Journal of the Korean
Data & Information Science Society
2001, Vol. 12, No. 1, pp. 27 ~ 40

잔차 분산을 이용한 선형회귀모형의 다중전환점 검정

이인석 [1] · 김종태 [2] · 이금자 [3]

요약

본 연구는 시간의 변화에 따라 여러 개의 전환점이 발생하여 선형회귀모형들이 여러 번 변화할 때의 변환시점을 Gasser, Stroke와 Jennen-Steinmez의 잔차분산 추정량을 이용하여 검정하고 실제의 몇 가지 모형을 제시하여 Graphic을 통하여 조사한 결과 여기서 제시한 방법이 더 효과적으로 자중전환점을 찾을 수 있었다.

주제어: 전환점, 잔차분산, GSJS 추정량

제 1 절 서론

회귀분석은 통계적 응용 분야에 있어 가장 중요한 분야들 중 하나이다. 회귀분석의 적용에 있어서, 만약 관심 있는 자료가 어떤 시간점(전환점)이후로부터 변화가 일어나는 구조를 가진다면, 한 개의 회귀모형을 가지고서 그 자료를 분석하는 것은 그 회귀모형을 잘 설명해 줄 수 없을 뿐 아니라 매우 잘못된 결론을 유도해 낼 것이다. 그러므로 전환점(changing point)의 검정이 선행되어진 후에 그 결과에 따라서 회귀분석을 하는 것이 보다 타당한 방법이다.

또한 잔차분산의 추정량이 회귀분석에서의 통계적 자료분석과 통계적 추론에 미치는 영향력을 생각해 볼 때 회귀함수의 추정에 대한 연구는 모수적, 비모수적인 측면에서 매우 많은 연구들이 진행되어 진 것에 비하여 잔차분산의 추정연구는 상대적으로 미약하였다. 비모수 회귀모형에 있어서의 차분에 기저 한 방법에 관한 잔차분산에 대한 연구는 Rice(1984)의 연구에 의하여 처음 시작된 후에 Gasser, Sroka와 Jennen-Steinmetz(1986)과 Hall, Kay와 Titterington(1990, 1991) 등에 의하여 연구되었으며, 회귀모형에서 전환점 문제들에 대해서도 많은 연구들이 진행되어져 왔다. 즉 Quandt(1958, 1960)는 한 개의 전환점을 기준으로 두 개의 회귀모형을 따르는 선형회귀모형의 추정과 검정에 기초한 우도비(likelihood ratio)에 대한 연구를 하였고 Brown, Durbin과 Evans(1975)는 중회귀모형에서 전환점들을 검정하기 위한 순환 잔차(residuals)를 이용한 방법을 소개하였으며 Kim(1998)은 시간의 변화에 따른 선형회귀모형에 대한 한 개의 전환점(changing point)에 대하여서만 연구하였다.

또한 통계적 전환점에 대한 문제는 통계 분석적인 측면에서 많은 관심의 대상이 되어 Sen과 Srivastava(1975)는 다변량 가우시안 관찰 값들의 수열 평균 벡터들의 전환점 문제에 대해 연구하였다. Chen과 Gupta(1997)는 우도함수 추정방법을 사용하여 다중평균벡터와 공분산 전환점들에 대한 문제를 연구했고, Chen(1998)은 Schwarz 정보판별함수를 사용하여 회귀모형에 대한 전환점에 대해 최근 연구하였다. 그러나 선형회귀모형에서의 전환점 문제

[1] 대구광역시 북구 산격동 1370 경북대학교 자연과학대학 통계학과 교수
[2] 경북 경산시 하양읍 내리리 15 대구대학교 이과대학 통계학과 부교수
[3] 대구광역시 북구 산격동 1370 경북대학교 대학원 통계학과

를 해결하기 위하여 최근까지 제시한 거의 대부분의 연구 방법은 선형회귀모형에서의 고전적인 오차 분산의 추정량에만 의존하였다. Gasser, Sroka와 Jennen-Steinmez(1986, GSJS)는 한 개의 전환점을 제시하였다지만 우리는 다중전환점에 대한 연구를 Gasser, Sroka와 Jennen- Steinmez(1986, GSJS)가 제시한 부분 선형 적합(local linear fit)에 기초한 잔차분산의 비모수 추정량을 이용하여 다중전환점 검정기법을 제시하고자 한다.

제 2 절 GSJS의 잔차 분산의 추정

반응변수 $y = \{y_1, y_2, \cdots y_n\}^T$, $y_1 = y(t_i)$은 다음의 회귀모형을 가진다고 가정하자.

$$y = f + \varepsilon \tag{2.1}$$

이 때 고정된 값 $t_1 < t_2 < \cdots < t_n$에 대하여 회귀함수 $f = (f(t_1), \cdots, f(t_n))^T$는 미지의 $\varepsilon = (\varepsilon_1, \cdots, \varepsilon_n)^T$는 $E(\varepsilon_i) = E(\varepsilon_i^3) = 0$이고 $Var(\varepsilon_i) = \sigma^2$와 $E(\varepsilon_i^4) < \infty$인 성질을 갖는다. Rice(1984)는 잔차의 분산 σ^2에 대한 일차 차분에 기저한 추정량을

$$\begin{aligned} \widehat{\sigma_R^2} &= \frac{1}{2(n-2)} \sum_{i=1}^{n-1} (y_{i+1} - y_i)^2 \\ &= \frac{1}{2(n-2)} \sum_{i=1}^{n-1} (\frac{y_{i+1} - y_i}{t_i + 1 - t_i})^2 (t_{i+1} - t_i)^2 \end{aligned} \tag{2.2}$$

으로 제시하였다. 추정량들의 성질에 있어서 고차 차분(higher order differences)들을 사용함으로 추정량의 편의(bias)를 감소시킬 수 있다. 이때 물론 분산의 값이 증가하는 결과는 감수해야 한다. 분산의 추정 뿐 아니라 모든 추정에 있어서 우리는 편의를 최소화하고 분산도 최소화하는 추정량을 가지길 원한다.

