

# 가법계절지수모형에서 예측에 미치는 이상치의 영향

편영숙<sup>1)</sup> 이재준<sup>2)</sup>

## 요약

시계열자료에는 흔히 대부분의 자료에서 벗어나는 이상치들이 포함되어 있는데, 이러한 자료는 관측치들 사이의 종속구조로 인해 분석과정에 영향을 끼칠 수 있고, 특히 연속시점에서 발생하는 경우에 그 영향이 매우 심각할 수 있다. 본 논문에서는 연속이상치(PO)가 Winters 계절지수모형의 분석과정에 미치는 영향을 유도하고, 예측 평균제곱오차(MSFE)를 구하여 연속이상치가 예측에 미치는 영향을 제시하였다. 또한, 실제자료를 이용하여 연속이상치의 영향을 실증적으로 분석하였다.

## 1. 서론

시계열 자료에는 개입(intervention)사건이나 단순한 입력오류에 의해 대부분의 자료에서 벗어나는 관측치들(outliers)이 포함될 수 있다. 이상치는 통계적 추론과정에 심각한 영향을 끼칠 수 있는데, 특히 시계열자료 사이에는 관측순서에 따른 특정의 종속관계가 존재하기 때문에 독립적 자료에서 보다 그 영향이 훨씬 더 심각할 수 있다. 따라서 보다 정확한 분석결과를 얻기 위해 이상치가 발생한 시점과 이유를 파악하고, 그 영향을 고려한 분석방법이 요구된다.

시계열의 이상치에 관련된 연구는 Fox(1972)가 AO(additive outlier), IO(innovation outlier) 등 단일이상치(single outliers)의 모형을 제안한 이래, 주로 ARIMA모형 하에서 단일이상치의 탐지방법이나 모수의 추정에 미치는 영향에 관한 연구가 주를 이루었다 (Chang 등(1988), Tsay(1988)). 시계열자료에서는 연속시점에서 이상치(연속이상치; patchy outliers(PO))가 발생하는 경우가 많은데, 이러한 이상치들이 분석과정에 미치는 영향은 더욱 심각하다고 알려지고 있다 (Martin과 Yohai(1986), Bruce와 Martin(1989)).

이상치가 예측에 미치는 영향에 관한 연구는 주로 참시계열이 ARIMA모형을 따르는 경우에 대하여 연구결과가 제시되었다. Ledolter(1988)는 ARIMA모형하에서 AO가 시계열 예측값에 미치는 영향의 크기를 유도하였고, Chen과 Liu(1993)은 이상영향이 포함된 자료와 이상영향이 제거된 참시계열(outlier free series)로부터 각각 계산된 두 예측값들의 차와 평균제곱근오차(root mean squared error)의 차를 이용하여 단일이상치가 예측에 미치는 영향을 구체적으로 제시한 바 있다. 이러한 결과는 AO, IO, LC(level change), TC(transient change)등 예측에 미치는 단일이상치의 영향에 국한되었는데, Lee 등(1996)은 ARIMA모형을 따르는 참시계열에 연속시점에서 발생한 이상영향이 포함되어 관측된 시계열을 개입모형으로 모형화하고, 이 모형을 이용하여 예측에 미치는 이상치들의 영향을 유도하여 제시한 바 있다.

1) (402-751) 인천시 남구 용현동 253, 인하대학교 통계학과, 석사

2) (402-751) 인천시 남구 용현동 253, 인하대학교 통계학과, 교수

평활법은 *ARIMA* 모형에 의한 예측에 비하여 예측의 정확성은 떨어지지만, 모형의 선택이나 모수의 추정등이 간편하여 아직도 널리 사용되는 예측방법이다. 특히, 평활법은 자료가 하나씩 관측될 때마다 갱신식에 의해 기계적으로 예측이 수행될 수 있고 예측에 소요되는 시간이 매우 짧기 때문에, 많은 변수를 단시간에 예측해야 하는 경우와 같이 *ARIMA* 모형이 이용될 수 없을 때 흔히 사용되는 방법이다. 평활법을 이용한 예측에서 일반적으로 단일이상치가 예측원점에서 발생한 경우에 그 영향이 심각할 수 있지만, 그러한 영향의 크기는 평활상수에 의해 결정되게 된다. 단순지수평활모형에서 단일이상치가 예측에 미치는 영향은 *IMA(1,1)* 모형에 의한 예측결과가 단순지수평활모형과 같다는 사실을 이용하여 제시된 바 있지만(Chen과 Liu(1993)), 일반적인 지수평활모형에서 연속이상치가 예측에 미치는 영향에 관한 연구결과는 전무한 실정이다. Winters의 계절지수모형에는 가법계절지수모형(Additive Seasonal Model: ASM)과 승법계절지수모형(Multiplicative Seasonal Model: MSM)이 있는데(Montgomery 등(1990)), 본 논문에서는 Winters의 가법계절지수평활모형에서 *PO*가 각 변동요인의 추정값과 시계열 예측값에 미치는 영향을 밝히고, 실제자료를 이용하여 그러한 연속이상치의 영향을 분석하였다.

## 2. 연속이상치가 예측에 미치는 영향

연속이상치는  $m$ 개의 연속 시점에서 *AO*가 발생한 것으로 정의될 수 있고(Lee 등(1996)), 따라서 시점  $T$ 에서  $T+m-1$ 까지  $m$ 개 시점에서 연속이상치(*PO*)가 포함된 관측시계열  $Y_t$ 는 다음과 같은 개입모형으로 표현될 수 있다.

$$Y_t = Z_t + \sum_{j=0}^{m-1} w_j I_t(T+j), t = 1, 2, \dots, n \quad (2.1)$$

단,  $Z_t$ 는 이상치가 없는 참시계열,  $Y_t$ 는 참시계열에 이상영향이 포함된 관측시계열,  $T$ 는 연속이상치 발생시점,  $w_j$ 는  $T+j$  시점 관측치에 내재된 이상영향의 크기, 그리고  $I_t(T)$ 는  $t = T$ 에서 1이고  $t \neq T$ 에서 0인 지시함수이다.

