시계열분석

제1장 시계열자료

시계열(time series)자료는 시간에 따라 관측된 자료로 시간에 영향을 받는다. 예를 들면 오늘의 주가가 한달전, 일주일전의 주가보다는 어제, 그제의 주가에 더 많은 영향을 받는것 처럼 가까운 관측시점일수록 관측된 자료들간에 상관관계가 커진다는 의미이다.

관측시점을 나타내는 시간의 간격은 일, 월, 분기, 년 등으로 구분한다. 예들 들면 종합주가지수, 환율 등은 매일, 경기종합지수, 산업생산지수, 실업률 등은 매월, 국내총생산(GDP) 등은 매분기로 작성되는 자료이다. 시계열 분석에서는 관측시점과 관측시점들 사이의 간격인 시차(time lag)가 중요한 역할을 하므로 일반적으로 시간 t를 이용하여 다음과 같이 표현한다.

$$Z_t \ : t = 1,\ 2,\ 3 \quad \text{또는} \quad Z_1,\ Z_2,\ Z_3$$

1. 시계열의 형태

시계열 분석시에 가장 선행되어야 할 일은 시계열 그림(time series plot)을 그려본다. 이것은 시간의 경과에 따라 시계열의 자료가 변하는 것을 그린 것으로 시간 t 를 가로축으로 하고 시계열의 관측값를 세로축으로 하여 그린다. 그림을 그리는 이유는 시계열이 가지는 특징을 쉽게 파악할 수 있어 해당 자료의 적합한 분석방법 선택에 도움을 주기 때문이다.

시계열 분석함에 있어 시계열 자료가 사회적 관습이나 환경 변화 등의 다양한 변동요인에 영향을 받는다고 보아 시계열을 여러가지 변동요인으로 각각 분해한 후 원시계열(original time series)을 분석하는 분해법(decomposition method)을 많이 사용되었다.

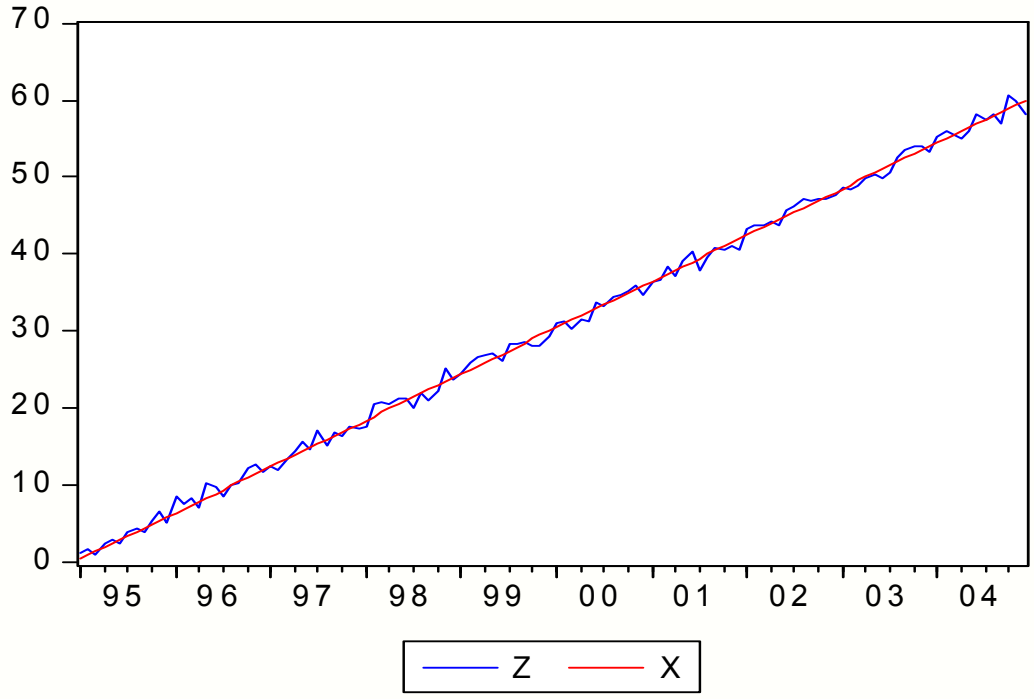
일반적으로 시계열에서 나타나는 변동에는 우연적으로 발생하는 불규칙변동(irregular variation)과 체계적(systematic)변동을 들 수 있다. 체계적변동 요인에는 장기간에 걸쳐 어떤 추세를 보이는 추세변동 (trend variation), 추세선을 따라 주기적으로 오르고 내림을 반복하는 순환변동 (cyclical variation), 그리고 계절적 요인이 작용하여 1년 주기로 나타나는 계절변동 (seasonal variation)이 있다.

