

제 12 장 개입분석(intervention analysis) 및 이상점의 탐지(outlier detection)

개입분석(intervention analysis)

시계열자료에 영향을 줄 수 있는 사건이 언제 발생했는지를 아는 경우 이들이 어떠한 형태로 자료에 영향을 미칠 지를 모형화 한 후 이들 효과를 추정하고 유의성 여부를 판단하는 분석

이상점(outlier)

사전에 알 수 없는 사건 또는 알 수 없는 이유들로 인해 영향을 받아 다른 패턴을 갖는 자료 모형의 적합에 영향을 주고 잘못 적합한 모형에 기초한 추론들은 문제점을 지님

● 이상점이 존재하는 경우의 대처법

이상점 탐지(outlier detection), Tsay (1984), Chang 등 (1988)

잔차를 이용하여 이상점의 존재 여부를 판단하고 원인을 규명하여 모형에 반영
로버스트(robust) 방법, Denby and Martin (1979)

이상점에 영향을 받지 않는 방법을 이용하여 자료를 분석

12.1 개입분석

개입(intervention) : 정책의 변화, 파업, 기름 파동, 상품광고 등과 같이 시계열 자료에 영향을 미칠 가능성이 있는 사건

개입모형(intervention model) 또는 간섭된 시계열모형(interrupted time series model) : 개입을 입력변수로 사용하는 시계열 모형

12.1.1 개입변수의 두 가지 형태

지시함수(indicator function)와 계단함수(step function)

(1) 지시함수(indicator function) 또는 펄스함수(ulse function) $P_t^{(T)}$

어떤 사건이 T 시점에서 발생하여 그 효과가 T 시점에만 영향을 미치는 경우

$$P_t^{(T)} = \begin{cases} 0, & t \neq T \\ 1, & t = T \end{cases}$$

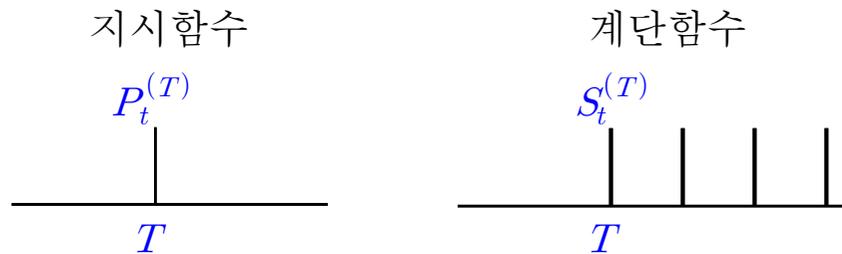
(2) 계단함수(step function) $S_t(T)$

어떤 사건이 T 시점에서 발생하여 그 효과가 발생시점 이후로 지속적으로 영향을 미치는 경우

$$S_t^{(T)} = \begin{cases} 0, & t < T \\ 1, & t \geq T \end{cases}$$

● 지시함수와 계단함수의 관계

$$P_t^{(T)} = S_t^{(T)} - S_{t-1}^{(T)} = (1-B)S_t^{(T)}$$



<그림 12.1> 개입변수의 형태

12.1.2 개입변수의 형태에 따른 반응형태

개입변수의 형태

$$v(B)B^b X_t = (v_0 + v_1 B + v_2 B^2 + \dots) B^b X_t$$

절약성의 원리에 의한 표현

$$\frac{w(B)}{\delta(B)} B^b X_t$$

X_t 는 입력변수

$$w(B) = w_0 - w_1 B - \dots - w_s B^s$$

$$\delta(B) = 1 - \delta_1 B - \dots - \delta_r B^r$$

b : 개입효과가 지체되어 반영되는 기간을 나타내는 지체모수

w_j : 개입의 초기 기대효과

δ_j : 개입의 영속적인 효과

$\delta(B) = 0$ 의 근들의 절대값은 1보다 크거나 같은 것으로 가정

근의 크기가 1인 경우에는 선형으로 증가하는 효과를 나타냄

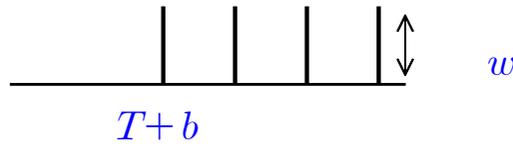
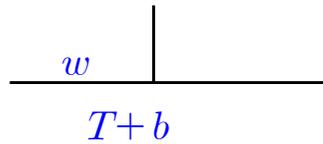
1보다 큰 경우에는 점차적으로 변화하는 효과를 나타냄