평활법은 주로 시계열의 불규칙 변동을 제거하고 추세를 파악하기 위해 사용되는데, 일반적으로 이상치가 존재하더라도 예측치는 영향을 크게 받지 않는 편이다. 그러나, Winters의 계절지수모형은 주기를 갖는 계절요인과 추세요인등이 포함되어 있기 때문에, 단순평활법과는 달리 훨씬 긴 시점들에서도 영향을 받게 될 것으로 예상된다. 본 논문에서는 *ASM*모형을 이용한 예측에서 *PO*가 각 요인들과 예측값에 미치는 영향을 구체적으로 유도하여 제시하였다.

### 2.1. WINTERS 가법계절모형

참시계열  $Z_t$ 는 주기가  $L$ 인 *ASM*을 따른다고 하자. 추세가 선형인 경우에  $Z_t$ 는 다음과 같이 표현된다.

$$\begin{aligned} Z_t &= M_t + S_t + I_t \\ &= (\beta_0 + \beta_1 t) + S_t + I_t \end{aligned} \quad (2.2)$$

단,  $M_t$ 는 절편과 기울기가 각각  $\beta_0, \beta_1$  인 선형추세요인,  $S_t$ 는 주기가  $L$ 인 계절요인, 그리고  $I_t$ 는 오차항으로서 평균이 0이고 분산이  $\sigma^2$  인 백색잡음과정을 따르는 불규칙요인이다. 식 (2.2)에서 계절요인  $S_t$ 는  $S_{t+L} = S_t$ 와  $\sum_{j=1}^L S_j = 0$  을 만족한다고 가정한다. 시점  $t$ 에서 계산되는  $M_t, \beta_1, S_t$  (수준(level), 추세(trend), 계절(season) 요인)의 추정치,  $\hat{\beta}_0(t), \hat{\beta}_1(t), \hat{S}(t)$ 는 다음과 같은 갱신식(undating equation)으로 표현된다.

$$\begin{aligned}\hat{\beta}_0(t) &= \gamma_1(Z_t - \hat{S}(t-L)) + (1-\gamma_1)(\hat{\beta}_0(t-1) + \hat{\beta}_1(t-1)) \\ \hat{\beta}_1(t) &= \gamma_2(\hat{\beta}_0(t) - \hat{\beta}_0(t-1)) + (1-\gamma_2)\hat{\beta}_1(t-1) \\ \hat{S}(t) &= \gamma_3(Z_t - \hat{\beta}_0(t)) + (1-\gamma_3)\hat{S}(t-L)\end{aligned}\quad (2.3)$$

단,  $\gamma_1, \gamma_2, \gamma_3$ 는 각각 해당 부분에 적용되는 평활상수(smoothing constant)이다. 식 (2.3)에서 일반적으로  $t$ 시점 수준(level)의 추정치는  $\hat{M}(t)$ 로 표기되는데, 본 논문에서는 현시점 좌표(current origin)의 기준에서 절편  $\beta_0$ 가 원시점 좌표(original origin)의 수준에 해당되므로  $t$ 시점의 수준 추정치를  $\hat{\beta}_0(t)$ 로 표기하였다.

## 2.2. 연속이상치가 예측과정에 미치는 영향

예측에 관련된 기호로서,  $T$ 를 이상치 발생시점,  $n$ 을 예측원점,  $l$ 을 예측시차(lead time),  $m$ 을  $PO$ 의 길이로 표기한다.  $PO$ 가 예측에 미치는 영향에 관한 문제는 정형화된 구조를 가정하지 않는  $PO$ 의 특성에 따라 관측치  $Y_t$ 에 포함될 이상영향을 구체적으로 미리 표현할 수 없고, 따라서 예측 대상은 참시계열  $Z_t$ 에 국한한다. 이 때, 예측원점  $n$ 에서  $l$ 시점 후의 참시계열값  $Z_{n+l}$ 의 예측값은

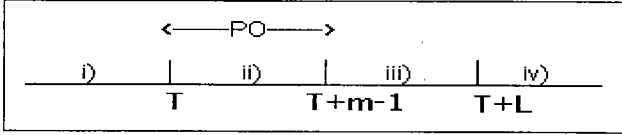
$$\hat{Z}_n(l) = (\hat{\beta}_0(n) + \hat{\beta}_1(n) \times l) + \hat{S}(n+l-L) \quad (2.4)$$

가 되는데, 자료에  $PO$ 가 포함되어 있는 경우에  $PO$ 의 영향을 제거하고 구한 모수의 추정값  $\hat{\beta}_0(n)_{PO}, \hat{\beta}_1(n)_{PO}, \hat{S}(n)_{PO}$ 를 식 (2.4)에 적용하여 예측값을 구하거나,  $PO$ 를 제거하지 않고 구한 추정값  $\hat{\beta}_0(n)_{NO}, \hat{\beta}_1(n)_{NO}, \hat{S}(n)_{NO}$ 으로 구하는 두 경우를 고려할 수 있다. 그러면, 식(2.3)으로 계산되는 두 경우의  $\beta_0, \beta_1, S_t$ 의 추정치는 다음과 같은 관계로 표현될 수 있다.