(1) 추세변동

추세변동이란 시계열자료가 갖는 장기적인 변화추세이다. 추세란 장기간에 걸쳐 지속적으로 증가 또는 감소하거나 혹은 일정한 상태를 유지하려는 성향을 의미한다. 그러므로 시계열 자료에서 짧은 기간 동안에는 추세 변동을 찾기란 어렵다. 추세변동은 짧은 기간 동안 급격히 변하는 것이 아니라 장기적인 추세 경향이 나타나는 것으로 직선이나 부드러운 곡선의 연장선으로 표시한다. 이러한 추세는 반드시 직선적일 필요은 없으며, 이차곡선의 형태 또는 S자형태의 추세를 가질 수도 있다. [그림 1]은 선형추세를 갖는 시계열의 예이다.

프로그램 예제 1: 선형추세를 갖는 시계열 산출과 그래프

```
wfcreate(wf="monthly") m  1995:01 2004:12

rndseed 1234
smpl @first @last
series tt=@obsnum
series x=0.5*tt
series z=0.5*tt+nrnd

graph zg.line(a) z x
```



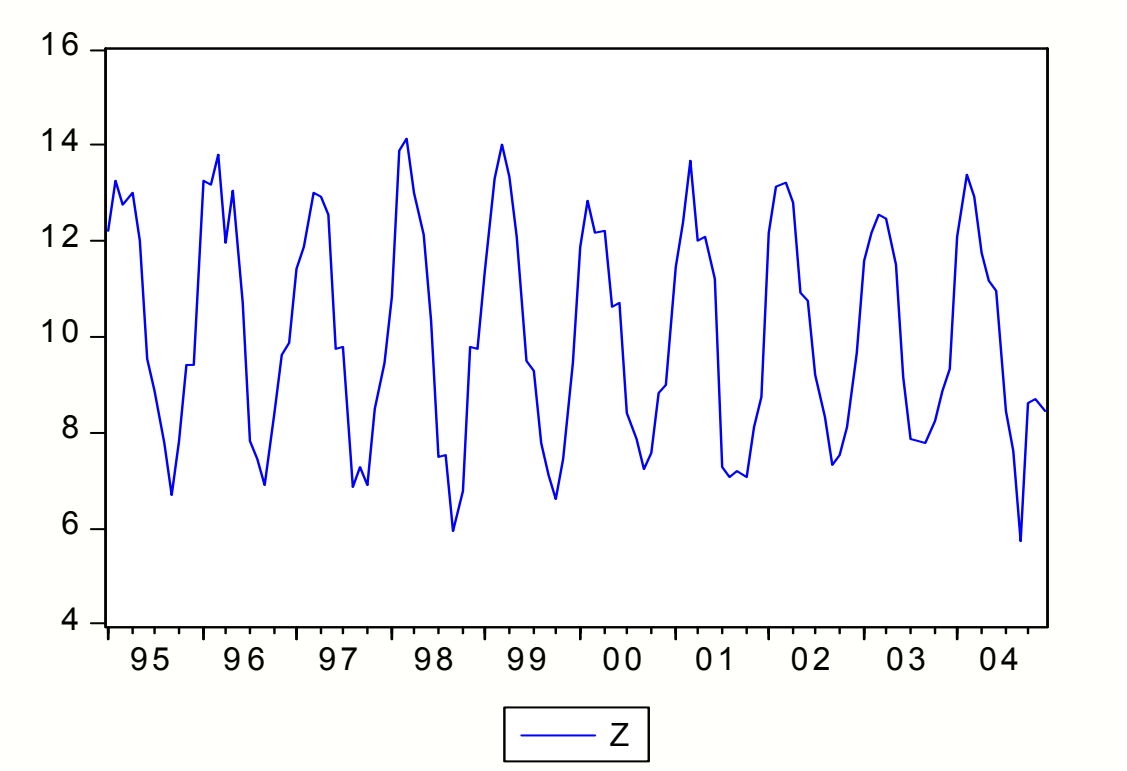
[그림 1] 추세성분을 갖는 시계열