개입의 효과가 반응변수에 반영되는 형태, Box and Tiao (1975)

- (1) 개입의 효과가 b 시차 후에 반영되며 그 크기가 w 로 일정한 경우
 개입의 효과가 즉시 반영되는 경우에는 $b=0$ 이 된다.

$$wB^b P_t^{(T)}$$

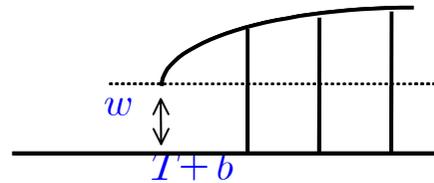
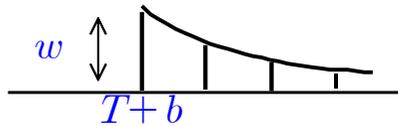
$$wB^b S_t^{(T)}$$



- (2) 개입의 효과가 b 시차 후에 반영되며 그 크기가 점차로 줄어드는 경우

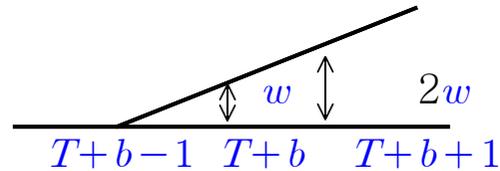
$$\frac{w}{(1-\delta B)} B^b P_t^{(T)}$$

$$\frac{w}{(1-\delta B)} B^b S_t^{(T)}$$



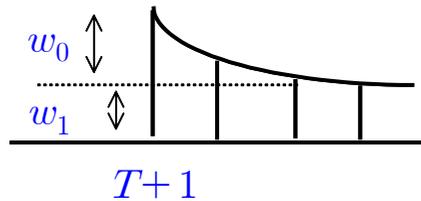
- (3) 개입의 효과가 b 시차 후에 반영되며 선형으로 증가하는 경우
 $\delta(B) = 0$ 의 근이 1인 경우 개입의 효과가 누적되어 선형추세를 가진다.

$$\frac{w}{1-B} B^b S_t^{(T)}$$

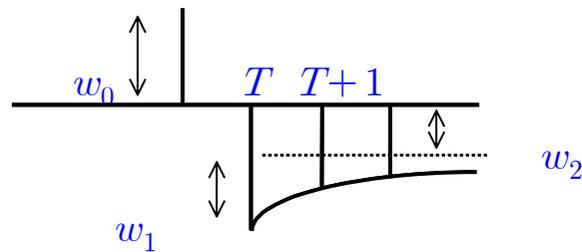


- (4) 반응변수에 반영되는 효과는 점차로 줄어드나 개입이 발생하기 이전의 수준과는 다른 새로운 수준으로 수렴하는 경우. 즉, 개입에 의해 반응변수의 수준에 항구적인 변화가 생기는 경우

$$\left[\frac{w_0}{1-\delta B} + \frac{w_1}{1-B} \right] B P_t^{(T)}$$



$$\left[w_0 + w_1 \frac{B}{1-\delta B} + w_2 \frac{B}{1-B} \right] P_t^{(T)}$$



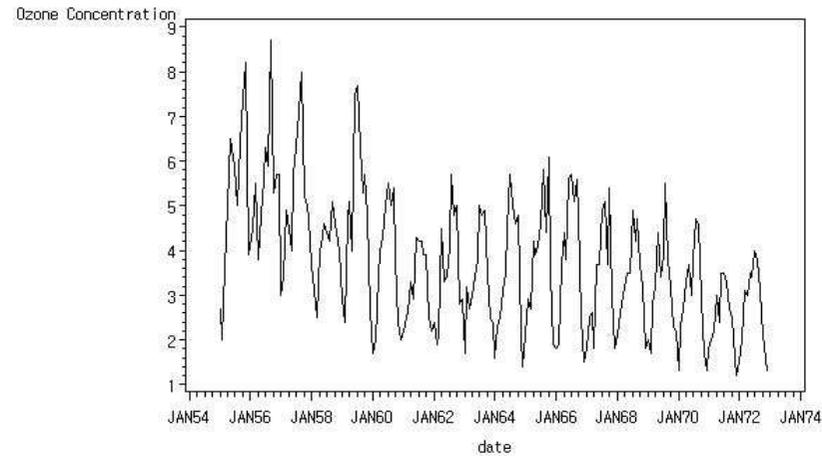
(5) 여러 개의 개입들이 결합되어 동시에 반응변수에 영향을 미치는 경우 다중개입모형(multiple intervention model)이라고 한다.