$$\begin{aligned}\hat{\beta}_0(t)_{PO} &= \hat{\beta}_0(t)_{NO} - \Delta\beta_0(t) \\ \hat{\beta}_1(t)_{PO} &= \hat{\beta}_1(t)_{NO} - \Delta\beta_1(t) \\ \hat{S}(t)_{PO} &= \hat{S}(t)_{NO} - \Delta S(t)\end{aligned}\quad (2.5)$$

여기에서, 시점  $t$ 에서  $PO$ 가 모수 추정치에 미치는 영향의 크기를 나타내는  $\Delta\beta_0(t), \Delta\beta_1(t), \Delta S(t)$ 는  $PO$ 의 발생시점( $T$ )와 길이( $m$ ), 계절요인의 주기( $L$ )에 따라 다음과 같이 표현될 수 있다.

Case 1 :  $m \leq L$  ( $PO$ 의 길이가 주기보다 짧은 경우)



i)  $PO$  발생전 ( $t \leq T - 1$ )

$$\Delta\beta_0(t) = \Delta\beta_1(t) = \Delta S(t) = 0$$

ii)  $PO$  발생중 ( $T \leq t \leq T + m - 1$ )

$$\Delta\beta_0(t) = \gamma_1\omega_{t-T} + (1 - \gamma_1)(\Delta\beta_0(t-1) + \Delta\beta_1(t-1))$$

$$\Delta\beta_1(t) = \gamma_2(\Delta\beta_0(t) - \Delta\beta_0(t-1)) + (1 - \gamma_2)\Delta\beta_1(t-1)$$

$$\Delta S(t) = \gamma_3(\omega_{t-T} - \Delta\beta_0(t))$$

iii)  $PO$  종료후이고  $PO$  시작 한 주기( $L$ ) 이내 ( $T + m \leq t \leq T + L - 1$ )

$$\Delta\beta_0(t) = (1 - \gamma_1)(\Delta\beta_0(t-1) + \Delta\beta_1(t-1))$$

$$\Delta\beta_1(t) = \gamma_2(\Delta\beta_0(t) - \Delta\beta_0(t-1)) + (1 - \gamma_2)\Delta\beta_1(t-1)$$

$$\Delta S(t) = -\gamma_3\Delta\beta_0(t)$$

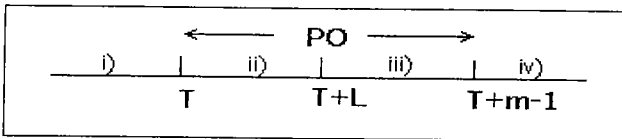
iv)  $PO$  시작시점에서 한 주기( $L$ ) 이후 ( $t \geq T + L$ )

$$\Delta\beta_0(t) = -\gamma_1\Delta S(t-L) + (1 - \gamma_1)(\Delta\beta_0(t-1) + \Delta\beta_1(t-1))$$

$$\Delta\beta_1(t) = \gamma_2(\Delta\beta_0(t) - \Delta\beta_0(t-1)) + (1 - \gamma_2)\Delta\beta_1(t-1)$$

$$\Delta S(t) = -\gamma_3\Delta\beta_0(t) + (1 - \gamma_3)\Delta S(t-L)$$

Case 2 :  $m \geq L$  ( $PO$ 의 길이가 주기보다 긴 경우)



i)  $PO$  발생전 ( $t \leq T - 1$ )

$$\Delta\beta_0(t) = \Delta\beta_1(t) = \Delta S(t) = 0$$

ii)  $PO$  발생중이고  $PO$  시작 한 주기 이내 ( $T \leq t \leq T + L - 1$ )

$$\Delta\beta_0(t) = \gamma_1\omega_{t-T} + (1 - \gamma_1)(\Delta\beta_0(t-1) + \Delta\beta_1(t-1))$$

$$\Delta\beta_1(t) = \gamma_2(\Delta\beta_0(t) - \Delta\beta_0(t-1)) + (1 - \gamma_2)\Delta\beta_1(t-1)$$

$$\Delta S(t) = \gamma_3(\omega_{t-T} - \Delta\beta_0(t))$$

iii)  $PO$  발생중이고  $PO$  시작 한 주기( $L$ ) 이후 ( $T + L \leq t \leq T + m - 1$ )

$$\Delta\beta_0(t) = \gamma_1(\omega_{t-T} - \Delta S(t-L)) + (1 - \gamma_1)(\Delta\beta_0(t-1) - \Delta\beta_1(t-1))$$

$$\Delta\beta_1(t) = \gamma_2(\Delta\beta_0(t) - \Delta\beta_0(t-1)) + (1 - \gamma_2)\Delta\beta_1(t-1)$$

$$\Delta S(t) = \gamma_3(\omega_{t-T} - \Delta\beta_0(t)) + (1 - \gamma_3)\Delta S(t - L)$$

iv) 종료 이후 ( $t \geq T + m$ )

$$\Delta\beta_0(t) = -\gamma_1\Delta S(t - L) + (1 - \gamma_1)(\Delta\beta_0(t - 1) + \Delta\beta_1(t - 1))$$

$$\Delta\beta_1(t) = \gamma_2(\Delta\beta_0(t) - \Delta\beta_0(t - 1)) + (1 - \gamma_2)\Delta\beta_1(t - 1)$$

$$\Delta S(t) = -\gamma_3\Delta\beta_0(t) + (1 - \gamma_3)\Delta S(t - L)$$

위 Case의 유도과정은 부록 1에 제시되어 있다.

