응용통계연구
제 5 권 제 1 호, 1992년

단순선형회귀에서의 변화점에 대한 연구†

정광모* · 한미혜

요 약

단순선형회귀모형에서 어떤 미지의 시점을 전후하여 회귀계수에 변화가 있었는지에 대한 통계적 가설을 검정하고 변화점의 추정방법을 논의한다. 이차형식의 통계량을 제안하고 그 근사분포 및 유의수준제어를 살펴보았다. 또한 하나의 예를 통하여 제안된 방법을 적용하고 우도비검정과 비교하였다.

1. 서 론

시간의 흐름에 따라 관찰되는 통계자료에서 어떤 사건이나 자연적인 현상에 의하여 자료의 평균이나 분산등과 같은 모수에 변화가 생겼는지를 분석하는 것은 통계학에서 중요한 관심분야가 된다. 변화점 문제에 대한 많은 연구들은 공변량을 포함하지 않는 자료에서 변화점 유무의 가설검정 및 변화점의 추정에 관하여 모수적 또는 비모수적측면에서 논의 되어 왔다. 이에 관한 참고문헌으로는 Page(1954), Chernoff와 Zacks(1964), Bhattacharyya와 Johnson(1968), Hinkely(1970, 1972), Hsu(1977, 1979), Worsley(1983a, 1986), Lombard(1987), James et al(1987) 등을 들 수 있다.

공변량을 포함하는 통계자료의 경우에 어떤 미지의 시점을 기준으로 반응변수와 공변량간의 회귀관계가 일정한지 아니면 미지의 시점을 전후하여 회귀계수의 변화가 있었는지에 대한 통계적 분석을 회귀모형의 변화점 문제라 하고, 미지의 시점을 변화점이라 한다. 예를 들면, 노동 또는 자본투입량과 이로부터 산출되는 총 생산량과의 회귀관계를 나타내는 계량경제모형에서, 이러한 회귀관계가 전체 시간구간에 있어서 일정한지를 검토하는 것은 회귀모형의 변화점 문제가 된다.

회귀모형에서의 변화점 문제는 Quandt(1958, 1960)에 의하여 최우추정법(maximum likelihood estimation)과 우도비검정법(likelihood ratio test)이 제안된 이래로 현재까지 연구가 계속되고 있다. 지금까지의 연구들은 오차항의 분포가 정규분포라는 가정하에서 이루어졌으며, 우도비검정법은 그 근사분포 조차 매우 복잡하여 쉽게 구해지지 않는 단점이 있다. 그 이유는 검정통계량의 분포가 공변량의

---

† 이 논문은 1990년도 문교부지원 한국학술진흥재단의 지방대육성 학술연구조성비에 의하여 연구되었음.

* (609-735) 부산직할시 금정구 장전동 부산대학교 통계학과

값의 위치에 의존하기 때문이며 이와같은 이유에서 근사값을 구하기 위한 방법들이 제안되었다. Quandt (1960)는 우도비검정통계량의 분포를 카이제곱분포로 근사시키는 것이 매우 좋지 않다는 사실을 발견하였으며, Worsley(1983b)는 Bonferroni 부등식에 의한 근사확률을 제시하였다. 또한, Kim과 Siegmund(1989)는 경계점교차확률(boundary crossing probability)을 이용하여 우도비검정법의 기각역에 대한 근사확률공식을 제안하였다. 우도비검정법과는 다른 방법으로 Brown et al(1975)은 정규성의 가정하에서 축차적잔차(recursive residual)를 이용한 누적합 통계량을 제안하고 그래프기법을 써서 변화점의 추정을 논의하였다.