$$Z_t = \sum_{i=1}^k \frac{w_i(B)B^{b_i}}{\delta_i(B)} I_{i,t} + \frac{\theta(B)}{\phi(B)} \varepsilon_t$$

$I_{i,t} = P_t^{(T)}$ 또는 $S_t^{(T)}$ 의 형태를 갖는 개입변수

12.1.3 개입모형의 적합 예

예제 12.1 자동차 배기가스로 인한 대기오염 조치들의 영향력 분석 , Box and Tiao (1975)
1955년부터 1972년 까지 LA 지역에서 시간별로 관측된 오존(O_3)의 월평균자료 분석



<그림 12.2> LA지역 Ozone 농도의 시계열그림(1955.1 - 1972.12)

- 대기오염을 줄이기 위한 여러 가지 조치

1960년에 개통된 Golden State 고속도로

휘발유에 포함된 대기오염의 원인인 활성 탄화수소의 양을 제한하는 법

1966년부터 판매되는 자동차는 오존의 발생량을 줄이는 엔진을 필수적으로 장착해야 하는 법안 통과

일련의 조치들이 LA 지역의 오존 발생량에 변화를 가져왔는가? 개입분석 적용

개입이 있기 전인 1955년부터 1959년까지의 시계열자료를 이용한 모형

$$(1 - B^{12})Z_t = (1 - \theta B)(1 - \Theta_{12}B^{12})\varepsilon_t$$

- 개입의 형태를 감안한 개입변수의 정의

고속도로의 개통과 휘발유에 포함된 오염원을 제한하는 법안

1960년 1월 이후 지속적으로 영향을 미친다고 보아 계단함수 형태로 정의

$$I_{1t} = \begin{cases} 0, & t < \text{January 1960} \\ 1, & t \geq \text{January 1960} \end{cases}$$

1966년부터 판매되는 자동차부터 새로 장착되는 엔진의 효과
여름철과 겨울철의 오존 발생에 미치는 영향이 다르다고 보아 계단함수로 정의

$$I_{2t} = \begin{cases} 1, & \text{summer (1966년 이후 6월 ~ 10월)} \\ 0, & \text{otherwise} \end{cases}$$

$$I_{3t} = \begin{cases} 1, & \text{winter (1966년 이후 11월 ~ 5월)} \\ 0, & \text{otherwise} \end{cases}$$

개입변수들을 포함한 시계열모형

$$Z_t = w_1 I_{1t} + \frac{w_2}{(1-B^{12})} I_{2t} + \frac{w_3}{(1-B^{12})} I_{3t} + \frac{(1-\theta_1 B)(1-\Theta_{12} B^{12})}{(1-B^{12})} \varepsilon_t$$