PO가  $l$ 시점 후의 예측값에 미치는 영향은 PO가 포함된 관측시계열  $\{Y_t\}$ 로 추정된  $\hat{\beta}_0(n)_{NO}, \hat{\beta}_1(n)_{NO}, \hat{S}(n+l-L)_{NO}$ 를 식 (2.4)에 적용하여 구한 예측값  $\hat{Z}_n(l)_{NO}$ 로부터 예측 오차가

$$\begin{aligned} Z_{n+l} - \hat{Z}_n(l)_{NO} &= I_{n+l} + (\beta_0 + \beta_1 \times n - \hat{\beta}_0(n)_{PO}) \\ &\quad + (\beta_1 - \hat{\beta}_1(n)_{PO}) \times l + (S_{n+l} - \hat{S}(n+l-L)_{PO}) \\ &\quad - (\Delta\beta_0(n) + \Delta\beta_1(n) * l + \Delta S(n+l-L)) \end{aligned} \quad (2.6)$$

가 되고,  $l$ 시점 후 예측오차의 평균제곱오차(MSFE)는 식 (2.7)과 같게 된다.

$$MSFE(l) = \sigma^2 \sum_{j=0}^{\infty} \alpha(l)_j^2 + (\Delta\beta_0(n) + \Delta\beta_1(n) + \Delta S(n+l-L))^2 \quad (2.7)$$

단,  $\alpha(l)_j$ 는 부록 2에서 정의된 것과 같이 표현된다. 이상치가 제거된 참시계열  $Z_t$ 를 이용하여  $Z_{n+l}$ 을 예측할 때, MSFE는  $\sum_{j=0}^{\infty} \alpha(l)_j^2$ 가 되므로 PO로 인한 MSFE의 상대적 증가량(IMSFE)는

$$IMSFE(l) = \frac{\Delta\beta_0(n) + \Delta\beta_1(n) \times l + \Delta S(n+l-L)^2}{\sigma^2 \sum_{j=0}^{\infty} \alpha(l)_j^2} \quad (2.8)$$

가 된다.

### 3. 자료분석

이 자료는 1983년 1월부터 1995년 8월까지 통계월보에 수록된 152개 월별 건축허가면적 자료로서, 분석과정에서는 제공된 변환자료(그림 3.1의 실선으로 표시)가 이용되었다. 이 자료는 그림 3.1에서 보여지듯이 1989년 2월부터 1991년 8월 사이에(75-104 번째) 주택 200만호 건설 시행으로 인해 다른 기간의 자료와 수준에서 차이(level change)가 남을 알 수 있다. 특히, 이 구간 전에는 계절적 특성이 명확하게 드러나고, 이 구간내에서는 수준변화는 있지만 계절성은 유지되는 반면에 이 구간 다음부터는 계절적 특성이 약해짐을 알 수 있다. Lee 등(1996)은 ARIMA모형하에서 이러한 정책의 영향이 시계열 예측에 미치는 영향을 제시하였다.

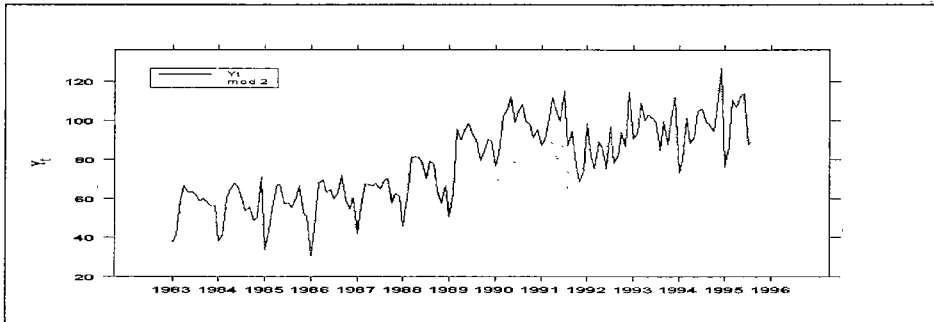


그림 3.1: 제공근 변환된 건축 허가 면적의 시계열 도표

### 3.1. 분석방법

본 분석에서는 1989년 2월부터 1991년 8월 사이의 자료를  $PO$ , 관측시계열  $Y_t$ 에서  $PO$ 의 영향  $\{\omega_j\}$ 를 제거하여 구한 시계열을 참시계열  $Z_t$ 라고 가정한다. 이 두 자료집단을 각각 Winters ASM에 적합시켜  $PO$ 가 예측과정에 미치는 영향을 분석하기로 한다. 단, Winters 계절지수모형은 모형의 특성상 이러한  $PO$ 의 구조를 명확히 밝히기 어렵기 때문에,  $PO$ 에 해당되는 구간에서 수준변화(level change)를 가정하고 그 크기를 추정하였다. 본 분석에서는  $\{\omega_j\}$ 를 추정하기 위하여, 주기와 같은 항의 이동평균(moving average : MA)를 구하고,  $PO$  구간 내에서 다른 구간의 자료에 영향을 가장 적게 받는 것으로 판단되는 구간의 MA값들의 평균으로  $\omega$ 를 추정하였다. 그 결과 89년 2월에서 90년 1월 사이에서는 모두  $\omega = 12$ 로, 90년 2월에서 91년 8월 사이는  $\omega = 22$ 로 이상영향의 크기가 추정되었다. 이로부터, 각 시점에서  $Y_t$ 에서 해당시점의  $\omega$ 를 감하여 참시계열  $Z_t$ 로 가정하였는데, 그 결과는 그림 1에서 점으로 표시된 시계열이 된다. 시점  $t$ 에서  $PO$ 가 포함된 원시계열  $\{Y_t\}$ 를 평활계수가  $\gamma_1 = \gamma_2 = \gamma_3 = 0.2$ 인 Winters ASM에 적용해 구한 모수(level, trend, season) 추정값을  $\hat{\beta}_0(t)_{NO}, \hat{\beta}_1(t)_{NO}, \hat{S}(t)_{NO}$ 라 하고,  $PO$ 의 영향이 제거된 참시계열  $\{Z_t\}$ 로 계산된 추정값들을  $\hat{\beta}_0(t)_{PO}, \hat{\beta}_1(t)_{PO}, \hat{S}(t)_{PO}$ 라 가정하였다. Winters ASM의 각 요인에 미치는  $PO$ 의 영향은 각 요인(모수)별로 그림 3.2~3.5에 제시되어 있다.