Macneill(1978)은 오차항의 분포에 대한 특정한 분포형을 가정하지 않고 다항회귀모형에서 단순한 잔차에 근거하여 이차형식의 통계량을 제안하였으며, 이러한 통계량의 극한분포가 Cramer-von Mises 통계량의 분포와 같은 형태를 갖게 된다는 사실을 보였다. 이때, 통계량의 분포를 구하는데 있어서는 브라운 브릿지 과정 (Brownian bridge process)을 이용하였고 회귀계수의 추정에서 최소제곱법을 적용하였다. 본 연구에서는 단순선형회귀모형에서 잔차 대신에 잔차의 순위를 이용하여 Macneill (1978)과 같은 형태의 이차형식 통계량을 제안하고 그 근사분포에 관하여 논의하고자 한다. 이와 같은 통계량은 공변량을 포함하지 않는 자료에서 Lombard(1987)에 의하여 연구된 바 있으며 그 근사분포가 쉽게 구해질 수 있고, 기존의 통계분포표를 이용하여 기각치를 구할 수 있는 특징이 있다. 제안된 통계량은 회귀계수의 추정치를 포함하기 때문에 추정된 회귀계수를 갖는 통계량의 성질에 관한 분포이론을 논의한다. 또한 모의실험을 통하여 귀무가설하에서의 유의수준 제어를 살펴보고 자료에 적용하여 제안된 통계방법의 실용성을 알아본다.

2. 통계적모형 및 기존의 검정법

2.1 통계적 모형

두 변수 x, y가 단순선형회귀관계를 가지면서 시간순서에 따라 관찰되었다고 하자. 변수 x는 공변량 또는 독립변수이고, y는 종속변수라 할 때 두 변수간의 관계가 전 시간구간에 대하여 일정한 회귀직선을 나타내는지 아니면 어떤 미지의 시점 t를 기준으로 회귀계수에 변화가 생겼는지에 관한 변화점 문제를 다루고자 한다. 변화점의 갯수는 많아야 한 번 일어난 단일 변화점모형을 가정할 때 이것을 통계적모형으로 나타내면

$$y_i = \begin{cases} \alpha_1 + \beta_1 x_i + \varepsilon_i, & \text{단, } 1 \le i \le t \\ \alpha_2 + \beta_2 x_i + \varepsilon_i, & \text{단, } t < i \le n \end{cases} \tag{2.1}$$

와 같다. 여기서 t는 미지의 변화점이고, ε_i는 서로 독립인 오차항으로서 연속인 분포함수 $F(\varepsilon)$를 따른다. 이의 확률밀도함수 $f(\varepsilon)$는 대칭이고 조건 $\int \{f'(\varepsilon)/f(\varepsilon)\}^2 dF(\varepsilon) < \infty$ 을 만족한다고 가정하자.

상수항 또는 기울기의 변화에 대한 대립가설

$H_1 : \alpha_1 \neq \alpha_2$ 또는 $\beta_1 \neq \beta_2$

에 대하여 귀무가설

$$H_0 : \alpha_1 = \alpha_2 , \quad \beta_1 = \beta_2$$

을 검정하고, 변화점 t의 추정방법을 다루고자 한다.

2.2 우도비검정법

우도비검정법은 Quandt(1960)에 의하여 제안되었고, Worsley(1983a)가 중선형회귀모형의 변화점 문제로 확장하였다.

고정된 t의 경우 우도비는

$$\Lambda = \frac{[\{t\hat{\sigma}_1^2 + (n-t)\hat{\sigma}_2^2\}/n]^{n/2}}{(\hat{\sigma}^2)^{n/2}} \tag{2.2}$$

와 같고, 여기서

$$\hat{\sigma}^2 = \sum_{i=1}^{n} (y_i - \hat{\alpha} - \hat{\beta} x_i)^2 / n,$$

$$\hat{\sigma}_1^2 = \sum_{i=1}^{t} (y_i - \hat{\alpha}_1 - \hat{\beta}_1 x_i)^2 / t,$$

$$\hat{\sigma}_2^2 = \sum_{i=t+1}^{n} (y_i - \hat{\alpha}_2 - \hat{\beta}_2 x_i)^2 / (n-t)$$