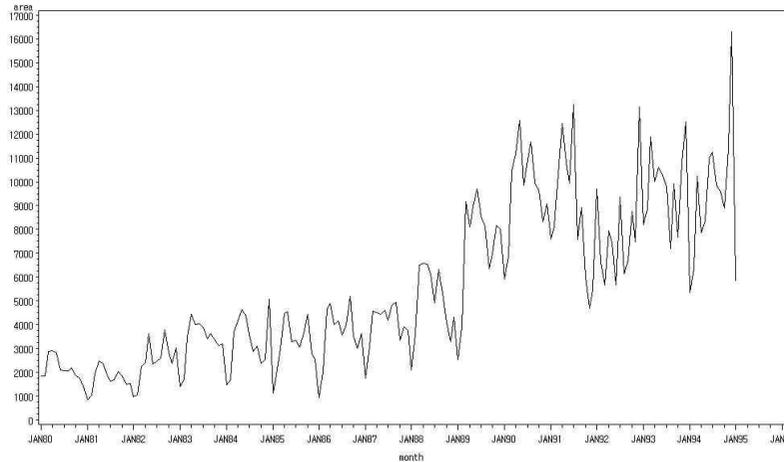
적합 결과 추정된 모수들, 괄호 속의 값은 표준편차

$$w_1 = -1.331 (0.192), \quad w_2 = -0.239 (0.060), \quad w_3 = -0.080 (0.050)$$

$$\theta_1 = -0.267 (0.067), \quad \Theta_{12} = 0.767 (0.060)$$

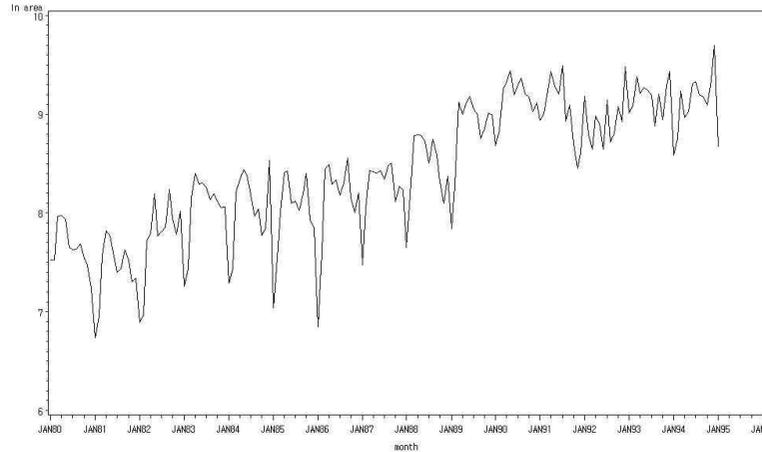
개입모형의 오차가 정규분포를 따른다는 가정 하에서 모수의 추정값들은 t -분포
개입효과들은 I_{3t} 를 제외하고는 모두 유의

예제 12.2 (건축허가면적) 1980년 1월부터 1995년 1월까지 건축허가면적 자료 이용한 개입분석
 6공화국 당시 분당, 일산 등의 건설과 관련된 주택 200만호 건설(1989년 2월 ~ 1991년 8월)이
 건축허가면적 자료에 영향을 미쳤는가?



<12.3> 건축허가면적 자료의 시계열그림 (1980.1-1995.1)

<12.4> 로그변환된 건축허가면적 자료의 시계열그림



1980년 1월부터 1989년 1월까지의 자료를 이용하여 적합한 시계열모형

$$(1-B)(1-B^{12})Z_t = (1-\theta B)(1-\Theta_{12}B^{12})\varepsilon_t$$

200만호 건설이라는 사건을 계단함수로 정의

$$S_t = \begin{cases} 0, & t = 1980.1 - 1989.1, 1991.9 - 1995.1 \\ 1, & t = 1989.2 - 1991.8 \end{cases}$$

개입의 효과가 발표 다음 달부터 반영되고($b = 1$)
 이 기간 동안 시계열의 수준에 변화를 가져왔다고 판단
 $w_0 S_{t-b}$ 의 형태로 모형에 반영

$$\ln(Z_t) = 0.260S_{t-1} + \frac{(1-0.641B)(1-0.491B^{12})}{(1-B)(1-B^{12})} \varepsilon_t$$

추정된 모수들이 모두 유의하고 잔차분석 결과도 큰 문제가 없음
 200만호 건설계획이 이 기간 동안 건축허가면적의 증가를 가져왔다고 판단

예제 12.3 (홍콩 항생지수) 1982년 7월 16일 1997년에 홍콩이 중국정부에 반환된다는 뉴스가
 홍콩의 주가지수에 어떠한 영향을 미치는 지? Wei (2006).

분석에 사용된 자료 : 1981년 7월 16일부터 1983년 9월 31일까지의 항생지수
 1981년 7월 16일부터 1982년 7월 15일까지의 주가지수를 이용한 모형

$$(1-B)Z_t = \varepsilon_t$$

뉴스의 효과를 계단함수로 정의

$$I_t = \begin{cases} 0, & t < 260 \text{ (July 16, 1982)} \\ 1, & t \geq 260 \text{ (July 16, 1982)} \end{cases}$$

뉴스가 주가의 수준에 즉각적인 영향을 미친다고 판단

$$Z_t = w_0 I_t + \frac{1}{(1-B)} \varepsilon_t$$

w_0 의 값이 유의하지 않아 개입의 효과는 없다고 판단

즉, 홍콩을 반환한다는 뉴스는 이미 오래 전부터 알려져 있던 악재로서 주가 수준에 큰 영향을 미치지 않은 것으로 보임

12.2 이상점 탐지(outlier detection)