### 3.2. 분석결과

본 분석은 분석방법에서 이론상 근거가 부족한 면이 있지만  $PO$ 가 예측과정에 미치는 영향에 대해 대략 다음과 같은 사실을 확인할 수 있다.

i) 그림 3.2의 수준 갱신값들의 변화과정을 비교할 때,  $PO$ 가 포함된 자료로 구한 수준의 추정치  $\hat{\beta}_0(t)_{NO}$ 는 89년 2월부터 급격히 증가하다가 91년 8월부터 급격히 감소하는데, 그 감소경향은 92년 중반까지 미침을 확인할 수 있다. 즉,  $PO$ 를 반영하지 않은 경우에 수

준 추정치는 PO 종료시점 이후에도 장기간에 걸쳐 영향을 받음을 알 수 있는데, 이는 Case 2의 iv)에 해당된다고 할 수 있다.

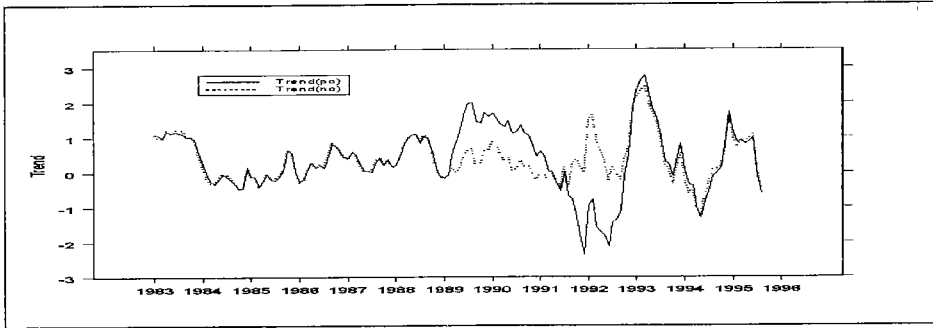


그림 3.2: trend 갱신값

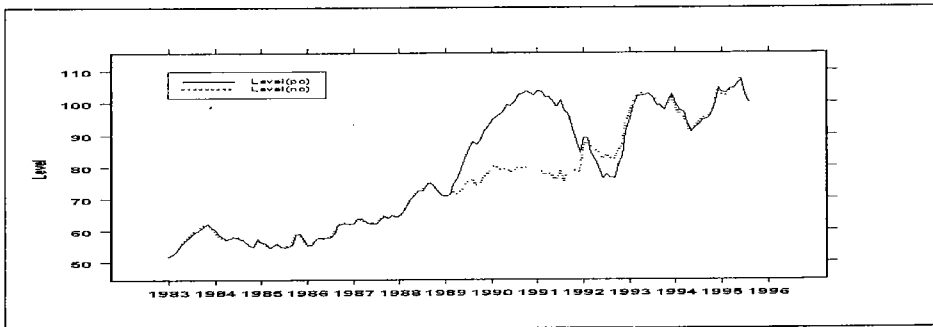


그림 3.3: level 갱신값

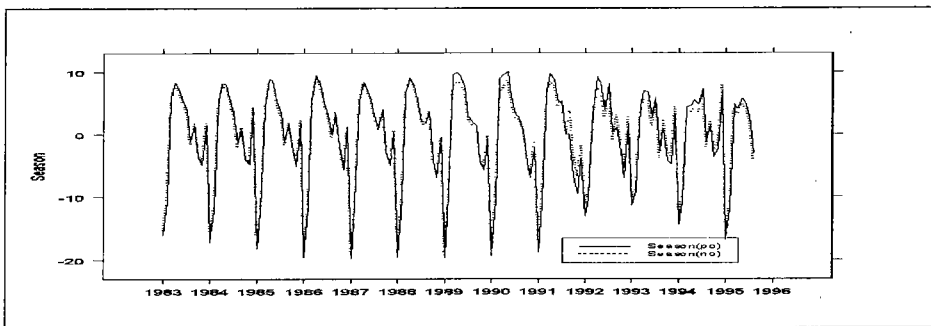


그림 3.4: season 갱신값

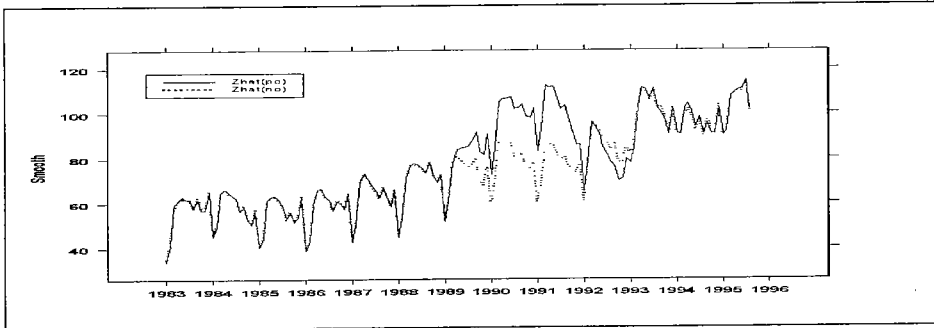


그림 3.5: 한시점 앞 예측 갱신값.