이다. 또한 $\hat{\alpha}$ 및 $\hat{\beta}$는 H_0이 참일 때 α, β의 최우추정량이고, $\hat{\alpha}_i$, $\hat{\beta}_i$ (i=1,2)는 H_1 하에서 시점 t 이전과 이후의 최우추정량을 나타낸다. 식 (2.2)는 시점 t 의 함수로서 $L(t) = -2\log\Lambda$라 할 때 우도비 검정통계량은

$$\begin{aligned} L &= \max_{2 \le t \le n-2} L(t) \\ &= \max_{2 \le t \le n-2} [\, n\log\hat{\sigma}^2 - n\log\{(t\hat{\sigma}_1^2 + (n-t)\hat{\sigma}_2^2)/n\} \,] \end{aligned} \tag{2.3}$$

와 같이 정의한다. 따라서 우도비검정법 L과 동치인 통계량으로

$$V = \max_{2 \le t \le n-2} (U_t / Q)$$

를 사용하기로 한다. 이 때, $Q = n\hat{\sigma}_2$이고, $U_t = n\hat{\sigma}^2 - (t\hat{\sigma}_1^2 + (n-t)\hat{\sigma}_2^2)$이다.

Quandt(1960)는 우도비검정법 L의 근사분포를 카이제곱분포로 추측하였으나 모의실험결과로부터 이러한 카이제곱분포근사가 매우 좋지않다는 사실을 논의하였다. 우도비검정법의 근사분포는 일반적으로 독립변수값의 배열에 의존하게 되므로 그 근사분포를 구하는 일은 힘들며, Hawkins(1980)의

주장에 의하면 우도비검정법의 근사분포는 알려져 있지 않다. Worsley(1983b)는 우도비검정법의 기각역을 결정하기 위한 방법으로 Bonferroni 부등식에 의한 유계(bound)값을 제안하였다. 또한, Kim과 Siegmund(1989)도 경계점교차확률의 근사공식을 써서 이에 대한 해결책을 제시하고 있으나 근사공식이 매우 복잡한 형태를 갖는다.

2.3 Macneill 검정법

Macneill(1978)은 다항회귀모형

$$y_i = \sum_{j=0}^{p} \beta_j x_i^j + \varepsilon_i, \quad i = 1, 2, \cdots, n$$

에서 오차항 ε_i가 평균 0, 분산 σ^2인 연속분포함수 $F(\varepsilon)$를 따를 때 회귀계수 $\underset{\sim}{\beta} = (\beta_0, \cdots, \beta_p)'$에 관한 변화점 문제를 논의하였다. 독립변수가 $x_i = i/n$일 때 β의 최소제곱추정량을 써서 잔차의 부분합 $\sum_{i=1}^{t} (y_i - \hat{y}_i)$에 관한 이차형식의 통계량

$$M = \sum_{t=1}^{n-1} w_n(t/n) \{\sum_{i=1}^{t} (y_i - \hat{y}_i)\}^2 \tag{2.4}$$

을 제안하였다. 식(2.4)에서 가중값 $w_n(t/n)$는

$$w_n(t/n) = \int_{(2t-1)/2n}^{(2t+1)/2n} \psi(u)du$$

와 같이 정의되며, $\psi(u)$는 음이 아닌 함수로서 조건

$$\int_0^1 u(1-u)\psi(u)du < \infty$$

을 만족한다.

통계량 M은 회귀분석결과에서 얻어지는 잔차 자체를 직접 이용하고 있기 때문에 계산상 편리한 장점이 있다. Macneill(1978)은 일반화된 브라운브릿지 $\{B_P(u) \mid 0 \le u \le 1\}$을 써서 통계량 $(n\sigma^2)^{-1}M$의 분포가 $n \to \infty$ 일 때

$$\int_0^1 \psi(u)\ B_P^2(u)du \tag{2.5}$$

에 수렴함을 보였다. 특별히 p=0인 경우 $\{B_P(u)\}$는 표준 브라운브릿지 과정이 되며 Anderson과 Darling(1952)은 $\psi(u)=1$일 때 식(2.5)의 확률분포를 구하였다. Macneill(1978)은 p=0, 1, 2, 3, 4, 5인 경우에 식(2.5)의 몇 가지 백분위값에 대한 근사값들을 계산하였다.