이상점(outlier) : 분석에 영향은 미치나 그 원인을 모르는 자료

이상점 탐지(outlier detection) : 이상점의 존재 여부를 판단

파업, 정책변화 및 시장구조의 급격한 변화 등과 같은 외적 요건 파악
자료를 잘못 읽거나 기입하는 과정에서 발생할 수 있는 오류 발견

이상점의 원인이 규명된다면 이를 개입의 형태로 보아 개입모형을 적합,
개입모형을 이용한 이상점을 탐지

Tsay (1988)와 Chang 등 (1988)

12.2.1 시계열에서의 이상점(outliers in time series), Fox (1972)

Z_t : 이상점에 의한 오염 가능성이 있는 실제 관측된 자료

X_t 는 이상점에 의해 오염되지 않은(outlier-free) 시계열, 관측할 수 없음

$\phi(B)X_t = \theta(B)\varepsilon_t$, ε_t 는 $N(0, \sigma_\varepsilon^2)$ 백색잡음

$$Z_t = X_t + w_i v_i(B) I_t^{(T)}, \quad i = A, I, L, C$$

w_i : 이상점의 효과

$v_i(B)$: 이상점의 종류에 따라 다른 형태를 갖는 함수

(1) Additive Outlier (AO)

이상점 효과 w_A 가 T 시점에서 오염되지 않은 시계열에 더해진 형태

$$Z_t = \begin{cases} X_t, & t \neq T \\ X_t + w_A, & t = T \end{cases}$$

또는

$$Z_t = X_t + w_A I_t^{(T)}, \quad \text{지시함수 } I_t^{(T)}$$

$$Z_t = \psi(B)\epsilon_t + w_A I_t^{(T)}, \quad \psi(B) = \theta(B)/\phi(B)$$

$$\pi(B)(Z_t - w_A I_t^{(T)}) = \epsilon_t, \quad \pi(B) = \phi(B)/\theta(B)$$

=> AO : 한 시점에서만 시계열의 관측값에 영향을 미치는 이상점
시계열을 관측하는 과정에서 발생하는 오차(measurement error)

자료를 옮겨 적는 과정에서 발생하는 오차(recording error)
 외생적 요인(exogenous cause)에 의한 효과를 반영하는 오차
 모수의 추정량 뿐만 아니라 잔차에도 영향을 미침

$$v(B) = 1$$

(2) Innovational Outlier (IO)

한 시점의 시계열의 관측값에만 영향을 미치지 않고 이상점이 발생한 이후의 관측값에 지속적인 영향을 미치는 형태의 이상점

$$Z_t = \begin{cases} X_t, & t \leq T \\ X_t + w_I \psi_j, & t = T+j, \quad j \geq 0 \end{cases}$$

또는

$$Z_t = \frac{\theta(B)}{\phi(B)} (\epsilon_t + w_I I_t^{(T)})$$

$$Z_t = X_t + \psi(B) w_I I_t^{(T)}$$

$$\pi(B) Z_t = w_I I_t^{(T)} + \epsilon_t$$

=> IO : ψ_j 를 통해 T 시점 이후의 모든 관측값 $Z_t, t \geq T$ 에 w_I 라는 효과가 더해짐
 오차 ε_T 에 영향을 주는 내생적 원인(endogenous cause)에 의해 발생하는 오차 X_t 의 $ARIMA$ 모형의 효과를 반영 이상점이 발생한 이후의 모든 관측값에 영향
 연속적으로 발생한 일련의 AO에 의한 효과로도 볼 수 있음, Tsay (1988).

$$v(B) = \frac{\theta(B)}{\phi(B)} = \psi(B)$$

(3) Level Shift (LS), Tsay (1988)

수준변화(LS) 이상점 : T 시점 이후의 모든 관측값의 수준에 항구적인 변화가 발생하는 경우

$$Z_t = \begin{cases} X_t & t < T \\ X_t + w_L, & t \geq T \end{cases}$$

또는

$$Z_t = X_t + \frac{1}{1-B} w_L I_t^{(T)}, \quad I_t^{(T)} = P_t^{(T)} : \text{지시함수}$$

$$= X_t + w_L S_t^{(T)}, \quad S_t^{(T)} : \text{계단함수}$$

$$\pi(B)(Z_t - w_L S_t^{(T)}) = \varepsilon_t$$

=> LS : T 시점 이후의 모든 관측값에 동일한 효과 w_L 을 더해 주어 시계열의 수준이 변화
 시스템에 변화가 오거나 관측기기에 변화가 생기는 경우
 IO의 효과가 ψ_j 에 의해 점차로 줄어드는 반면 LS는 항구적인 수준의 변화
 AO는 한 시점의 수준에만 변화가 생기는 반면 LS는 항구적인 수준의 변화

$$v(B) = \frac{1}{1-B}$$

(4) Temporary Change (TC), Tsay (1988)