또한 Gasser, Sroka와 Jennen, Steinmetz(1986)의 연구에서, 그들이 제시한 2차 차분(second order differences)을 이용한 잔차분산의 추정량은 편의와 분산을 줄이는 작용을 하였다. 이들이 제시한 추정량을 GSJS 추정량이라고 이름을 짓고 $\widehat{\sigma_{GSJS}^2}$이라고 표기하겠다. GSJS의 추정량을 공부하기 앞서, 먼저 차분에 기저한 분산 추정에 있어서 의사잔차(pseudo-residuals), $\tilde{\varepsilon_i}$에 대하여 살펴보자. 의사잔차 $\tilde{\varepsilon_i}$는 t_{i-1}, t_i, t_{i+1}의 계획된 점들(design points)의 연속적인 삼중관계 (continuous triples)에 의해 얻어지는데 일직선상의 두 개의 외곽 관찰값 (outer observations)들을 결합하고 난 다음 이 일직선과 그 중앙의 관찰값 $y(t_i)$ 사이의 차분을 계산한다. 수식적으로 표현하면

$$\begin{aligned} \tilde{\varepsilon_i} &= \frac{t_{i+1} - t_i}{t_{i+1} - t_{i-1}} y(t_{i-1}) + \frac{t_i - t_{i-1}}{t_{i+1} - t_{i-1}} y(t_{i+1}) - y(t_i) \\ &= a_i y(t_{i-1}) + b_i y(t_{i+1}) - y(t_i), \quad i = 2, \cdots, n-1 \end{aligned} \tag{2.3}$$

이다.

GSJS 추정량의 기본적인 개념은 위의 잔차들이 $f = 0$일 때 $\tilde{\varepsilon_i^2}$이 σ^2에 대한 불편성을 갖게 하기 위하여 정규화(normalization)을 시키고 이러한 정규화된 잔차의 평균을 σ^2의 추정량으로서 사용한다는 것이다. 이제 의사 잔차를 행렬로 표현하여 잔차의 분산에 대한 GSJS

추정량을 구하기 위하여 식(2.3)을 표현하기 위해 $(n-2)\times n$ 행렬 A를 다음과 같이 두자.

$$A=\begin{bmatrix} a_2 & -1 & b_2 & 0 & \cdots & 0 \\ 0 & a_3 & -1 & b_3 & \ddots & \vdots \\ \vdots & & \ddots & & & 0 \\ 0 & \cdots & 0 & a_{n-1} & -1 & b_{n-1} \end{bmatrix}$$

또한 B는 $(n-2)\times(n-2)$의 대각행렬로서

$$\begin{aligned} B &= \operatorname{diag}\left[\frac{1}{\sqrt{a_2^2+b_2^2+1}},\frac{1}{\sqrt{a_3^2+b_3^2+1}},\cdots,\frac{1}{\sqrt{a_{n-1}^2+b_{n-1}^2+1}}\right] \\ &= \operatorname{diag}[c_2,c_3,\cdots,c_{n-1}] \end{aligned}$$

로 표현된다. 그러면 위의 두 행렬을 이용하여 $D=BA$는 $(n-2)\times n$행렬로서

$$D=\begin{bmatrix} c_2a_2 & -c_2 & c_2b_2 & 0 & \cdots & 0 \\ 0 & c_3a_3 & -c_3 & c_3b_3 & \ddots & \vdots \\ \vdots & & \ddots & & & 0 \\ 0 & \cdots & 0 & c_{n-1}a_{n-1} & -c_{n-1} & c_{n-1}b_{n-1} \end{bmatrix}$$

가 된다. 그러므로 식 (2.3)의 의사 잔차 $\tilde{\varepsilon}=Dy=Df+D\varepsilon$는

$$\tilde{\varepsilon}=Dy=\begin{bmatrix} c_2a_2y_1 & -c_2y_2 & +c_2b_2y_3 \\ c_3a_3y_2 & -c_3y_3 & +c_3b_3y_4 \\ \vdots & \vdots & \vdots \\ c_{n-1}a_{n-1}y_{n-2} & -c_{n-1}y_{n-1} & +c_{n-1}b_{n-1}y_n \end{bmatrix}$$

으로 표현되며 의사잔차 $\tilde{\varepsilon}$의 잔차분산의 추정량 $\widehat{\sigma^2_{GSJS}}$은

$$\widehat{\sigma^2_{GSJS}}=\frac{\tilde{\varepsilon}^T\tilde{\varepsilon}}{tr(D^TD)}=\frac{y^TD^TDy}{tr(D^TD)}=\frac{1}{(n-2)}\sum_{i=2}^{n-1}\tilde{\varepsilon}_i^2 \qquad (2.4)$$

으로 나타난다. 여기서 $\tilde{\varepsilon}_i$는 $\tilde{\varepsilon}_i=c_i(a_iy_{i-1}+b_iy_{i+1}-y_i)$이다.

또한 행렬 W를 $W=D^TD$로 두면 σ^2에 대한 추정량들의 일반적인 형태는

$$\widehat{\sigma^2_{GSJS}}=y^TWy/tr(W)$$

의 이차형태(quadratic form)로 표현되어 진다. 여기서 W는 대칭이며 양정치(positive-definite) 행렬이며 분모에 있는 $tr(W)$은 추정량이 $f=0$일 때 σ^2에 대한 불편성 추정량임을 보장하여 준다.

더욱이 (2.1)에 있는 모형에 대한 GSJS 추정량을 세부적으로 표현하여 보면

$$\widehat{\sigma^2_{GSJS}}=(f^TWf+2f^TW\varepsilon+\varepsilon^TW\varepsilon)/tr(W) \qquad (2.5)$$

이다. 위의 분산의 추정량은 이차형태를 가진 세 부분으로 나누어 생각하면 (2.5)에 있는 첫 번째 부분인 $f^TWf/tr(W)$은 양의 편의를 나태내고, 두 번째 부분 $2f^TW\varepsilon/tr(W)$은 평균이 0이고 분산이 $4\sigma^2f^TW^2f/\{tr(W)\}^2$인 정규분포를 가지며, 마지막 부분인 $\varepsilon^TW\varepsilon/tr(W)$ 은 평균 σ^2과 분산 $2\sigma^4tr(W)/[tr(W)]^2$을 가지는 σ^2의 자연적인 추정량(natural estimator)이다. GSJS 추정량 $\widehat{\sigma^2_{GSJS}}$의 불편성을 보장해 주는 가정 $f=0$은 매우 강한 가정이다. 그러나 실제적으로 요구되어지는 가정은 $f=0$가 아니라 $f^TWf=0$이다. 이 이유는 $f=0$가 되는 경우는 물론이고 $f\neq 0$인 경우에도 불편성이 거의 성립이 되어지기 때문이다.