위치모수나 척도모수에 관한 변화점 문제에서 Lombard(1987)는 관찰값의 순위를 이용하여 식(2.4)와 같은 형태의 통계량을 제안하고, 단일변화점의 경우에 통계량의 분포가 Cramer-von Mises 통계량의 극한분포와 같아짐을 보였다.

3. 제안된 검정법

3.1 검정통계량

변화점에 관한 모형 (2.1)에서 잔차를 직접 이용하는 대신에 잔차의 순위를 써서 검정통계량을 정의하고자 한다. 잔차 $e_i = y_i - \hat{\alpha} - \hat{\beta} x_i$ 의 순위를 R_i 라 할 때, R_i 는 $y_i - \hat{\beta} x_i$ 의 순위와 같은 값을 갖게 되므로 상수항의 추정치와는 무관하게 된다. 식(2.4)와 같은 형태의 통계량을

$$Q = \sum_{t=1}^{n-1} (1/n) \{ \sum_{i=1}^{t} a(R_i) \}^2 \tag{3.1}$$

와 같이 제안한다. 여기서 a(i)는 함수 $\phi(u), 0 \le u \le 1$에 의하여

$$a(i) = \phi(i/(n+1)) - \int_0^1 \phi(u) du$$

로 정의된 점수이며, $\phi(u)$는 조건 $0 < \int \phi^2(u) du < \infty$ 을 만족한다.

통계량 Q는 식(2.4)의 통계량 M에서 잔차를 순위로 바꾸어 표현하였으며 가중치는 1/n로 두었다. 통계량의 계산에는 회귀계수 β의 추정이 필요하며 추정방법으로는 최소제곱추정량 또는 Hodges-Lehmann 추정량을 고려할 수 있다. Hodges-Lehmann 추정량은 Adichie(1967)의 계산공식에 의하여 구해졌으며 독립변수 값에 대한 몇 가지 약한 조건 아래에서 $\sqrt{n}$ - 일치추정량이다. 또한 최소제곱추정량도 유사한 조건하에서 이와 같은 성질을 갖는다.

식(3.1)의 정의에서 $\phi(u)$를 윌콕슨 점수로 하면, 이 때 부분합

$$\sum_{i=1}^{t} a(R_i) = \sum_{i=1}^{t} R_i/(n+1) - t/2 \tag{3.2}$$

은 시점 t 이전까지의 순위합과 동치인 통계량이 된다. 통계량 Q의 근사분포는 윌콕슨 점수에 국한하여 논의하고자 한다.

3.2 근사분포

Lombard(1987)는 공변량을 갖지 않는 변화점모형에서 식(3.1)과 같은 통계량의 근사분포를 논의하였다. 회귀계수 β가 주어진 경우에 통계량 Q의 근사분포는 브라운브릿지를 써서 설명된다. 식

$$B_n(u) = \begin{cases} (nA^2)^{-1/2} \sum_{i}^{t} a(R_i), & \text{단, } t/n \le u < (t+1)/n, \ 1 \le t < n \\ 0, & \text{그 밖에} \end{cases} \tag{3.3}$$

라 놓으면, Billingsley(1968)의 Theorem 24.1에 의하여 H_0 이 참이고 $n \to \infty$ 일 때 $B_n(u)$의 분포는

표준 브라운브릿지 B(u)에 수렴한다. 식(3.3)에서 $A^2=\sum_i^n \{\phi(i/(n+1))-1/2\}/n$이다. 분포의 수렴에 관한 연속사상정리(continuous mapping theorem)로부터 $(nA^2)^{-1}Q$의 분포는

$$\int_0^1 B^2(u)du \tag{3.4}$$

에 수렴한다. 식(3.4)의 분포는 Cramer-von Mises 통계량의 극한분포로서 Anderson과 Darling(1952)에 의하여 분포표가 구해졌다.