관측값의 수준에 생긴 변화가 단기간 지속되고 또한 변화의 크기도 점차적으로 줄어드는 경우

$$Z_t = X_t + \frac{1}{1-\delta B} w_C I_t^{(T)}, \quad 0 < \delta < 1$$

$$\pi(B)(Z_t - \frac{1}{1-\delta B} w_C I_t^{(T)}) = \varepsilon_t$$

=> TC : 초기에는 시계열의 수준에 영향을 주나 점차적으로 그 효과가 줄어드는 경우
감쇠(damping) 효과 δ 의 크기에 의해 지속적으로 감소
지속되는 기간과 크기가 δ 의 크기에 의해 좌우됨

$$v(B) = \frac{1}{1-\delta B}$$

참고 : AO, LS와 TC의 관계

$$\text{AO} : \delta \simeq 0$$

$$\text{LS} : \delta \simeq 1$$

(5) Consecutive AO's or Patches of Outliers, Bruce and Martin (1989), Lee (1990)

이상점이 연속적으로 발생하는 경우 연속적인 AO가 발생했다고 생각

$$Z_t = X_t + \sum_{j=1}^m w_j I_t^{(T_j)}$$

IO와 TC 등이 연속적인 AO에 의해 근사 될 수 있음

하나의 시계열에는 여러 종류의 이상점이 영향을 미치는 경우, Chang (1982)과 Tsay (1988)

$$Z_t = X_t + \sum_{i=1}^k w_i v_i(B) I_t^{(T_i)}$$

12.2.2 이상점 탐지(outlier detection)

이상점이 의심되는 경우 이상점의 종류와 발생시점을 찾아내어 이를 모형에 반영
오염되지 않은 시계열의 모형은 알고 있다고 가정, 즉, $\phi(B)$ 와 $\theta(B)$ 를 안다고 가정
이상점의 발생시점을 찾아내고 이상점의 종류를 결정

Tsay (1988) : 이상점이 잔차에 미치는 영향을 고려

(1) 이상점이 잔차에 미치는 영향

오염되지 않은 시계열 X_t 의 모형을 안다고 가정하고 이상점이 잔차(residual)에 미치는 영향
: $\pi(B)$ 를 이용하여 추정

$$e_t = \pi(B)Z_t \quad , \quad \pi(B) = \phi(B)/\theta(B) = (1 - \pi_1 B - \pi_2 B^2 - \dots)$$

● $\pi(B)$ 를 이용한 이상점의 효과

$$AO : e_t = w_A \pi(B) I_t^{(T)} + \varepsilon_t$$

$$IO : e_t = w_I I_t^{(T)} + \varepsilon_t$$

$$LS : e_t = w_L \frac{1}{1-B} \pi(B) I_t^{(T)} + \varepsilon_t$$

$$TC : e_t = w_C \frac{1}{1-\delta B} \pi(B) I_t^{(T)} + \varepsilon_t$$

이상점의 종류에 따라 $Y_{i,t}$, ($i = A, I, L, C$) 정의

$$Y_{i,(T+k)} = \begin{cases} -\pi_k, & AO \\ 0, & IO \\ 1 - \sum_{j=1}^k \pi_j, & LS \\ \delta^k - \sum_{j=1}^{k-1} \delta^{k-j} \pi_j - \pi_k, & TC \end{cases}$$

잔차 : $e_t = w_i Y_{i,t} + \varepsilon_t$

AO의 경우 :

$$\begin{bmatrix} e_1 \\ \vdots \\ e_{T-1} \\ e_T \\ e_{T+1} \\ \vdots \\ e_n \end{bmatrix} = w_A \begin{bmatrix} 0 \\ \vdots \\ 0 \\ 1 \\ -\pi_1 \\ \vdots \\ -\pi_{n-T} \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_{T-1} \\ \varepsilon_T \\ \varepsilon_{T+1} \\ \vdots \\ \varepsilon_n \end{bmatrix}$$

