T_r}{2\sigma_r^2} + \frac{e_r'e_r}{2\sigma_r^4}\Big|_{\hat{\theta}} = -\frac{T_r}{2\,\hat{\sigma}^2}\left[1 - \frac{\hat{e_r}'\hat{e_r}/T_r}{\hat{\sigma}^2}\right]$$

이다. 여기서, $\hat{e_r}$는 r번째 그룹의 OLS잔차이다. 따라서 일차미분벡터를

$$S(\hat{\theta}) = \{S(\hat{\beta}),\ S(\hat{\sigma}_1^2),\ S(\hat{\sigma}_2^2),\ \cdots,\ S(\hat{\sigma}_G^2)\} \tag{2.9}$$

로 표기하면 $S(\hat{\beta}) = 0$이고,

$$S(\hat{\sigma}_r^2) = -\frac{T_r}{2\,\hat{\sigma}^2}\left[1 - \frac{\hat{e_r}'\hat{e_r}/T_r}{\hat{\sigma}^2}\right],\quad r = 1,\ 2,\ \cdots,\ G$$

이다. 또한, 로그우도함수가 식 (2.6)과 같은 경우 정보행렬은

$$I = \begin{bmatrix} X'\Omega^{-1}X & 0 \\ 0 & J \end{bmatrix} \tag{2.10}$$

이고, J 는 원소가 $J_{rs} = E\left[-\frac{\partial^2 \log L}{\partial\sigma_r^2\,\partial\sigma_s^2}\right]$, $r, s = 1, 2, \cdots, G$인 $G\times G$인 행렬이다.

원소 $J_{r,s}$는 Lemma 2.2의 식 (2.8)을 이용하면 다음과 같다.

$$J_{r,s} = \frac{1}{2}\, tr[\Omega^{-1}(\partial\Omega/\partial\sigma_r^2)\Omega^{-1}(\partial\Omega/\partial\sigma_s^2)] \tag{2.11}$$

그룹화 이분산 행렬 Ω는 대각행렬이므로 $r \neq s$인 경우는 $J_{rs}=0$이 되며,

$$\Omega^{-1}\left(\frac{\partial\Omega}{\partial\sigma_r^2}\right)\Omega^{-1}\left(\frac{\partial\Omega}{\partial\sigma_s^2}\right) = \begin{bmatrix} 0 & & \cdots & & 0 \\ & \ddots & & & \\ & & 0 & & \\ \vdots & & & \sigma_r^{-4} I_{T_r} & & \vdots \\ & & & & 0 & \\ & & & & & \ddots \\ 0 & & \cdots & & & 0 \end{bmatrix} \tag{2.12}$$

의 트레이스를 구하면 $r=s$ 인 경우에는 $J_{r,s} = \dfrac{T_r}{2\sigma_r^4}$이 된다. 따라서 그룹화 이분산인 경우의 정보행렬은 다음과 같다.

$$I = \begin{bmatrix} X'\Omega^{-1}X & 0 & 0 & \cdots & 0 \\ 0 & \dfrac{T_1}{2\sigma_1^4} & 0 & \cdots & 0 \\ \vdots & \vdots & & \ddots & 0 \\ 0 & 0 & \cdots & 0 & \dfrac{T_G}{2\sigma_G^4} \end{bmatrix} \tag{2.13}$$

식 (2.13)을 이용하여 귀무가설 $\Omega=\sigma^2 I$하에서 정보행렬을 계산하면 다음과 같이 주어진다.

$$I(\widehat{\theta}) = \begin{bmatrix} \dfrac{X'X}{\widehat{\sigma}^2} & 0 & 0 & \cdots & 0 \\ 0 & \dfrac{T_1}{2\,\widehat{\sigma}^4} & 0 & \cdots & 0 \\ \vdots & \vdots & & \ddots & 0 \\ 0 & 0 & \cdots & 0 & \dfrac{T_G}{2\,\widehat{\sigma}^4} \end{bmatrix} \tag{2.14}$$

따라서 LM검정통계량은 다음과 같다.

$$LM = S(\widehat{\theta})' I(\widehat{\theta})^{-1} S(\widehat{\theta}) = \frac{1}{2}\sum_{r=1}^{G} T_r \left(1 - \frac{\widehat{\sigma_r}^2}{\widehat{\sigma}^2}\right)^2 \tag{2.15}$$