정리 2.1.[Kim (1998)] 회귀모형 (2.1)에 있는 조건에 대하여, 만약 의사 잔차 $\widetilde{\varepsilon}=Dy$가 정의되어 진다면 $\widehat{\sigma^2_{GSJS}}-\sigma^2=O_p(\frac{1}{\sqrt{n}})$이다.

제 3 절 다중전환점 검정

시간의 변화에 따라서 선형회귀모형들의 변화에 따라서 선형회귀모형들의 변환이 일어나는 전환점들을 검정하기 위한 가설은 다음과 같다.

H_0 : 한 개의 전환점도 발생하지 않는다 vs. H_1 : 한 개 이상의 전환점들이 발생한다

만약 한 개 이상의 전환점이 발생한다면 우리는 정확한 변화의 시점, 즉 전환점들을 정확히 찾아 낼 수 있는가? 하는 것이 우리들의 중요한 관심이 될 것이다. 위의 가설을 검정하기 위하여 $y=X\beta+\varepsilon$의 선형회귀모형을 고려하자. 여기서 $X=(X_1,\cdots,X_n)$는 $n\times(p+1)$의 행렬로 $(p+1)$개의 미지의 회귀벡터이며, $\varepsilon=(\varepsilon_1,\varepsilon_2,\cdots,\varepsilon_n)$는 평균이 0이고 분산이 σ^2인 정규분포로서 각각 독립인 랜덤오차이다. 또한 설명변수 $Y=(y_1,\cdots,y_n)$는 $N(X_i\beta,\sigma^2)$에 따르는 확률변수이다. 그러면 선형회귀모형에서 전환점의 존재여부를 알기 위하여 $u_y=(u_{y1},u_{y2},\cdots,u_{yn})$일 때 한 개의 모형만을 가지는 귀무가설은 $H_0: u_y=X\beta$이고, 이에 대한 대립가설은 $H_1: \mu_y=X\beta_1, i=1,\cdots,k;\ \mu_y=X\beta_2, i=k+1,\cdots,n$이다. 이러한 다중전환점 가설에 대한 검정을 위하여 2절에서 소개한 GSJS의 잔차분산의 추정량의 성질을 이용한다. 시간점 t_k, $k=3,4,\cdots,n-1$이 변함에 따라 GSJS의 잔차분산의 추정량을 이용한 통계량

$$\widehat{\sigma^2_{GSJS}}(k)=\frac{1}{(k-2)}\sum_{i=2}^{k-1}\widetilde{\varepsilon_i^2}=\frac{1}{(k-2)}\sum_{i=2}^{k-1}c_i(a_iy_{i-1}+b_iy_{i+1}-y_i) \tag{3.1}$$

의 변화는 선형회귀모형의 전환점들에서 발생하는 모형의 변화에 따라 GSJS의 잔차분산의 추정량의 변화는 점프의 형태로 일어남을 관찰할 수 있다. 즉, k의 변화에 따른 GSJS의 잔차분산의 추정량이 어느 시점 k에서 점프가 일어난다면, 우리는 $\widehat{k}=k+1$에서 전환점을 찾을 수 있다. 따라서 본 논문에서는 다중전환을 찾기 위한 통계량으로

$$\begin{aligned}\widehat{\sigma^2_{LK}}(k) &= \frac{1}{(k-2)}\sum_{i=2}^{k-1}\widetilde{\varepsilon_i^2}+\frac{1}{(n-k-2)}\sum_{i=k+2}^{n-1}\widetilde{\varepsilon_i^2}\\ &= \frac{1}{(k-2)}\sum_{i=2}^{k-1}c_i(a_iy_{i-1}+b_iy_{i+1}-y_i)\\ &\quad+\frac{1}{(n-k-2)}\sum_{i=k+2}^{n-1}c_i(a_iy_{i-1}+b_iy_{i+1}-y_i)\end{aligned} \tag{3.2}$$

을 제시한다. 그러면 이 방법에 의해 다중 전환점을 찾는 방법은 시간의 변화에 따라 $\widehat{\sigma^2_{LK}}(k)$의 값이 어느 시점 k에서 점프의 형태로 일어나며, 또한 $\hat{k} = k+1$에서 한 개이상의 전환점을 찾을 수 있다.

제 4 절 다중전환점들에 대한 모의 실험

Chen and Gupta(1997)에 의한 전환점은

$$\hat{\sigma}(k) = \frac{1}{k}\sum_{i=1}^{k}(y_i - \overline{y_1})^2 + \frac{1}{n-k}\sum_{i=k+1}^{n}(y_i - \overline{y_1})^2 + n(1+log2\pi) + 2logn \qquad (4.1)$$

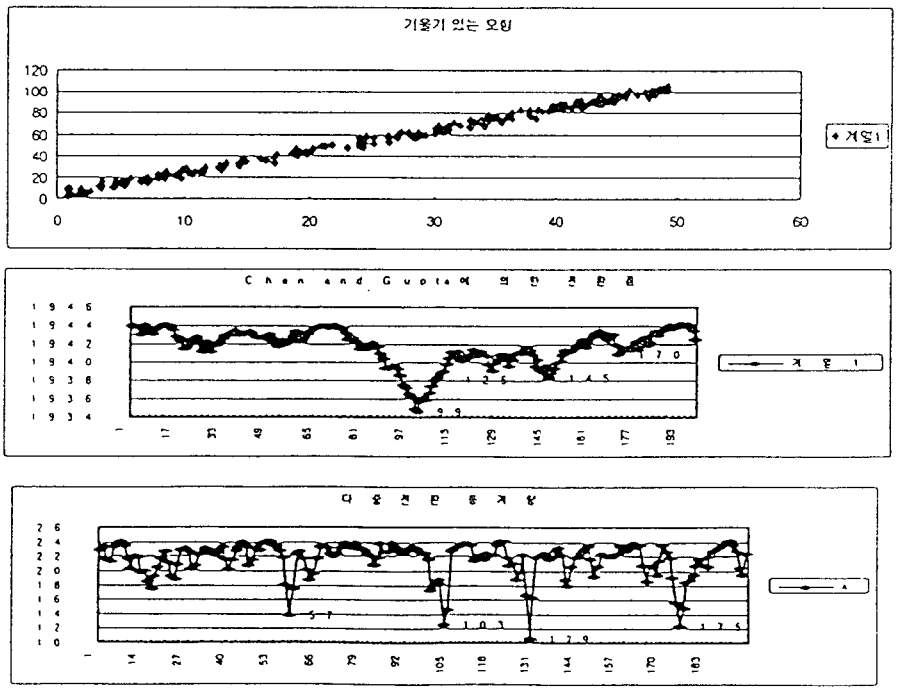
이다.