ii) 그림 3.3의 추세 갱신값들의 변화과정을 비교할 때,  $PO$ 가 포함된 자료로 구한 추세 값은 89년 2월부터 급격히 증가하며, 89년 중반부터는 별로 변화가 크지 않다가 90년 후반 부터는 급격히 감소하는 경향을 보이는데, 이 감소경향은 91년 말까지 계속되는 것으로 보인다.  $PO$ 를 고려한 추정값(점선)과 비교해 볼 때, 그 영향은 94년경까지 장기간에 걸쳐 나타나는 것으로 보인다.

iii) 그림 3.4로부터 1991년 중반부터 계절요인이 약해짐을 확인할 수 있고,  $PO$ 의 처리 여부에 따라 장기간에 걸쳐 약간의 차이가 남을 알 수 있다 ( $\Delta S(t) = -3.8 \sim 2.6$ ). 또한, 그림 3.5로부터 한 시점 앞 예측값에 미치는  $PO$ 의 영향을 대략적으로 파악할 수 있는데, 그 영향은  $PO$  발생기간 동안 증폭되며 그 이후 93년 중반까지  $PO$  종료이후 장기간에 걸쳐 나타남을 알 수 있다.

#### 4. 요약 및 토론

본 논문은 Winters 계절지수평활모형을 이용한 시계열자료의 분석과정에서  $PO$ 의 영향을 연구한 결과로서, 모형의 수준, 추세, 그리고 계절요인의 추정에서  $PO$ 를 고려한 경우와 그렇지 않은 경우로부터 계산되는 추정치들 간의 차  $\Delta\beta_0(t)$ ,  $\Delta\beta_1(t)$ ,  $\Delta S(t)$ 를 유도하였다. 본문의 Case 1과 Case 2에 요약되어 있듯이, 각 시점에서 모수 추정에 미치는  $PO$ 의 영향은  $PO$ 발생기간내 시점에서의 이상영향의 크기  $\{\omega_j\}$ 들과 이전 시점들에서 모수 추정에 미친  $PO$ 의 영향  $\Delta W(t)$ 에 의해 표현될 수 있음을 밝혔다. 특히, 계절지수모형에서는  $\Delta S(t-L)$  항에 의해  $PO$ 종료후 장기간에 걸쳐 그 영향이 미칠 수 있음을 확인할 수 있는데, 그 영향은  $PO$ 발생기간내 시점에서의 이상영향의 크기와 형태  $\{\omega_j\}$ 에 의해 결정됨을 알 수 있다.

시계열자료의 분석에서  $PO$ 의 영향은 본문의 Case들의 결과를 반복적으로 적용하여 구할 수 있다. 그러나, 본 논문에서는 상식적인 방법으로 원자료에서  $PO$ 의 영향을 제거하여 참시계열이라 가정하고, 이 두 자료집단을 모형에 적용하여  $PO$ 의 영향을 확인하였다. 이러한 분석은 이론적 근거가 부족하므로 대략적인  $PO$ 의 영향을 파악할 수 있으나 결과에 대한 신뢰성에는 한계가 있을 수밖에 없다.



## 참고문헌

- [1] Bruce, A. G. and Martin, R. D. (1989). Leave-k-out diagnostics for time series (with discussion), *Journal of the Royal Statistical Society Series B*, Vol. 51, 363-424.
- [2] Chang, I., Tiao, G. C., and Chen, C. (1988). Estimation of time series parameters in the presence of outliers, *Technometrics*, Vol. 30, 193-204.
- [3] Chen, C. and Liu, L. M. (1993). Forecasting times series with outliers, *Journal of Forecasting*, Vol. 12, 13-35.
- [4] Fox, A. J. (1972). Outliers in time series, *Journal of the Royal Statistical Society Series B*, Vol. 32, 337-645.
- [5] Ledolter, J. (1988). The effect of additive outliers on the forecasts from *ARIMA* models, *International Journal of Forecasting*, Vol. 5, 231-240.
- [6] Lee, Jae June and Pyun, Young Sook (1996). The effect of patchy outliers in time series forecasting, *The Korean Journal of Applied Statistics*, Vol. 9, 125-137.
- [7] Martin, R. D. and Yohai, V. J. (1986). Influential functionals for time series (with discussion), *Annals of Statistics*, Vol. 14, 781-818.
- [8] Montgomery, D. C., Johnson, L. A., and Gardiner, J. S. (1990). *Forecasting and time series analysis(2nd ed.)*, McGraw-Hill, New York.
- [9] Sweet, A. L. (1985). Computing the variance of the forecast error for the Holt-Winters seasonal models, *Journal of Forecasting*, Vol. 4, 235-243.
- [10] Tsay, R. S. (1988). Outliers, level shifts, and variance changes in time series, *Journal of Forecasting*, Vol. 7, 1-20.