여기서, $\widehat{\sigma_r}^2 = \widehat{e_r}'\widehat{e_r}/T_r$로 r번째 그룹의 분산추정량이며, $\widehat{\sigma}^2$은 제약적인 최대우도추정량으로 다음과 같다.

$$\widehat{\sigma}^2 = \frac{(y-X\widehat{\beta})'(y-X\widehat{\beta})}{T}, \quad \widehat{\beta} = (X'X)^{-1}X'y.$$

LM검정통계량은 귀무가설 하에서 근사적으로 자유도가 G-1인 카이제곱분포를 따른다.

정리 2.3: 모형 (2.1)에서 그룹 이분산을 검정하기 위한 LR 검정 통계량은 다음과 같이 구해진다.

$$l = \frac{L(\widehat{\theta})}{L(\widetilde{\theta})} = \frac{\widetilde{\sigma_1}^{T_1}\,\widetilde{\sigma_2}^{T_2}\cdots\widetilde{\sigma_G}^{T_G}}{\widehat{\sigma}^{T}}$$

여기서, $\widetilde{\sigma_r^2}$ 는 비제약적인 최대우도추정량을 나타낸다.

증명: β와 σ_r^2의 비제약적인 최대우도추정량을 구하기 위해서는 모수에 대하여 로그우도함수 (2.6)을 일차 미분한 결과가 필요하다. 식 (2.6)의 로그우도함수를 β에 대하여 일차미분한 결과는

$$\frac{\partial \log L_1}{\partial \beta} = X'\Omega^{-1}y - X'\Omega^{-1}X\beta \qquad)$$

이고, $\partial \log L_1/\partial\beta = 0$를 만족하는 β의 최대우도추정량은

$$\widetilde{\beta} = (X'\Omega^{-1}X)^{-1}X'\Omega^{-1}y$$

이다. 그러나 일반적으로 Ω는 알려지지 않으므로 Ω의 일치추정치 S를 이용한 다음과 같은 β의 Feasible Generalized Least Squares(FGLS)추정량을 사용한다.

$$\widetilde{\widetilde{\beta}} = (X'S^{-1}X)^{-1}X'S^{-1}y \tag{2.16}$$

여기서 S는 다음과 같고

$$S = \begin{bmatrix} \widetilde{\sigma}_1^{\,2} I_{T_1} & 0 & \cdots & 0 \\ 0 & \widetilde{\sigma}_2^{\,2} I_{T_2} & & \vdots \\ \vdots & & \ddots & \\ 0 & \cdots & 0 & \widetilde{\sigma}_G^{\,2} I_{T_G} \end{bmatrix} \tag{2.17}$$

이를 구하기 위해서 G개의 대각원소 σ_r^2를 추정해야 한다.

로그우도함수 (2.6)을 σ_r^2에 대하여 일차미분한 결과는

$$\frac{\partial \log L_1}{\partial \sigma_r^2} = -\frac{T_r}{2\sigma_r^2} + \frac{1}{2\sigma_r^4}(y_r - X_r\beta)'(y_r - X_r\beta)$$

이며, $\partial \log L_1/\partial\sigma_r^2 = 0$으로 만족하는 σ_r^2의 최대우도추정량은 다음과 같이 얻어진다.

$$\widetilde{\sigma}_r^{\,2} = \frac{(y_r - X_r\widetilde{\beta})'(y_r - X_r\widetilde{\beta})}{T_r}, \quad r = 1, 2, \cdots, G \tag{2.18}$$