같은 모형으로 식(3.2)과 식(4.1)의 다중전환점을 비교해 보았다.

4. 1 전환점이 3개인 경우

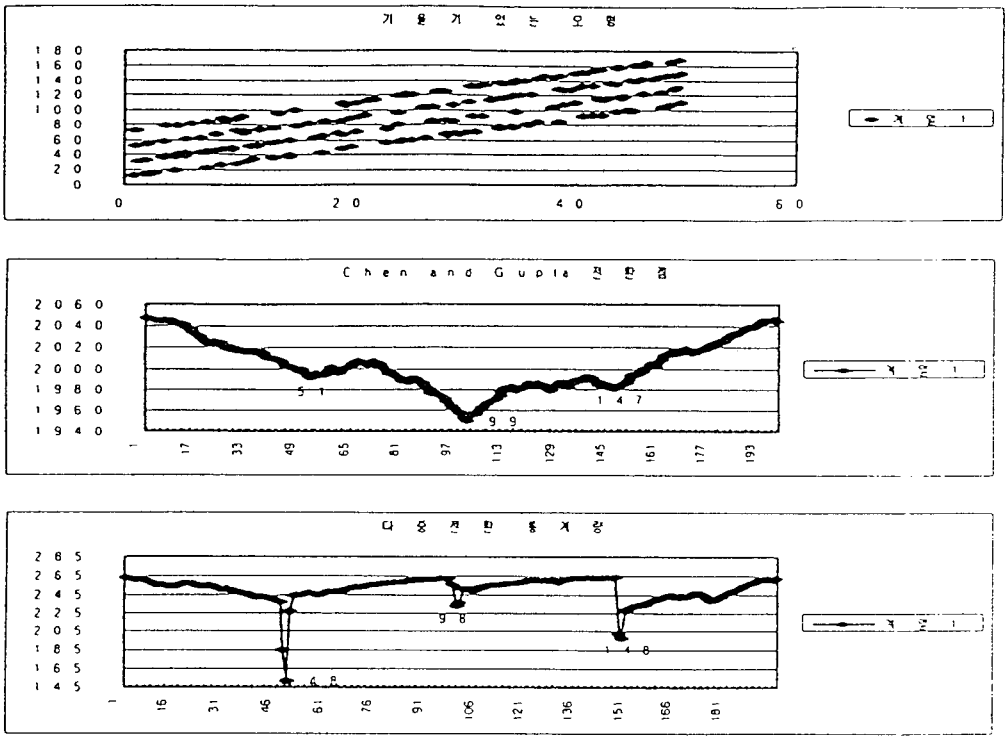
[모형1] : $y_1 = 1 + 2x_1 + \varepsilon_1 \quad y_2 = 3 + 2x_2 + \varepsilon_2$

$y_3 = 5 + 2x_3 + \varepsilon_3 \quad y_4 = 7 + 2x_4 + \varepsilon_4$



<그림1> 절편이 다른 경우의 전환점이 3개인 경우
(절편이 1, 3, 5, 7일 때 $\varepsilon \sim N(0,1)$)

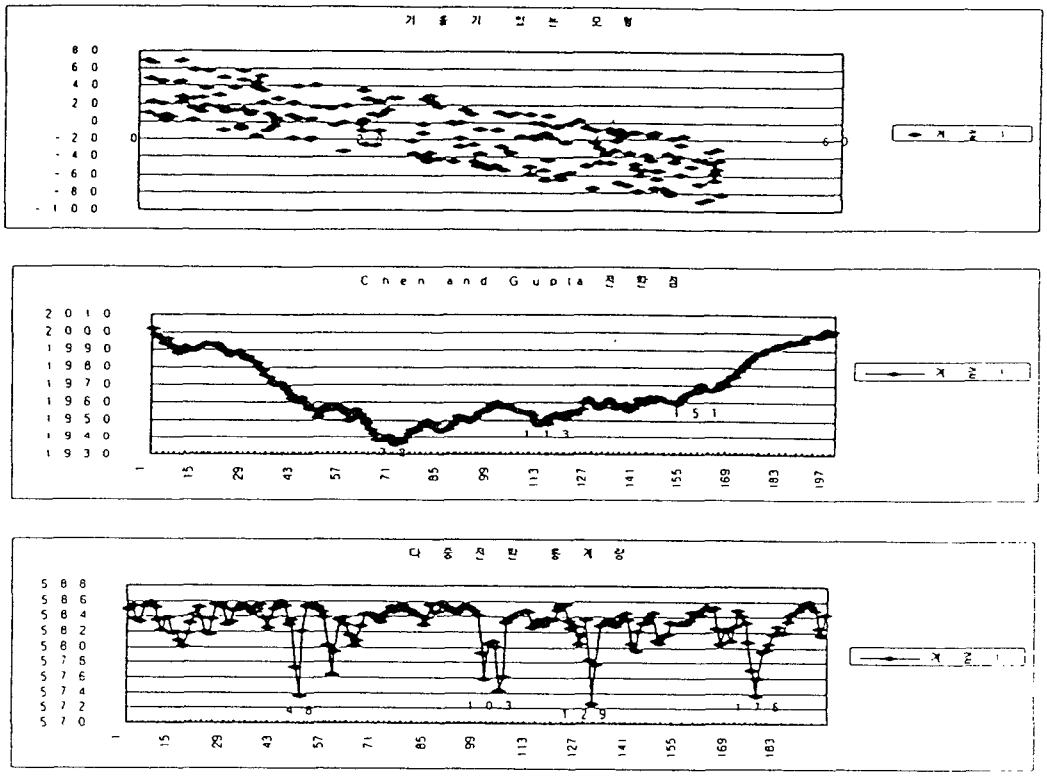
[모형2] : $y_1 = 10 + 2x_1 + \varepsilon_1 \quad y_2 = 30 + 2x_2 + \varepsilon_2$
$y_3 = 50 + 2x_3 + \varepsilon_3 \quad y_4 = 70 + 2x_4 + \varepsilon_4$