[ 1998년 10월 접수, 1999년 4월 최종수정 ]

## 부록

- [1] Case 1과 Case 2의 유도(sketch of derivations)

표기를 단순하게 하기 위해,  $t$ 시점에서  $\hat{W}(t)_{PO}$  (단,  $W = \beta_0, \beta_1, S_t$ )를  $\hat{W}(t)_{NO}$ 와 남은 항들의 합으로 표현하고, 남은 항들을 그 시점에서  $W$ 가  $PO$ 로 부터 받는 영향이라고 생각하여  $-\Delta W(t)$  라고 표기한다. 즉

$$\hat{W}(t)_{PO} = \hat{W}(t)_{NO} - \Delta W(t), \quad W = \beta_0, \beta_1, S_t \quad (A.1)$$

$\hat{\beta}_0(t), \hat{\beta}_1(t), \hat{S}(t)$ 는 주기와 관련되어 계산되는데, 시점  $t$ 가  $PO$  발생시작시점  $T$ 로부터 한주기( $L$ )이내이면  $\hat{S}(t-L)_{PO} = \hat{S}(t-L)_{NO}$ 가 만족되고, 그 이후에는(즉,  $t \geq T+L$ ) 그 관계가 성립되지 않으므로  $\Delta W(t)$ 에  $\Delta S(t-L)$  항이 나타나게 된다.

Case 1. ( $m < L$ 의 경우)

i)  $PO$  발생전 ( $t \leq T-1$ )의 경우 이상치가 없기 때문에 두 자료집단으로 계산되는 각 요인의 추정값들은 같고, 따라서 두 추정치의 차는 모두 0이 된다.

$$\Delta\beta_0 = \Delta\beta_1(t) = \Delta S(t) = 0$$

ii)  $PO$  발생중 ( $T \leq t \leq T+m-1$ )

시점  $T$ 에서는  $Z_T = Y_T - \omega_0$  이고,  $\hat{W}(T-1)_{PO} = \hat{W}(T-1)_{NO}$  이므로

$$\begin{aligned} \hat{\beta}_0(T)_{PO} &= \gamma_1\{Z_T - \hat{S}(t-L)_{PO}\} + (1-\gamma_1)\{\hat{\beta}_0(T-1)_{PO} + \hat{\beta}_1(T-1)_{PO}\} \\ &= \gamma_1\{Y_T - \omega_0 - \hat{S}(t-L)_{NO}\} + (1-\gamma_1)\{\hat{\beta}_0(T-1)_{NO} + \hat{\beta}_1(T-1)_{NO}\} \\ &= \hat{\beta}_0(T)_{NO} - \gamma_1\omega_0 \\ \hat{\beta}_1(T)_{PO} &= \gamma_2\{\hat{\beta}_0(T)_{PO} - \hat{\beta}_0(T-1)_{PO}\} + (1-\gamma_2)\hat{\beta}_1(T-1)_{PO} \\ &= \gamma_2\{\hat{\beta}_0(T)_{NO} - \gamma_1\omega_0 - \hat{\beta}_0(T-1)_{NO}\} + (1-\gamma_2)\hat{\beta}_1(T-1)_{NO} \\ &= \hat{\beta}_1(T)_{NO} - \gamma_2\gamma_1\omega_0 \\ \hat{S}(T)_{PO} &= \gamma_3\{Z_T - \hat{\beta}_0(T)_{PO}\} + (1-\gamma_3)\hat{S}(T-L)_{PO} \\ &= \gamma_3\{Y_T - \omega_0 - (\hat{\beta}_0(T)_{NO} - \gamma_1\omega_0)\} + (1-\gamma_3)\hat{S}(T-L)_{NO} \\ &= \hat{S}(T)_{NO} - \gamma_3(\omega_0 - \gamma_1\omega_0) \end{aligned} \tag{A.2}$$

가 된다. 따라서,  $T$  시점에서  $\beta_0, \beta_1, S_t$ 의 추정치가 받은  $PO$ 의 영향은  $\Delta\beta_0(T) = \gamma_1\omega_0$ ,  $\Delta\beta_1(T) = \gamma_2\Delta\beta_0(T)$ ,  $\Delta S_T = \gamma_3(\omega_0 - \Delta\beta_0(T))$ 가 된다. 같은 방법으로,  $T+1$  시점에서의  $\hat{W}(T+1)_{PO}$ 는 (A.2)의 결과에 의해

$$\begin{aligned} \hat{\beta}_0(T+1)_{PO} &= \gamma_1\{Y_{T+1} - \omega_1 - \hat{S}(T+1-L)_{PO}\} \\ &\quad + (1-\gamma_1)\{\hat{\beta}_0(T)_{PO} + \hat{\beta}_1(T)_{PO}\} \\ &= \hat{\beta}_0(T+1)_{NO} - \gamma_1\omega_1 - (1-\gamma_1)\{\Delta\beta_0(T) + \Delta\beta_1(T)\} \\ \hat{\beta}_1(T+1)_{PO} &= \gamma_2\{\hat{\beta}_0(T+1)_{PO} - \hat{\beta}_0(T)_{PO}\} + (1-\gamma_2)\hat{\beta}_1(T)_{PO} \\ &= \hat{\beta}_1(T+1)_{NO} - \gamma_2\{\Delta\beta_0(T+1) - \Delta\beta_0(T)\} - (1-\gamma_2)\Delta\beta_1(T) \\ \hat{S}(T+1)_{PO} &= \gamma_3\{Y_{T+1} - \omega_1 - \hat{\beta}_0(T+1)_{PO}\} + (1-\gamma_3)\hat{S}(T+1-L)_{PO} \\ &= \hat{S}(T+1)_{NO} - \gamma_3(\omega_1 - \Delta\beta_0(T+1)) \end{aligned}$$

가 되므로  $\Delta W(T+1)$ 은  $\Delta\beta_0(T+1) = \gamma_1\omega_1 + (1-\gamma_1)\{\Delta\beta_0(T) + \Delta\beta_1(T)\}$ ,  $\Delta\beta_1(T+1) = \gamma_2\{\Delta\beta_0(T+1) - \Delta\beta_0(T)\} + (1-\gamma_2)\Delta\beta_1(T)$ ,  $\Delta S(T+1) = \gamma_3\{\omega_1 - \Delta\beta_0(T+1)\}$ 가 된다.