식 (2.18)을 구하기 위해서는 최대우도추정치 $\widetilde{\beta}$가 필요하다. 그러나 $\widetilde{\beta}$은 Ω가 미지의 값이기 때문에 구할 수가 없다. 따라서 $\partial \log L_1/\partial\beta = 0$, $\partial \log L_1/\partial\sigma_r^2 = 0$을 동시에 만족시키는 최대우도추정량을 구해야 한다. 이는 비선형 관계이므로 Newton-Raphson과 같은 수치해석적인 방법을 사용하여 구할 수 있다. FGLS추정량 $\widetilde{\widetilde{\beta}}$를 구하기 위한 초기치로 OLS추정량을 이용하며, 이를 식 (2.18)에 대입하여 $\widetilde{\sigma}_r^{\,2}$를 구한다. 다시 $\widetilde{\sigma}_r^{\,2}$를 식 (2.17)에 대입하여 $\widetilde{\widetilde{\beta}}$를 구한다. 이러한 과정을 수렴할 때까지 반복하여 최종적인 추정치 $\widetilde{\widetilde{\beta}}$와 $\widetilde{\sigma}_g^{\,2}$를 구한다. 그러므로 본 연구에서는 FGLS추정량을 사용하여 다음과 같이 $\widetilde{\sigma}_r^{\,2}$을 구할 것이다.

$$\widetilde{\sigma}_r^{\,2} = \frac{(y_r - X_r\widetilde{\widetilde{\beta}})'(y_r - X_r\widetilde{\widetilde{\beta}})}{T_r}, \quad r = 1, 2, \cdots, G \tag{2.19}$$

따라서 β와 σ_r^2의 비제약적인 최대우도추정량 식 (2.16)과 (2.19)를 식 (2.6)에 대입하면 다음과 같다.

$$L(\widetilde{\theta}) = \frac{1}{(2\pi)^{T/2}|S|^{1/2}} \exp\left\{-\frac{1}{2}(y - X\widetilde{\widetilde{\beta}})'S^{-1}(y - X\widetilde{\widetilde{\beta}})\right\} \tag{2.20}$$

따라서 LR 검정통계량은

$$l=\frac{L(\widehat{\theta})}{L(\widetilde{\theta})}=\frac{\widetilde{\sigma}_1^{\,T_1}\,\widetilde{\sigma}_2^{\,T_2}\cdots\,\widetilde{\sigma}_G^{\,T_G}}{\widehat{\sigma}^{\,T}} \tag{2.21}$$

이고, 다음과 같은 LR검정통계량을 유도한다.

$$LR \;=\; -2\log\, l \;=\; 2\left[\,T\log\widehat{\sigma}-(\,T_1\log\widetilde{\sigma}_1+T_2\log\widetilde{\sigma}_2+\cdots+T_G\log\widetilde{\sigma}_G)\right] \tag{2.22}$$

이는 귀무가설 하에서 근사적으로 자유도가 G-1인 카이제곱분포를 따른다.

마지막으로 Wald검정통계량은 다음과 같이 구할 수 있다.

$$W=(\widetilde{\theta}-\theta_0)'I(\widetilde{\theta})(\widetilde{\theta}-\theta_0) \tag{2.23}$$

여기서 $\widetilde{\theta}$는 대립가설 하에서의 θ이다. 즉 $\widetilde{\theta}=(\widetilde{\beta},\ \widetilde{\sigma}_1^{\,2},\cdots,\ \widetilde{\sigma}_G^{\,2})$이다. 또한 정보행렬은 식 (2.13)에 $\widetilde{\theta}$를 대입하면 다음과 같다.

$$I=\begin{bmatrix} X'S^{-1}X & 0 & 0 & \cdots & 0 \\ 0 & \dfrac{T_1}{2\widetilde{\sigma}_1^4} & 0 & \cdots & 0 \\ \vdots & \vdots & & \ddots & 0 \\ 0 & 0 & \cdots & 0 & \dfrac{T_G}{2\widetilde{\sigma}_G^4} \end{bmatrix} \tag{2.24}$$

여기서 대각원소 $\widetilde{\sigma}_r^{\,2}$는 식 (2.19)과 같다.

$$\widetilde{\sigma}_r^{\,2}=\frac{(y_r-X_r\widetilde{\beta})'(y_r-X_r\widetilde{\beta})}{T_r},\quad r=1,2,\cdots,G \tag{2.25}$$

이를 Wald검정통계량 식 (2.23)에 대입하면

$$W=(\widetilde{\beta}-\widehat{\beta})'X'S^{-1}X(\widetilde{\beta}-\widehat{\beta})+\frac{1}{2}\sum_{g=1}^{G}T_r\left(1-\frac{\widehat{\sigma}^2}{\widetilde{\sigma}_r^{\,2}}\right)^2$$

이며 이는 귀무가설 하에서 근사적으로 자유도가 G-1인 카이제곱분포를 따른다.