그러나 회귀계수 β는 미지의 값을 나타내므로 추정된 값을 사용하면 통계량 Q는 추정된 값에 의존하게 된다. 식(3.2)에 정의된 부분합은 추정값을 포함한 U-통계량 $U(\hat{\beta})$와 동치이다. 여기서

$$U(\hat{\beta}) = \frac{1}{t(n-t)} \sum_{(i,j)\in C_n} h[(y_i-\hat{\beta}x_i-y_j+\hat{\beta}x_j)(d_i-d_j)] \tag{3.5}$$

와 같다. 단, d_i 는 $i \le t$일 때 1이고, $i > t$일 때는 0이며, $C_n = \{(i,j) \mid 1\le i,\ j\le n,\ d_i > d_j\}$이다. 그리고 h(u)는 커널(kernel)로서 u가 양수일 때 1, 그 밖에 −1의 값을 갖는다. 식(3.5)은 Randles(1984)의 식(4.2)에 정의된 통계량과 유사한 형태가 되므로 그의 전개방법을 적용한다. $U(\gamma)$의 기대값 $E_\beta[U(\gamma)]$를 $\mu(\gamma)$라 놓고 오차항의 차 $\varepsilon_1-\varepsilon_2$의 분포함수를 G라 할 때, $\mu(\gamma)$의 미분(differential)은

$$d\mu(\gamma) \mid_{\gamma=\beta} = (\bar{x}_1-\bar{x}_2)\int h(u)\ g'(u)du \tag{3.6}$$

과 같다. 단, $\bar{x}_1$, $\bar{x}_2$는 각각 t 이전과 이후의 평균값이다.

추정량 $\hat{\beta}$가 $\sqrt{n}$ - 일치추정량일 때 Randles(1984)의 Theorem A.9에 의하면 식(3.6)이 0에 근사할 때 $U(\hat{\beta})$과 $U(\beta)$는 같은 극한분포를 갖게 된다. 독립변수 x_i 의 값이 랜덤하게 관찰된 경우에 이와 같은 조건이 성립하게 되고, 이 때 제안된 통계량 Q의 근사분포는 식(3.4)와 같게 된다.

변화점의 추정방법에 관하여는 Lombard(1987)에 의하여 사용된 부분합도표를 통하여 변화점의 대략적인 위치를 파악한다. 즉, 시점 t와 부분합 $S_t=\sum_{i=1}^t a(R_i)/A$의 관계 (t, S_t)를 그래프로 나타내고, S_t 의 변화상태가 증가에서 감소 또는 그 반대현상이 뚜렷하게 나타나는 점의 위치를 변화점으로 추정하고자 한다. 이러한 추정방법은 통계적 품질관리에서 잘 알려진 누적합도표의 일종이며 다음절의 보기를 통하여 적용방법을 설명한다.

4. 모의실험 및 보기

4.1 모의실험을 통한 유의수준의 비교

제안된 검정통계량 Q의 근사분포가 식(3.4)에 의하여 주어짐을 앞 절에서 논의하였는데, 근사분포의 기각치를 사용한 검정법이 유의수준을 얼마나 잘 제어하는지 모의실험을 통하여 비교하고자 한다.

오차항의 모집단 분포는 균일분포, 정규분포, 이중지수분포 및 코쉬분포로 택하였고, 표본크기는 n=15, n=25 인 경우로 하였다. 독립변수 x_i의 값은 1부터 n까지의 정수값을 랜덤하게 치환하여 얻

어졌으며, 회귀모형의 상수항은 0으로 두었다. 그리고, 기울기의 값은 1/2, 1 및 2인 세 가지 경우에 대하여 반복횟수를 1,000번 시행하여 기각역에 속하게 되는 횟수를 실험유의확률로 나타내었다.