$T \leq t \leq T + m - 1$ 에서는  $\Delta W(t)$ 는  $T+1$ 시점에서와 같은 방법으로 반복(iteration)해서 계산될 수 있고, 따라서 본문의 Case 1-ii)의 결과를 얻게 된다.

iii)  $PO$  종료후이고  $PO$ 시작 한 주기 이내( $T + m \leq t \leq T + L - 1$ )의 시점  $t$ 에서는  $Y_t$ 에 이상영향( $\omega$ )이 포함되어 있지 않고, 또한  $\hat{S}(t - L)_{PO} = \hat{S}(t - L)_{NO}$ 가 성립하므로  $\Delta W(t)$ 는 본문의 Case 1-iii)와 같이 유도된다.

iv)  $t \geq T + L$ 의 시점에서는  $Y_t$ 에 이상영향( $\omega$ )이 포함되어 있지 않고  $PO$  발생시점에서 한 주기가 지났으므로,  $\hat{S}(t - L)_{PO} = \hat{S}(t - L)_{NO} - \Delta S(t - L)$ 가 만족된다. 따라서, 같은 방법으로  $\Delta W(t)$ 를 유도하면 Case 1-iv)의 결과를 얻게 된다.

Case 2. ( $m \geq L$ 의 경우)

$m \geq L$ 인 경우도  $\hat{W}(t)_{PO} = \hat{W}(t)_{NO} - \Delta \hat{W}(t)$ 로 표현할 수 있으며,  $PO$ 의 영향은 Case 1의 경우와 같은 방법으로  $\Delta W(t)$ 를 유도하여 표현할 수 있다.

[2] (2.7)식에서  $\alpha(l)_i$ 의 정의

$l = 1$ 에서는  $\alpha(l)_i$ 는  $\alpha_i$ 로 표기하며 다음과 같이 계산된다.(Sweet(1985))

$$\begin{aligned} \alpha_0 &= 1 \\ \alpha_1 + (\gamma_1\gamma_2 - (1 - \gamma_1))\alpha_0 &= -1 \\ \alpha_j + (\gamma_1\gamma_2 - (1 - \gamma_1))\alpha_{j-1} + \gamma_1\gamma_2 \sum_{i=0}^{j-2} \alpha_i &= -\gamma_3(1 - \gamma_1)\delta_{j,L} \quad 2 \leq j \leq L \\ \alpha_j + (\gamma_1\gamma_2 - (1 - \gamma_1))\alpha_{j-L} + \gamma_1\gamma_2 \sum_{i=j-L}^{j-2} \alpha_i + (\gamma_3(1 - \gamma_1) - 1)\alpha_{j-L} + (1 - \gamma_1)(1 - \gamma_3)\alpha_{j-L-i} & \\ &= \delta_{j,L+1} \quad L + 1 \leq j \end{aligned}$$

단,  $\delta_{i,j}$ 는 Kronecker delta이다.

$l \geq 2$ 에서는  $\alpha(l)_j$ 는  $\alpha(l)_j = \psi(l)_j + \phi(l)_j$ 가 되는데  $\psi(l)_j$ 는

$$\psi(l)_j = \begin{cases} 1 & j = 0 \\ \alpha_j + \gamma_1 \sum_{i=1}^{\min(i,l-1)} (1 + i\gamma_2)\alpha_{j-i} & j > 0 \end{cases}$$

와 같고  $\phi(l)_j$ 는 다음과 같다.

i)  $1 \leq l \leq L$

$$\phi(l)_j = 0, \quad j \leq 0$$

ii)  $L + 1 \leq l \leq 2L$

$$\phi(l)_j = \begin{cases} 0 & 0 \leq j \leq L - 1 \\ (1 - \gamma_1)\gamma_3 \alpha_{j-L} & j \leq L \end{cases}$$

$$\text{iii) } kL + 1 \leq l \leq (k + 1)L$$

$$k \geq 1$$

$$\phi(l)_j = \begin{cases} 0 & 0 \leq j \leq L - 1 \\ (1 - \gamma_1)\gamma_3 \sum_{i=1}^c \alpha_{j-iL} & L \leq j \leq kL - 1, j = cL + \gamma, 0 \leq \gamma \leq L - 1, c \geq 1 \\ (1 - \gamma_1)\gamma_3 \sum_{i=1}^k \alpha_{j-iL} & kL \leq j \end{cases}$$

## Influence of Patchy Outliers on the Forecast of Winters Seasonal Model

Young Sook Pyun<sup>1)</sup> Jae June Lee<sup>2)</sup>

### ABSTRACT

Unusual observations are not rare in time series data. They, especially outliers in sequence, can influence on the analysis of the time series, because of the dependent structure among observations. This paper focuses on the derivation of the impact of patchy outliers(*PO*) on the analysis of time series in Winters seasonal exponential smoothing models. In addition to that, we derived the mean square of the  $l$  step-ahead forecast error to measure the impact of *PO* on forecasts, and analyzed a real data set to roughly evaluate the impact of *PO* on the forecasting procedure.

---

1) M.S., Department of Statistics, Inha University, 253 Younghyun-dong, Nam-ku, Inchen 402-751, Korea

2) Professor, Department of Statistics, Inha University, 253 Younghyun-dong, Nam-ku, Inchen 402-751, Korea