3. 모의 실험

이 장에서는 그룹화 이분산 검정을 위하여 유도한 LR, LM, Wald검정통계량을 이용하여 세 가지 검정의 검정력을 모의실험을 통하여 비교하고자 한다. 모의실험모형은 다음과 같은 단순선형회귀모형이다.

$$y=\beta_0 1+\beta_1 X_1+e \tag{3.1}$$

여기서, y는 $T\times 1$인 종속변수벡터이며, **1**은 모든 원소가 1인 $T\times 1$단위벡터이다. 또한 X_1은 하나의 독립변수를 가진 $T\times 1$ 독립변수벡터이며, β_0와 β_1은 상수인 모수이고, e는 $T\times 1$인 오차벡터로써 정규분포N(0, 1)에서 발생시켰다. 모의실험을 위한 대립가설은 두 가지 형태로 첫 번째 가설은

$$H_1:\ \sigma_r^2=\delta^{\,r-1}\sigma^2\ ,\quad r=1,2,\cdots,G \tag{3.2}$$

이며, δ는 분산의 상대비율 $\sigma^2_{r-1}/\sigma^2_r$이다. 즉, 이 가설은 그룹별 분산이 δ만큼의 일정한 크기로 증가하는 것을 말한다. 실험에서 δ는 0.1에서 1까지 0.2의 간격(δ=0.1, 0.3, 0.5, 0.7, 0.9)으로 변화되며, $\delta=1$인 경우는 각 통계량의 명목유의수준이 계산된 것이다. 이때 독립변수 X_1은 균등분포(0, 1)에서 생성하였다. 두 번째 대립가설은

$$H_2 : \sigma^2_r = \sigma^2 \, \overline{X}_{1r}^{\ 2} \ , \ r=1,2,\cdots,G \tag{3.3}$$

으로, $\overline{X}_{1r}^{\ 2}$는 g번째 그룹내의 독립변수의 평균을 제곱한 것이다. 이 가설은 그룹화 이분산이 그룹내 독립변수의 평균값의 제곱에 비례하는 경우를 말한다. 이를 검정하기 위해서 X_1의 분포가 X_{1A} ~ 균등분포(0, 10), X_{1B} ~ 균등분포(0, 20), X_{1C} ~ 균등분포(0, 30), X_{1D} ~ 균등분포(0, 40) 인 네 가지 경우로 실험을 하였다. 또한 β_0와 β_1의 초기치는 1로 주었다. 전체 표본크기(T)는 30, 50, 60, 90, 120개이며, 그룹의 크기(G)는 2, 3, 4, 5개이다. 그리고 마지막 그룹을 제외한 각 그룹의 개체 수 T_r, $r=1,2,\cdots,G-1$는 전체 표본크기를 그룹 수로 나눈 값의 정수부분으로 동일하며, 마지막 그룹의 개체 수는 $T_G = T - \sum_{r=1}^{G-1} T_r$로 설정하였다. 각 경우의 반복수는 1000회로 하였다.

<표 1>에서 δ=1은 명목유의수준을 0.05로 하는 경우 모의실험을 통해 계산된 LM, LR, Wald검정의 명목유의수준이다. 이항분포의 정규분포 근사를 이용하면 명목유의수준이 0.05인 경우, 1000번의 반복 중에서 모의실험된 유의수준이 0.036보다 작거나 0.064보다 크게 나타날 가능성은 5%미만이다. 이와 같은 관점에서 살펴보면 LM검정은 (T, G)=(50, 3), (50, 4), (60, 4)인 경우를 제외하고, 명목유의수준을 제대로 유지하고 있는 것을 알 수 있다. 그러나 LR검정은 T<120인 경우 명목유의수준을 과대추정하고 있는 것으로 나타났다. 특히, 이는 그룹의 수가 많아질수록 심해지는데 T=30, G=5인 경우에 LR검정의 모의실험된 유의수준은 0.131로 매우 높게 나타났다. Wald검정은 표본수가 작고 그룹수가 많아질수록 LR검정보다 더 심하게 과대추정을 하는 경향이 있다.