<그림2> 절편이 다른 경우의 전환점이 3개인 경우
(절편이 10, 30, 50, 70일 때 $\varepsilon \sim N(0,1)$)

<그림1>에서 Chen and Gupta 전환점 99, 145, 170과 다중전환 통계량 57, 103, 129, 175를 전환점으로 잡을 수 있으며 <그림2>에서는 Chen and Gupta 전환점 51, 99, 147과 다중전환 통계량 48, 98, 148을 전환점으로 잡을 수 있다. 그러므로 모형1과 모형2에서 기울기는 같고 절편이 다른 모형 4개를 만들었을 때 전환점이 3개정도 생기는 것을 알 수 있으며 모형1에서는 Chen and Gupta 전환점이 모형2에서는 다중전환 통계량이 더 정확한 전환점을 찾아냄을 알 수 있다.

[모형3] $: y_1 = 10 + (-2)x_1 + \varepsilon_1 \quad y_2 = 30 + (-2)x_2 + \varepsilon_2$
$y_3 = 50 + (-2)x_3 + \varepsilon_3 \quad y_4 = 70 + (-2)x_4 + \varepsilon_4$

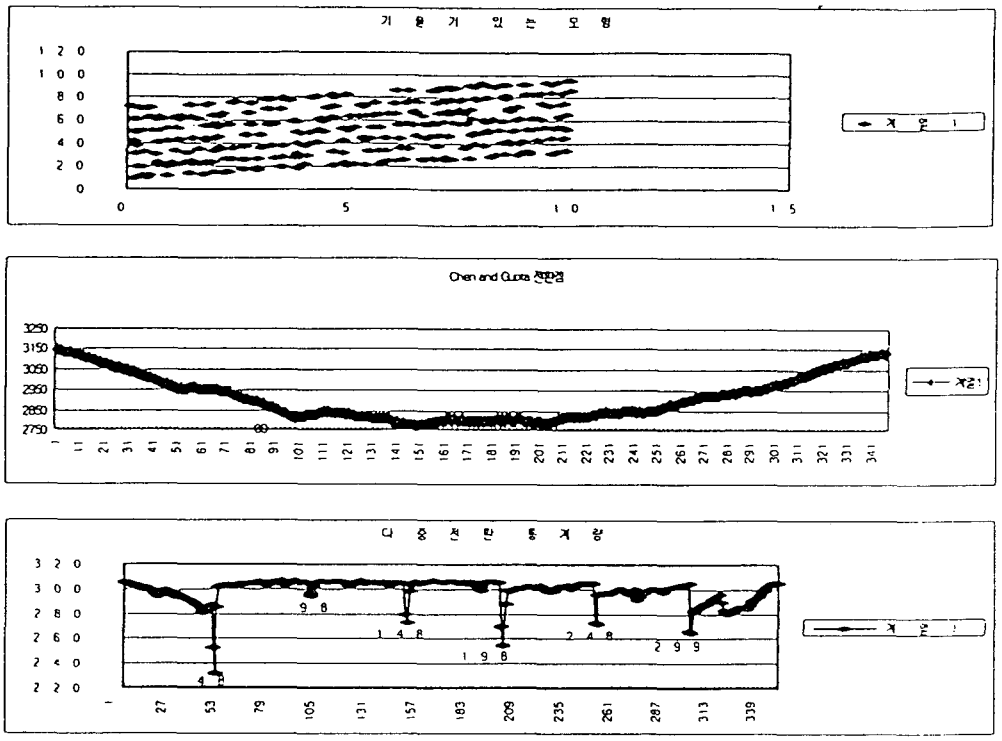


<그림3> 절편이 다르면서 기울기가 음인 경우의 전환점이 3개인 경우($\varepsilon \sim N(0,4)$)

<그림3>에서 Chen and Gupta 전환점 72, 113, 151과 다중전환 통계량 48, 103, 129, 175를 전환점으로 잡을 수 있다. 모형3에서는 모형2와 절편은 같게 기울기는 음의 값을 가지게 만들어 보았다. 기울기가 양, 음에 상관없이 전환점은 3개 정도 찾을 수 있었으며 Chen and Gupta 전환점보다 다중전환 통계량이 더 정확하게 나타남을 알 수 있다.