제안된 검정법 Q의 기각치는 Anderson과 Darling(1952)의 분포표 Table 1을 활용하여 구하였고, 우도비 검정법의 기각역은 정규분포에서 10,000번 반복한 모의실험결과의 백분위수를 사용하였다. 또한, Macneill 검정 M의 기각값은 Macneill(1978)의 Table 2에서 p=1인 경우의 값을 택하였다. 모의실험 결과는 다음 표1, 표2와 같다. 표의 Q통계량에서 $\hat{\beta}_{HL}$ · $\hat{\beta}_{LSE}$는 각각 β의 Hodges-Lehmann 추정값 및 최소제곱추정값을 사용한 경우를 나타낸다.

표1. n=15 ; β=2.0 ; 유의수준 α=0.05(0.10)인 경우

통계량	균일분포	정규분포	이중지수분포	코쉬분포
Q				
$\hat{\beta}_{HL}$	0.041(0.090)*	0.042(0.088)	0.043(0.090)	0.042(0.092)
$\hat{\beta}_{LSE}$	0.044(0.091)	0.046(0.093)	0.045(0.095)	0.041(0.090)
L	0.029(0.061)	0.048(0.097)	0.095(0.158)	0.252(0.286)
M	0.379(0.474)	0.375(0.479)	0.379(0.492)	0.391(0.517)

*괄호안의 값은 유의수준 α=0.10일 때임.

표2. n=25 ; β=2.0 ; 유의수준 α=0.05(0.10)인 경우

통계량	균일분포	정규분포	이중지수분포	코쉬분포
Q				
$\hat{\beta}_{HL}$	0.038(0.093)	0.045(0.092)	0.040(0.091)	0.044(0.088)
$\hat{\beta}_{LSE}$	0.043(0.094)	0.041(0.088)	0.043(0.089)	0.031(0.083)
L	0.034(0.070)	0.059(0.105)	0.101(0.153)	0.255(0.294)
M	0.363(0.468)	0.359(0.491)	0.380(0.483)	0.407(0.534)

표 1, 표 2의 결과로부터 제안된 검정법 Q의 기각값은 명목유의 수준의 제어에 있어서 오차항의 분포에 관계없이 양호한 반면, 우도비 검정법은 정규분포보다 두터운 꼬리를 갖는 이중지수분포와 코쉬분포에서 유의수준을 제어하지 못함을 알 수 있다. 또한 Macneill검정은 전반적으로 유의수준을 전혀 제어하지 못하기 때문에 독립변수값이 랜덤하게 관찰된 경우에는 사용할 수 없음을 알 수 있다. 균일분포의 경우에 우도비검정법은 다소 보수적(conservative)이다. 명목유의수준이 α =0.01인 경우에는 제안된 검정법의 실험유의확률이 작게 나타났고, β =0.5, β =1인 경우도 위 표 1, 표 2에 제시된 것과 유사한 값이므로 지면관계상 이에 대한 표는 생략한다.

4.2 적용사례

앞에서 논의된 검정법들중 제안된 검정법과 우도비검정법을 다음과 같은 자료에 적용하여 그 결과를 비교하고자 한다. Macneill검정은 독립변수가 x_i = i라는 제약조건 하에서 논의되었기 때문에 인용된 자료에는 적합하지 못하여 비교에서 제외하였다.