<표 1>은 대립가설 식(3.2)에 대한 검정력으로 LM, LR, Wald검정의 검정력은 표본크기(T)와 그룹수(G)가 커질수록 증가하는 경향을 보여주고 있다. 예를 들어, T=50, δ=0.5에서 G=2인 경우 LM검정의 검정력은 0.356인데 반하여 G=5인 경우의 LM검정의 검정력은 0.921로 나타나 그룹수가 커질수록 검정력은 높아진다는 것을 알 수 있다. 이러한 현상은 이분산이 2개인 경우보다는 이분산이 5개인 경우가 귀무가설을 기각할 가능성이 더 크기 때문으로 해석할 수 있다. 그리고 δ가 감소할수록 검정력이 증가하는데 증가량은 세 가지 검정 모두 비슷하다. 전반적으로 LR검정법의 검정력이 LM검정법의 검정력보다 약간 높게 나타나고 있다. 그러나 LR 검정 통계량을 이용하기 위해서는 많은 계산이 요구되어지는 단점이 있다.

<표 2> 는 대립가설 식 (3.3)에 대한 검정력이다. 결과를 살펴보면 전반적으로 표본크기가 증가할수록 모든 검정법들의 검정력이 증가하고 있으나 T=60인 경우 X_{1A}에서의 검정력은 T=50인 경우보다 작은 값을 가지고 있다. 그리고 그룹수가 증가함에 따라 검정력은 전반적으로 증가하지만 일부 경우에서는 증가하다가 점차 감소하는 경향이 나타났다. 그 원인을 살펴보면 이 가설은 이분산이 독립변수의 값에 의존하므로 G개의 그룹간 독립변수의 평균값 $\overline{X}_1$의 일부가 동일하거나 비

슷한 값을 갖는 경우가 발생할 가능성이 있다. 이런 경우 실질적인 이분산의 수와 그룹수가 다르게 되어 검정통계량의 값이 작아지므로 검정력이 감소하는 현상이 나타나는 것이다. 예를 들어

<표 1> LM, LR, Wald검정의 명목유의수준 및 검정력

T \ G		2			3			4			5		
	δ	LM	LR	W	LM	LR	W	LM	LR	W	LM	LR	W
30	1.0	0.038	0.068	0.123	0.039	0.089	0.198	0.041	0.123	0.367	0.040	0.131	0.380
	0.9	0.034	0.068	0.133	0.041	0.102	0.241	0.039	0.131	0.463	0.052	0.148	0.475
	0.7	0.071	0.131	0.245	0.101	0.207	0.467	0.148	0.318	0.775	0.193	0.363	0.840
	0.5	0.194	0.273	0.455	0.366	0.534	0.833	0.486	0.762	0.982	0.634	0.849	0.995
	0.3	0.521	0.653	0.824	0.788	0.934	0.995	0.903	0.995	1.000	0.975	1.000	1.000
	0.1	0.950	0.998	0.998	0.994	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
50	1.0	0.042	0.061	0.097	0.034	0.069	0.138	0.031	0.064	0.181	0.044	0.086	0.211
	0.9	0.040	0.063	0.097	0.043	0.077	0.201	0.055	0.094	0.276	0.063	0.141	0.367
	0.7	0.127	0.169	0.270	0.150	0.239	0.506	0.231	0.347	0.767	0.366	0.512	0.887
	0.5	0.356	0.423	0.588	0.562	0.705	0.936	0.784	0.906	0.955	0.921	0.983	1.000
	0.3	0.793	0.844	0.948	0.963	0.993	1.000	0.993	1.000	1.000	0.999	1.000	1.000
	0.1	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
60	1.0	0.050	0.069	0.091	0.046	0.076	0.121	0.033	0.061	0.133	0.040	0.073	0.163
	0.9	0.048	0.071	0.097	0.058	0.081	0.155	0.057	0.106	0.243	0.076	0.125	0.325
	0.7	0.154	0.182	0.270	0.202	0.274	0.493	0.367	0.454	0.803	0.463	0.594	0.907
	0.5	0.390	0.451	0.613	0.711	0.797	0.949	0.865	0.941	0.997	0.969	0.993	1.000
	0.3	0.855	0.886	0.956	0.994	0.999	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	0.1	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
90	1.0	0.038	0.048	0.061	0.044	0.067	0.092	0.050	0.073	0.117	0.058	0.069	0.140
	0.9	0.074	0.083	0.110	0.064	0.090	0.137	0.062	0.099	0.234	0.910	0.122	0.290
	0.7	0.195	0.222	0.311	0.367	0.413	0.602	0.527	0.618	0.882	0.685	0.755	0.960
	0.5	0.604	0.634	0.756	0.897	0.935	0.990	0.990	0.998	1.000	0.999	1.000	1.000
	0.3	0.982	0.988	0.996	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	0.1	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
120	1.0	0.048	0.056	0.074	0.051	0.058	0.079	0.047	0.052	0.081	0.046	0.066	0.116
	0.9	0.067	0.078	0.102	0.068	0.093	0.134	0.091	0.111	0.217	0.118	0.150	0.293
	0.7	0.266	0.281	0.343	0.447	0.494	0.696	0.675	0.732	0.992	0.844	0.877	0.988
	0.5	0.743	0.761	0.855	0.966	0.978	0.998	0.999	1.000	1.000	1.000	1.000	1.000
	0.3	0.996	0.997	0.998	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	0.1	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000