4. 2 전환점이 6개인 경우

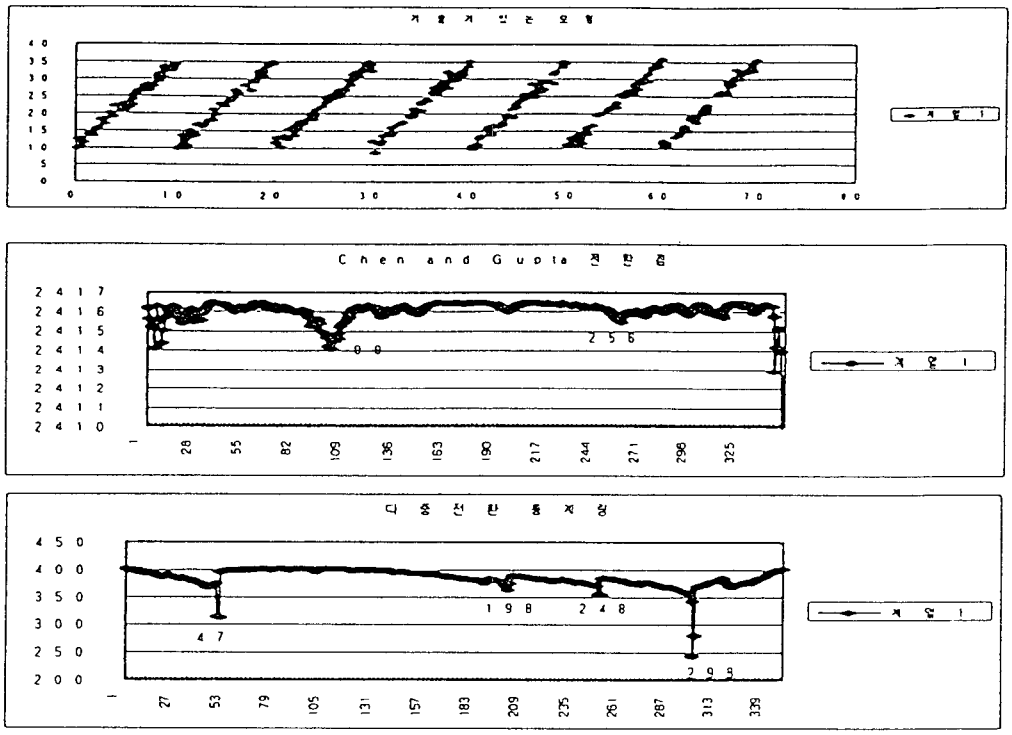
[모형4] : $y_1 = 10 + 2.5x_1 + \varepsilon_1$ $y_2 = 20 + 2.5x_2 + \varepsilon_2$ $y_3 = 30 + 2.5x_3 + \varepsilon_3$
$y_4 = 40 + 2.5x_4 + \varepsilon_4$ $y_5 = 50 + 2.5x_5 + \varepsilon_5$ $y_6 = 60 + 2.5x_6 + \varepsilon_6$
$y_7 = 70 + 2.5x_7 + \varepsilon_7$



<그림4> 절편이 다르면서 기울기가 양인 경우의 전환점이 6개인 경우($\varepsilon \sim N(0,1)$)

<그림4>에서 Chen and Gupta 전환점 99, 147과 다중전환 통계량 48, 98, 148, 198, 248, 299를 전환점으로 잡을 수 있다. 기울기는 같고 절편이 다른 모형 7개를 만들었을 때 전환점이 6개정도 생기는 것을 알 수 있으며 Chen and Gupta 전환점보다 다중전환 통계량이 더 정확한 전환점을 찾아냄을 알 수 있다.

[모형5] : $y_1 = 10 + 2.5x_1 + \varepsilon_1$ $y_2 = 10 + 2.5x_2 + \varepsilon_2$ $y_3 = 10 + 2.5x_3 + \varepsilon_3$
$y_4 = 10 + 2.5x_4 + \varepsilon_4$ $y_5 = 10 + 2.5x_5 + \varepsilon_5$ $y_6 = 10 + 2.5x_6 + \varepsilon_6$
$y_7 = 10 + 2.5x_7 + \varepsilon_7$

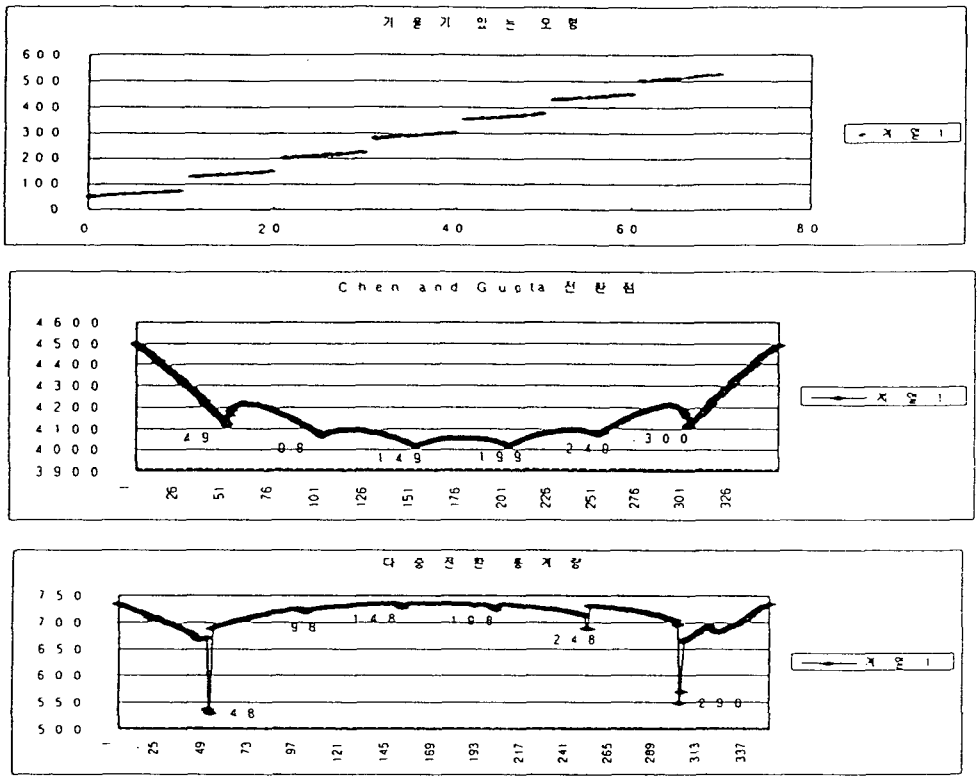


<그림5> 절편이 같으면서 기울기가 음인 경우의 전환점이 6개인 경우

($\varepsilon \sim N(0,1)$, $x_1 \sim x_1$, $x_2 \sim x_2 + 10$, $x_3 \sim x_3 + 20$, $x_4 \sim x_4 + 30$,
$x_5 \sim x_5 + 40$, $x_6 \sim x_6 + 50$, $x_7 \sim x_7 + 60$)

<그림5>에서 Chen and Gupta 전환점 99, 256과 다중전환 통계량 48, 198, 248, 298을 전환점으로 잡을 수 있다. 기울기, 절편이 같은 모형 7개를 만들고 x의 구간을 나누었을 때 Chen and Gupta 전환점은 2개, 다중전환 통계량은 4개의 전환점을 가지며 다중전환 통계량이 더 정확한 전환점을 찾아냄을 알 수 있다.