를 설명변수로 하고 w_i 를 모수로 하는 회귀모형을 적합

최소제곱법에 의한 이상점의 효과 \hat{w}_A , \hat{w}_I 와 분산 $Var(\hat{w}_A)$, $Var(\hat{w}_I)$

$$\hat{w}_A = \frac{e_t - \sum_{j=1}^{n-T} \pi_j e_{T+j}}{\sum_{j=0}^{n-T} \pi_j^2}, \quad Var(\hat{w}_A) = \frac{\sigma_\varepsilon^2}{\sum_{j=0}^{n-T} \pi_j^2}$$

$$\hat{w}_I = e_T, \quad Var(\hat{w}_I) = Var(e_T) = Var(w_I I_t^{(T)} + \varepsilon_t) = \sigma_\varepsilon^2$$

(2) 검정통계량(Test statistics)

잔차를 이용한 각종 진단에서의 어려운 점

- ① 실제로 이상점이 발생했는지, 발생했다면 그 발생시점이 언제인가를 알 수 없다.
- ② 이상점이 발생했다 하더라도 이상점의 종류가 어떤 것 인지를 알 수 없다.
- ③ 여러 개의 이상점이 발생할 수 있다.
- ④ 마지막으로 실제모형(true model)을 알 수 없어 추정된 모형도 효율이 떨어질 수 있다.

①-③ : 적합된 모형의 잔차를 이용하여 그 종류를 구분

④ : 반복적인 절차를 필요로 함

● 이상점의 탐지를 위한 가설검정

H_0 : Z_T 는 AO, IO, LS, TC가 아니다, 즉, 이상점이 아니다.

vs H_1 : Z_T 는 AO, H_2 : Z_T 는 IO, H_3 : Z_T 는 LS, H_4 : Z_T 는 TC

Tsay (1988) : 가능도비검정통계량(likelihood ratio test statistics) 제안

$$\lambda_{i,t} = \frac{\hat{w}_{i,t}}{\sqrt{\text{Var}(\hat{w}_{i,t})}}, \quad i = AO, IO, LS, TC, \quad t = 1, 2, \dots, n$$

t : 이상점의 발생시점

● Tsay(1988)에 의한 반복 추정 절차(iterative estimation procedure)

이상점이 없다는 가정 하에서 σ_ϵ^2 를 포함한 모수를 추정, 잔차 구함

$$\hat{e}_t = \hat{\pi}(B)Z_t = \frac{\hat{\phi}(B)}{\hat{\theta}(B)}Z_t, \quad \hat{\sigma}_\epsilon^2 = \frac{1}{n} \sum_{t=1}^n \hat{e}_t^2$$

다음의 단계를 거쳐 이상점을 탐지하고 모형에 반영

- ① 이상점의 탐지(outlier detection)
- ② 이상점의 조정(outlier adjustment)
- ③ 조정된 시계열을 이용한 모수의 추정

1 단계

- ① 4 종류의 검정통계량의 값을 구함, $\lambda_{i,t}$, $t = 1, 2, \dots, n$, $i = IO, AO, LS, TC$
- ② 절대값이 가장 큰 검정통계량 $\hat{\lambda}_T$ 를 선택

$$\hat{\lambda}_T = \max_t \max_i \{ |\hat{\lambda}_{i,t}| \}$$

$\hat{\lambda}_T > C$: 미리 결정된 기각값(critical value), 경험적으로 3과 4 사이의 값을 추천
=> 시점 T 에서 선택된 종류의 이상점이 발생했다고 판단

$\hat{\lambda}_T < C$: 이상점이 없다고 판단

- ③ 이상점의 효과를 추정하고 이를 이용하여 잔차를 조정(adjust)
조정된 잔차를 이용하여 σ_ϵ^2 의 추정량 $\tilde{\sigma}_\epsilon$ 를 구함

$$\tilde{Z}_t = Z_t - \hat{w}_i \frac{\omega(B)}{\delta(B)} I_t^{(T)}$$

$$\tilde{e}_t = \hat{e}_t - \hat{w}_i \frac{\omega(B)}{\delta(B)} \hat{\pi}(B) I_t^{(T)}, \quad \tilde{\sigma}_\epsilon^2 = \frac{1}{n} \sum_{t=1}^n \tilde{e}_t^2$$

2 단계

- ① 조정된 잔차와 $\tilde{\sigma}_\varepsilon^2$ 를 이용하여 $\hat{\lambda}_{i,t}$ 를 구함
- ② 모든 이상점이 탐지될 때까지 1 단계를 반복적으로 수행