<표 2> LM, LR, Wald검정의 검정력

T \ G		2			3			4			5		
		LM	LR	W	LM	LR	W	LM	LR	W	LM	LR	W
30	X_{1A}	0.293	0.390	0.819	0.270	0.471	0.917	0.232	0.441	0.962	0.225	0.425	0.954
	X_{1B}	0.213	0.312	0.848	0.225	0.433	0.966	0.215	0.482	0.996	0.234	0.478	0.996
	X_{1C}	0.482	0.632	0.975	0.388	0.616	0.991	0.378	0.632	0.999	0.344	0.640	0.999
	X_{1D}	0.573	0.701	0.986	0.628	0.900	1.000	0.500	0.860	0.999	0.505	0.860	1.000
50	X_{1A}	0.501	0.590	0.932	0.475	0.654	0.976	0.445	0.674	0.991	0.441	0.646	0.998
	X_{1B}	0.617	0.667	0.969	0.626	0.775	0.994	0.620	0.818	0.999	0.588	0.803	0.999
	X_{1C}	0.634	0.687	0.982	0.652	0.813	0.998	0.645	0.851	1.000	0.617	0.855	1.000
	X_{1D}	0.803	0.854	0.997	0.845	0.930	1.000	0.848	0.961	1.000	0.801	0.958	1.000
60	X_{1A}	0.484	0.540	0.898	0.466	0.622	0.964	0.469	0.638	0.976	0.477	0.673	0.966
	X_{1B}	0.777	0.815	0.990	0.835	0.910	1.000	0.778	0.902	1.000	0.787	0.908	1.000
	X_{1C}	0.788	0.825	0.989	0.847	0.926	1.000	0.788	0.917	1.000	0.829	0.940	1.000
	X_{1D}	0.864	0.892	0.998	0.905	0.963	1.000	0.903	0.969	1.000	0.913	0.987	1.000
90	X_{1A}	0.529	0.559	0.918	0.664	0.743	0.983	0.661	0.765	0.988	0.637	0.768	0.994
	X_{1B}	0.876	0.890	0.995	0.915	0.951	1.000	0.925	0.974	1.000	0.926	0.986	1.000
	X_{1C}	0.978	0.983	0.999	0.987	0.998	1.000	0.995	1.000	1.000	0.994	0.999	1.000
	X_{1D}	0.973	0.978	1.000	0.997	0.999	1.000	0.998	1.000	1.000	0.995	1.000	1.000
120	X_{1A}	0.814	0.835	0.982	0.804	0.848	0.992	0.813	0.869	0.996	0.791	0.878	1.000
	X_{1B}	0.987	0.990	1.000	0.997	0.998	1.000	0.990	0.994	1.000	0.993	0.996	1.000
	X_{1C}	0.995	0.996	1.000	0.997	0.998	1.000	0.997	1.000	1.000	1.000	1.000	1.000
	X_{1D}	0.998	1.000	1.000	0.998	0.999	1.000	0.999	1.000	1.000	1.000	1.000	1.000