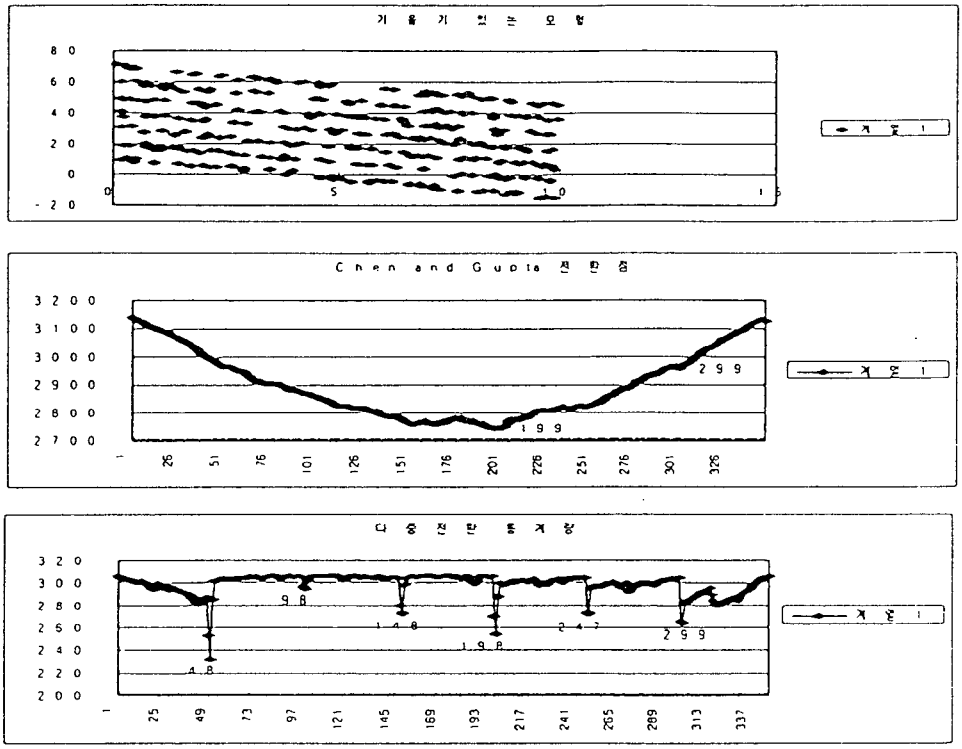
[모형6] : $y_1 = 50 + 2.5x_1 + \varepsilon_1 \quad y_2 = 100 + 2.5x_2 + \varepsilon_2 \quad y_3 = 150 + 2.5x_3 + \varepsilon_3$
$y_4 = 200 + 2.5x_4 + \varepsilon_4 \quad y_5 = 250 + 2.5x_5 + \varepsilon_5 \quad y_6 = 300 + 2.5x_6 + \varepsilon_6$
$y_7 = 350 + 2.5x_7 + \varepsilon_7$



<그림6> 절편이 큰 경우의 전환점이 6개인 경우($\varepsilon \sim N(0,1)$)

<그림6>에서 Chen and Gupta 전환점 49, 98, 149, 199, 249, 300과 다중전환 통계량 48, 98, 148, 198, 248, 298을 전환점으로 잡을 수 있다. 모형6은 기울기는 모형5와 같고 절편은 크게 잡은 모형 7개를 만들었다. 절편을 크게 잡았더니 Chen and Gupta 전환점과 다중전환 통계량이 거의 비슷한 점에서 전환점을 가지는 것을 알 수 있다.

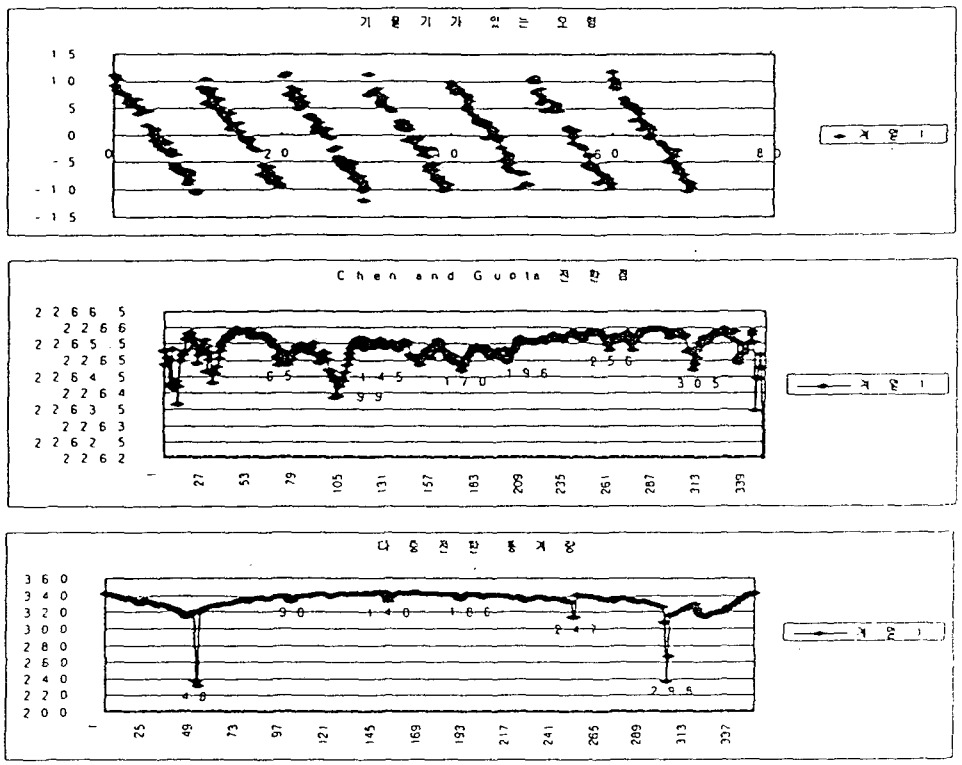
[모형7] : $y_1 = 10 + (-2.5)x_1 + \varepsilon_1$ $y_2 = 20 + (-2.5)x_2 + \varepsilon_2$ $y_3 = 30 + (-2.5)x_3 + \varepsilon_3$
$y_4 = 40 + (-2.5)x_4 + \varepsilon_4$ $y_5 = 50 + (-2.5)x_5 + \varepsilon_5$ $y_6 = 60 + (-2.5)x_6 + \varepsilon_6$
$y_7 = 70 + (-2.5)x_7 + \varepsilon_7$