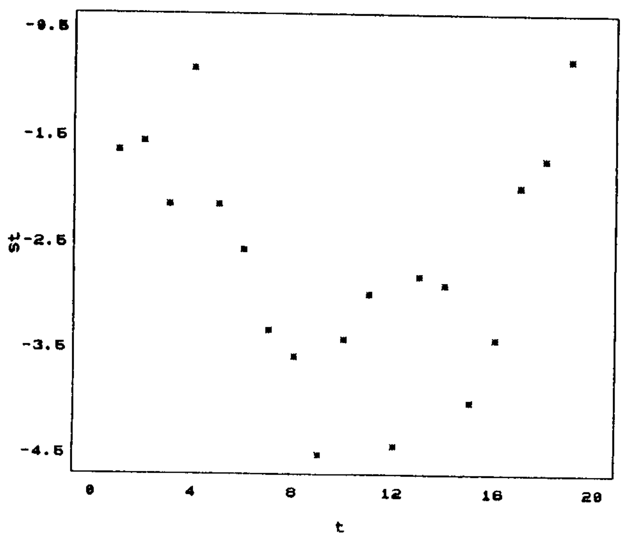
보기 표 3에 주어진 자료는 Quandt(1958)에 의하여 만들어진 가상 자료이며, 정규분포를 따르는 오차항을 가정하고 독립변수 값은 1에서 20까지의 정수를 랜덤하게 치환하여 얻어졌다. 변화점을 갖는 자료로서 변화점을 t=12로 한 것이다.

표 3. Quandt의 자료

t	1	2	3	4	5	6	7	8	9	10
x_t	4	13	5	2	6	8	1	12	17	20
y_t	3.473	11.555	5.714	5.710	6.046	7.650	3.140	10.312	13.353	17.197

t	11	12	13	14	15	16	17	18	19	20
x_t	15	11	3	14	16	10	7	19	18	9
y_t	13.036	8.264	7.612	11.802	12.551	10.296	10.014	15.472	15.650	9.871

표 3의 자료 전체에 대한 최소제곱추정값은 $\hat{\alpha}$ =3.21, $\hat{\beta}$ =0.64이고, 결정계수 R^2=0.90이다. 이 때 최소제곱추정값을 사용한 Q통계량에 의한 가설검정의 결과 유의확률 p값이 0.079로 나타났다.



〈그림 1〉 Quandt 자료의 누적합 도표

그림 1은 누적합도표로서 (t, S_t)의 관계를 나타낸다. 단, $S_t = \sum_{i=1}^{t} a(R_i)/A$이다. 그래프의 변화 상태로부터 변화점은 대략 t=8부터 t=12 사이에 위치함을 알 수 있다. Worsley(1983b)의 Table 2에 의하면 우도비검정법의 유의확률 p는 0.05보다 작게 되며, 이 때 추정된 변화점은 t=12이다. 표 3의 자료에서 t=12 이전의 최소제곱추정값은 $\hat{\alpha}=2.22$, $\hat{\beta}=0.69$이고, 결정계수 $R^2=0.95$인 반면 t=13 이후의 자료에 대한 $\hat{\alpha}=5.91$, $\hat{\beta}=0.48$, $R^2=0.93$으로서 특히 상수항의 변화가 두드러짐을 알 수 있다. 또한, 변화점 이전과 이후의 자료를 따로 분석한 경우가 전체자료에 대한 결정계수의 값보다 크게 되어 모형이 더욱 잘 적합된 것을 나타내준다.

5. 맺음말

공변량을 포함한 변화점문제에서 단순선형회귀모형을 가정하고 변화점의 유무에 대한 검정법 및 변화점의 추정방법을 논의하였다. 기존의 우도비검정법에 대한 근사분포가 쉽게 구해질 수 없는 단점을 갖는 데 반하여 제안된 이차형식의 검정법은 기존의 이용가능한 근사분포를 활용하여 검정기각역을 결정할 수 있는 특징이 있다. 추정된 모수를 포함한 통계량의 근사분포를 논의하고 실제의 자료에 적용하여 제안된 통계적 방법의 활용방안을 살펴보았다.

단순선형회귀모형에 관한 변화점문제를 좀 더 일반적인 회귀모형에 확장할 수 있는 지에 대한 것은 본 연구에서 논의되지 않았다. 또한, 단일변화점문제를 다중변화점 문제로 확장하는 것 등도 앞으로의 연구과제로 남겨두고자 한다.