T=30, X_{1A}에서 G=5인 경우의 검정력은 G=3인 경우의 검정력보다 작다. 왜냐하면 G=5인 경우의 그룹내 표본크기는 6이고, G=3인 경우의 표본크기는 10이 되므로 G=3인 경우보다 G=5인 경우에 $\overline{X}_{1A}$의 값들 중 일부가 비슷한 값을 갖는 경우가 더 많이 발생하여 G=5에서 검정력이 감소한 것이다. 위 예에서 보았듯이 이러한 경향은 그룹내 표본크기가 작고, 균등분포의 구간이 좁을 경우에 더 잘 나타난다. 더불어 4가지 유형 중 X_{1D}에서의 검정력이 가장 크게 나타났으며, LM검정에 비하여 LR검정의 검정력이 높게 나타나고 있다. 이와 같은 차이는 표본크기가 증가할수록 다소 감소되어 T=90, 120인 경우에는 X_{1A}를 제외하고 거의 같은 수준의 검정력을 보여주고 있다. Wald검정은 표본수와 그룹수가 증가함에 따라 증가하지만 그 양은 LM과 LR에 비하여 작다. 또

한 X_{1A}인 경우에 가장 낮은 검정력을, X_{1D}인 경우에 가장 높은 검정력을 갖지만 검정력의 차이는 크지 않다.

4. 결 론

본 연구에서는 선형회귀모형에서 이분산의 형태 중 그룹화 이분산에 대하여 이를 검정하기 위하여 LR, LM과 Wald검정통계량을 유도하였다. 또한 두 가지의 모의실험을 통하여 세가지 검정법에 대하여 검정력을 비교하였다. 모의실험의 결과에서 LM검정법은 명목유의수준을 제대로 유지하고 있는 반면, LR검정법은 명목유의수준을 과대추정하는 경향을 Wald검정통계량은 표본수가 작을수록 그룹수가 많아질수록 더욱 과대추정을 하는 면을 보여주었다. 두 가지 형태의 대립가설에 대한 검정력의 비교 결과에서 표본크기와 그룹수가 증가함에 따라 세 가지 모든 검정법의 검정력이 증가하였다. 그러나 그룹간 분산이 독립변수에 의존하는 가설의 일부결과에서 검정력이 증가하다가 점차 감소하는 경향이 나타났다. 이러한 결과를 통해 선형회귀모형에서 그룹화 이분산을 검정하기 위하여 LR 검정 통계량을 이용하는 경우 이를 구하는 과정이 매우 복잡하며 유의수준 또한 과대 추정하는 경향이 있다 따라서 귀무가설 하에서 LM검정법을 사용하는 것이 적절할 것이다. 이들의 선택기준은 계산상의 편리함에도 있다. 예를 들어, 식 (2.5)에서 알 수 있듯이 LM 검정통계량은 OLS잔차를 이용하므로 비교적 계산상으로 구하기 쉬운 이점이 있다.

참고문헌

[1] Breusch, T.S. and Pagan, A.R. (1979), A simple test for heteroscedasticity and random coefficient variation, *Econometrica*, Vol. 47, 1287-1294.

[2] Buse, A. (1982), The likelihood ratio, wald, and lagrange multiplier tests: An expository note, *The American Statistician,* Vol. 36, 153-157.

[3] Goldfeld, S.M. and Quandt, R.E. (1965), Some tests for heteroscedasticity, *Journal of the American Statistical Association,* Vol. 60, 539-547.

[4] Hartley, H.O. and Rao, J.N.K. (1967), Maximum likelihood estimation for the mixed analysis of variance model, *Biometrika,* Vol. 54, 93-108.

[5] Harvey, A.C. and Phillips, G.D.A. (1974), A comparison of the power of some tests of heteroscedasticity in the general linear model, *Journal of Econometrics,* Vol. 2, 307-316.

[6] Hemmerle, W.J. and Hartley, H.O. (1973), Computing maximum likelihood estimates for the mixed model using the W-transformation, *Technometrics,* Vol. 15, 819-831.

[2004년 1월 접수, 2004년 8월 채택]