<그림7> 절편이 다르면서 기울기가 음인 경우의 전환점이 6개인 경우($\varepsilon \sim N(0,1)$)

<그림7>에서 Chen and Gupta 전환점 199, 299과 다중전환 통계량 48, 98, 148, 198, 247, 299를 전환점으로 잡을 수 있다. 모형7은 모형4를 기울기가 음의 값을 갖게 만들었다. <그림4>와 비교할 때 Chen and Gupta 전환점은 값은 다르지만 2개를 찾았고 다중전환 통계량은 거의 같은 점에서 전환점 6개를 찾아냄을 알 수 있다. 기울기가 양, 음의 값에 상관없다는 것을 알 수 있으며 다중전환 통계량이 더 정확한 전환점을 찾아냄을 알 수 있다.

[모형8] : $y_1 = 10 + (-2.5)x_1 + \varepsilon_1$ $y_2 = 10 + (-2.5)x_2 + \varepsilon_2$ $y_3 = 10 + (-2.5)x_3 + \varepsilon_3$
$y_4 = 10 + (-2.5)x_4 + \varepsilon_4$ $y_5 = 10 + (-2.5)x_5 + \varepsilon_5$ $y_6 = 10 + (-2.5)x_6 + \varepsilon_6$
$y_7 = 10 + (-2.5)x_7 + \varepsilon_7$



<그림8> 절편이 같으면서 기울기가 음인 경우의 전환점이 6개인 경우

($\varepsilon \sim N(0,1)$, $x_1 \sim x_1$, $x_2 \sim x_2 + 10$, $x_3 \sim x_3 + 20$, $x_4 \sim x_4 + 30$,
$x_5 \sim x_5 + 40$, $x_6 \sim x_6 + 50$, $x_7 \sim x_7 + 60$)

<그림8>에서 Chen and Gupta 전환점에서 정확한 전환점은 65, 99, 305이고, 나머지 145, 170, 196, 256은 전환점이라고 보기가 힘들다. 다중전환 통계량 48, 247, 298에서 정확한 전환점을 가지고 나머지 98, 148, 186은 전환점이라고 보기가 힘들지만 Chen and Gupta 전환점보다 전환점을 찾기가 더 쉽다. 모형5와 기울기값만 음으로 바꾼 모형8에서도 다중전환 통계량이 더 정확한 전환점을 찾아냄을 알 수 있다.

위의 예에서 본 것 같이 기울기가 있는 모형에서 전환점이 3개인 경우는 Chen and Gupta 전환점과 다중전환 통계량이 거의 차이가 없이 비슷하게 찾아주지만 전환점이 6개인 경우는 기울기가 양, 음의 값에 상관없이 같은 전환점을 찾아주며 절편이 커질수록 더 정확한

전환점을 찾음을 알 수 있다. Chen and Gupta 전환점보다 다중전환 통계량이 더 정확한 전환점을 찾아주는것을 알 수 있다.

제 5 절 결론

이 논문에서는 다중 전환점을 찾는 방법은 시간의 변화에 따라 다중전환 통계량의 값이 어느 시점 k에서 점프의 형태로 일어나며, 우리는 $\hat{k} = k + 1$에서 한 개이상의 전환점을 찾을 수 있다는 다중 전환점을 가지는 통계량을 제시하였다. 다중 전환점을 가지는 Chen and Gupta 전환점과 제시한 전환점을 비교했을 때 Chen and Gupta 전환점보다 제시한 전환점이 기울기(양, 음), 절편에 관계없이 더 정확한 전환점을 찾아 주는 것을 알 수 있다.

참고문헌

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

2. Chen, J. and Gupta, A.K. (1997). Testing and locating variance changepoints with application to stock price, *Journal of the American Statistical Association*, Vol. 92, 739-747.

3. Chen, J. (1998). Testing for a change point in linear regression models, *Communications in Statistics Theory and Method*, Vol. 27, 2481-2493.

4. Gasser, T., Sroka, L. & Jennen-Steinmetz, C. (1986). Residual variance and residual pattern in nonlinear regression. *Biometrika* Vol. 73, 625-33.

5. Hall, P., Kay, J. W. & Titterington, D. M. (1990). Asymptotically optimal difference-based estimation of variance in nonparametric regression. *Biometrika*, Vol. 77, 521-28.

6. Hall, P., Kay, J. W. & Titterington, D. M. (1991). On estimation of noise variance in two-dimensional signal processing. *Advanced Applied Probability*, Vol. 23. 115-123.

7. Kim Jong-Tae (1998). 비모수 회귀모형의 차분에 기저한 분산추정에 대한 고찰, *The Korean Communications in Statistics*, Vol.5, 121-131.

8. Quandt, R.E. (1958). The estimation of the parameters of a linear regression system obeys two separate regimes, *Journal of the American Statistical Association*, Vol. 53, 873-880.

9. Quandt, R.E. (1960). Tests of the hypothesis that a linear regression system obeys two separate regimes, *Journal of the American Statistical Association*, Vol. 55, 324-330.

10. Rice, J. (1984). Bandwidth choice for nonparametric kernel regression. *Annals of Statistics*, Vol. 12, 1215-30.

11. Sen, A.K. and Srivastava, M.S. (1975). On test for detecting change in mean, *Annals of Statistics*, Vol. 3, 461-464.

Testing for a multiple change point residual variance in regression model

In Suk Lee [4] · Jong Tae Kim [5] · Kum Ja Lee [6]

Abstract

The purpose of this study is to test a multiple change point in the regression model with the passage of time, using the estimated residual variance figure suggested by Gasser, Sroka and Jennen - Steinmez (GSJS).

As a result of the simulation, it is showed that there is a jump change of the estimated residual variance figure at that time of change point.

The way to analyze a intuitive multiple change point through graphics is more effective and accurate than any other existing ways.

Key Words and Phrases: Change point, Residual variance, GSJS estimator

[4]Professor, Dept of Statistics, Kyungpook National University
[5]Associate Professor, Dept of Statistics, Teagu University
[6]Graduate, Dept of Statistics, Kyungpook National University