3 단계

- ① T_1, \dots, T_k 시점에서 k 개의 이상점을 잠정적으로 식별한다.
- ② 아래의 모형을 이용하여 이상점의 효과 w_1, w_2, \dots, w_k 와 시계열모형의 모수를 동시에 추정

$$Z_t = \sum_{j=1}^k w_j v_j(B) I_t^{(T_j)} + \frac{\theta(B)}{\phi(B)} \varepsilon_t$$

새로운 잔차를 구하고 분산을 추정

$$\tilde{e}_t^{(1)} = \hat{\pi}^{(1)}(B) [Z_t - \sum_{j=1}^k \hat{w}_j v_j(B) I_t^{(T_j)}] \text{ 와 } \tilde{\sigma}_\varepsilon^2$$

마지막으로 추가적인 이상점이 탐지되지 않을 때까지 1-3 단계를 반복

$$Z_t = \sum_{j=1}^k \hat{w}_j \hat{v}_j(B) I_t^{(T_j)} + \frac{\hat{\theta}(B)}{\hat{\phi}(B)} \varepsilon_t$$

이상점 탐지절차의 문제점

① 기각값(critical value)은 어떻게 결정하는가?

검정통계량의 분포를 알 수 없어 정확한 유의 수준을 만족하는 기각값을 구할 수 없음
Chang (1982)과 Tsay (1988) 등은 3.0 과 4.0 사이의 값을 추천

② 탐지절차가 이상점이 없다고 가정하고 식별된 ARMA 모형에 기초하므로 반복 추정절차가 효율적이 아닐 수 있음

이상점이 존재할 경우 추정된 모수들이 편향(bias)될 수 있어 이상점 탐지를 위한 반복 추정절차가 영향을 받음

이상점에 의해 영향을 받지 않는 로버스트 추정방법을 사용하는 것이 대안, Matin (1980)

③ 이상점들의 다른 이상점의 효과와 중첩되어(mask) 식별되지 않을 수 있음

예제 12.1 (계속) ozone 자료를 이용한 이상점 탐지

60 번째에 해당되는 1960년 1월부터 -1.31 만큼 수준에 변화가 있는 것으로 탐지
이 시점부터 여러 가지 형태의 변화가 있었음을 확인하고 적합한 형태의 개입모형을 적합
11번째와 21번째 관측값도 AO로 탐지 : 특별한 이유를 발견할 수 없음

관측값을 옮겨 적는 과정이나 관측기기에 혹시 문제가 있었는지 자세한 조사가 필요함

12.3 SAS 프로그램 및 설명

12.3.1 개입분석을 위한 SAS 프로그램

```
Proc ARIMA ;  
  identify var=반응변수 crosscorr=개입변수 ;  
  estimate p=1 q=1 input=(  $b \ \$ (w-lags) / (\delta-lags)$  개입변수  
) ;  
  forecast options ;  
run;
```

(A) $wB^2P_t^{(T)}$: input = (2 \$ P)

$wB^2S_t^{(T)}$: input = (2 \$ S)

(B) $\frac{w}{(1-\delta B)}BP_t^{(T)}$: input = (1 \$ /(1) P)

$\frac{w}{(1-\delta B)}B^2S_t^{(T)}$: input = (2 \$ /(1) S)

$$(C) \quad \frac{w}{1-B} B^2 S_t^{(T)} = w B^2 T_t^{(T)}, \quad T_t^{(T)} = \begin{cases} t-T & t \geq T \\ 0 & o.w \end{cases} : \text{input} = (2 \$ T)$$

$$(D) \quad \left(\frac{w_0}{1-\delta B} + \frac{w_1}{1-B} \right) B^2 P_t^{(T)} = \frac{w_0}{1-\delta B} B^2 P_t^{(T)} + w_1 B^2 S_t^{(T)} :$$

input = (2 \$ / (1) P , 2 \$ S)

$$\left(w_0 + \frac{w_1}{1-\delta B} B^2 + \frac{w_2}{1-B} B^2 \right) P_t^{(T)} = w_0 P_t^{(T)} + \frac{w_1}{1-\delta B} B^2 P_t^{(T)} + w_2 B^2 S_t^{(T)} :$$

input = (P , 2 \$ / (1) P , 2 \$ S)

12.3.2 이상점의 탐지를 위한 SAS 프로그램

3 종류의 이상점(AO, LS, TC)만을 지정할 수 있음

```
proc ARIMA ;
  identify var=variables ;
  estimate p=1 q=1 ;
  outlier options ;
  forecast options ;
run;
```