(2) 계절변동

시계열자료에서 보통 계절적 영향과 사회적관습에 따라 1년 주기로 발생하는 변동요인을 계절변동이라 하고, 보통 계절에 따라 순환하며 변하는 특성을 지닌다. 그런데 계절변동이 순환변동과 다른 점은 순환주기가 짧다는 점이다. [그림 2]는 계절변동요인만 있는 시계열의 예이다.

프로그램 예제 2: 계절변동요인을 갖는 시계열 산출과 그래프

```
wfcreate(wf="monthly") m  1995:01 2004:12

rndseed 1234
smpl @first @last
series tt=@obsnum
genr z=10+3*@sin((2*3.14*tt)/12)+0.8*nrnd

graph g2.line(a) z
```

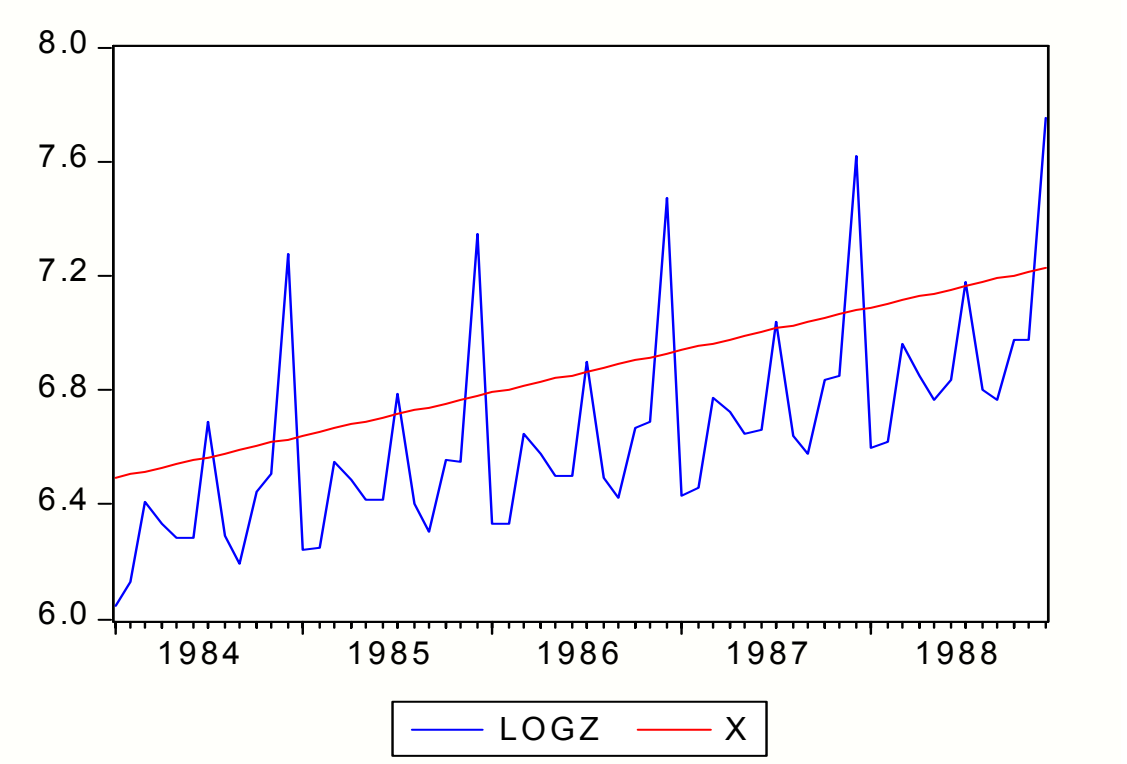


[그림 2] 계절변동요인을 갖는 시계열

그러나 대부분의 경제관련 시계열들은 [그림 3]과 같이 추세와 계절요인을 동시에 포함하고 있다. 이는 경제가 성장함에 따라 백화점판매액, 해외여행자수, 청량음료 등과 같이 계절상품 판매량 자료들이 시간의 변화에 따라 증가하기 때문이다.