참 고 문 헌

[1] Adichie, J.N.(1967), "Estimation of regression parameter based on rank tests", *Annals of Mathematical Statistics*, 38, 894-904.

[2] Anderson, T.W. and Darling, D.A.(1952), "Asymptotic theory for certain goodness of fit criteria based on stochastic processes", *Annals of Mathematical Statistics*, 23, 193-212.

[3] Bhattacharyya, G.K. and Johnson, R.A.(1968), "Nonparametric tests for shift at unknown time point", *Annals of Mathematical Statistics*, 39, 1731-1743.

[4] Billingsley, P.(1968), *Convergence of Probability Measure*, John Wiley and Sons, Inc.

[5] Brown, R.L., Durbin, J., and Evans, J.M.(1975), "Techniques for testing the constancy of regression relationships over time", *Journal of Royal Statistical Society*, B, 37, 149-192.

[6] Chernoff, H., and Zacks, S.(1964), "Estimating the current mean of a normal distribution which is subject to changes in time", *Annals of Mathematical Statistics*, 35, 999-1018.

[7] Hawkins, D.M.(1980), "A note on continuous and discontinuous segmented regression", *Technometrics*, 22, 443-444.

[8] Hinkley, D.V.(1970), "Inference about the change-point in a sequence of random variables", *Biometrika*, 57, 1-17.

[9] Hinkley, D.V.(1972), "Time-ordered classification.", *Biometrka*, 59, 509-523.
[10] Hsu, A.A.(1977), "Tests for variance shift at an unknown time point", *Applied Statistics*, 26, 279-284.
[11] Hsu, D.A.(1979), "Detecting shifts of parameter in gamma sequences with applications to stock price and air traffic flow analysis", *Journal of the American Statistical Association*, 74, 31-40.
[12] James, B., James, K.L. and Siegmund, D.(1987), "Tests for a changepoint", *Biometrika*, 74, 71-84.
[13] Kim, H.J. and Siegmund, D.(1989), "The likelihood ratio test for a changepoint in simple linear regression", *Biometrika*, 76, 409-423.
[14] Lombard, F.(1987), "Rank Tests for changepoint problem", *Biometrika*, 74, 615-624.
[15] Macneill, I.B.(1978), "Properties of sequences of partial sums of polynomial regression residuals with applications to tests for change of regressions at unknown times", *The Annals of Statistics*, 6, 422-433.
[16] Page, E.S.(1954), "Continuous inspection schemes", *Biometrika*, 41, 100-115.
[17] Quandt, R.E.(1958), "The estimation of the parameter of a linear regression system obeying two separate regimes", *Journal of the American Statistical Association*, 53, 873-880.
[18] Quandt, R.E.(1960), "Tests of the hypothesis that a linear regression system obeys two separate regimes", *Journal of the Americal Statistical Association*, 55, 324-330.
[19] Randles, R.H.(1984), "On tests applied to residuals", *Journal of the American Statistical Association*, 79, 349-354.
[20] Worsley, K.J.(1983a), "The power of likelihood ratio and cumulative sum tests for a change in a binomial probability", *Biometrika*, 70, 455-464.
[21] Worsley, K.J.(1983b), "Testing for a two-phase multiple regression", *Technometrics*, 25, 35-42.
[22] Worsley, K.J.(1986), "Confidence regions and tests for a changepoint in a sequence of exponential family random variables", *Biometrika*, 73, 91-104.

A Study on Change-Points in Simple Linear Regression†

Kwang Mo Jeong* and Mee Hyea Han

ABSTRACT

A testing and estimation procedure is considered for changes at unknown time point in simple linear regression model. A test statistic of quadratic form is suggested. We also discuss the asymptotic distribution and its level control. The proposed method is compared with the likelihood ratio test through a example.

† Resrarch supported in part by the Education of Ministry of Korea
* Department of Statistics, Pusan National University, Pusan, 609-735, Korea