프로그램 예제 3: 추세와 계절변동요인을 갖는 시계열 그래프

```
wfcreate(wf="monthly") m  1984:01 1988:12
read(t=dat,norect)  e:\times\depart1.dat z
genr logz=log(z)
genr x=-289.7361+0.000409*@date
graph g3.line(a) logz x
```

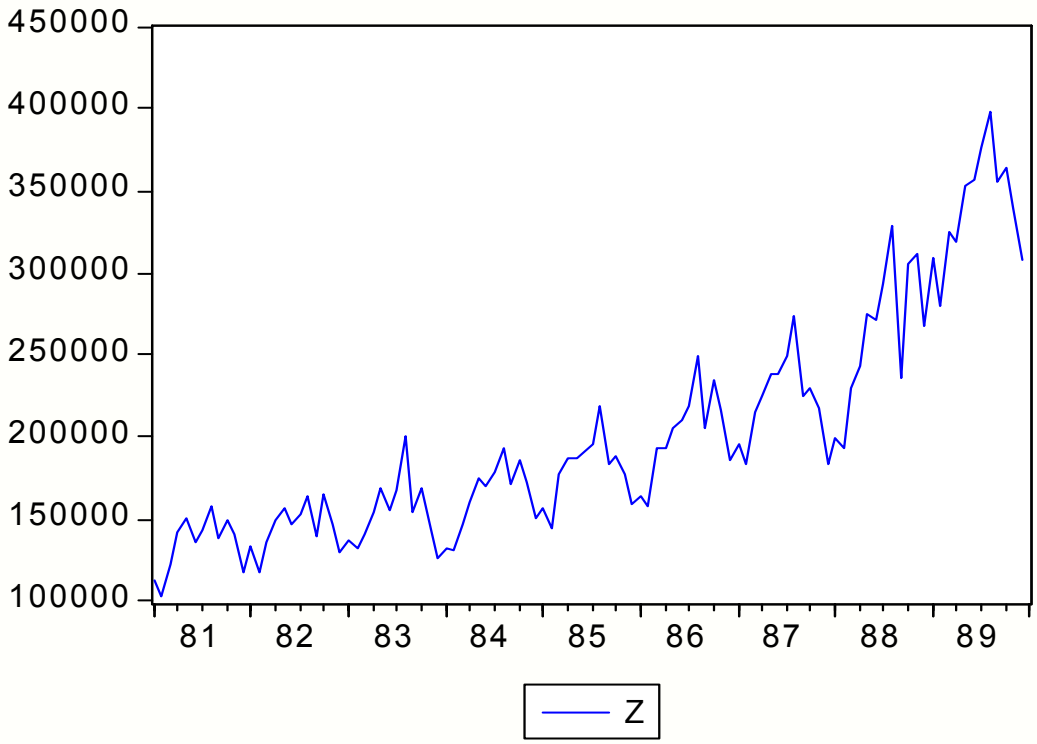


[그림 3] 추세와 계절변동요인을 갖는 시계열

[그림 3]에서는 시간의 경과에 따라 자료의 변동폭에 변화가 없으나 [그림 4]는 시간의 증가에 따라 변동 폭이 커지는 시계열의 예이다. 경제시계열자료들이 이와같은 추세를 보여준다.

프로그램 예제 4: 시간의 변화에 따라 변동폭이 커지는 시계열 그래프

```
wfcreate(wf="monthly") m  1981:01 1990:12
read(t=dat,norect)  e:\times\koreapass.dat z
graph g4.line(a) z
```



[그림 4] 시간의 변화에 따라 변동폭이 커지는 시계열 그래프 프로그램

(3) 순환변동

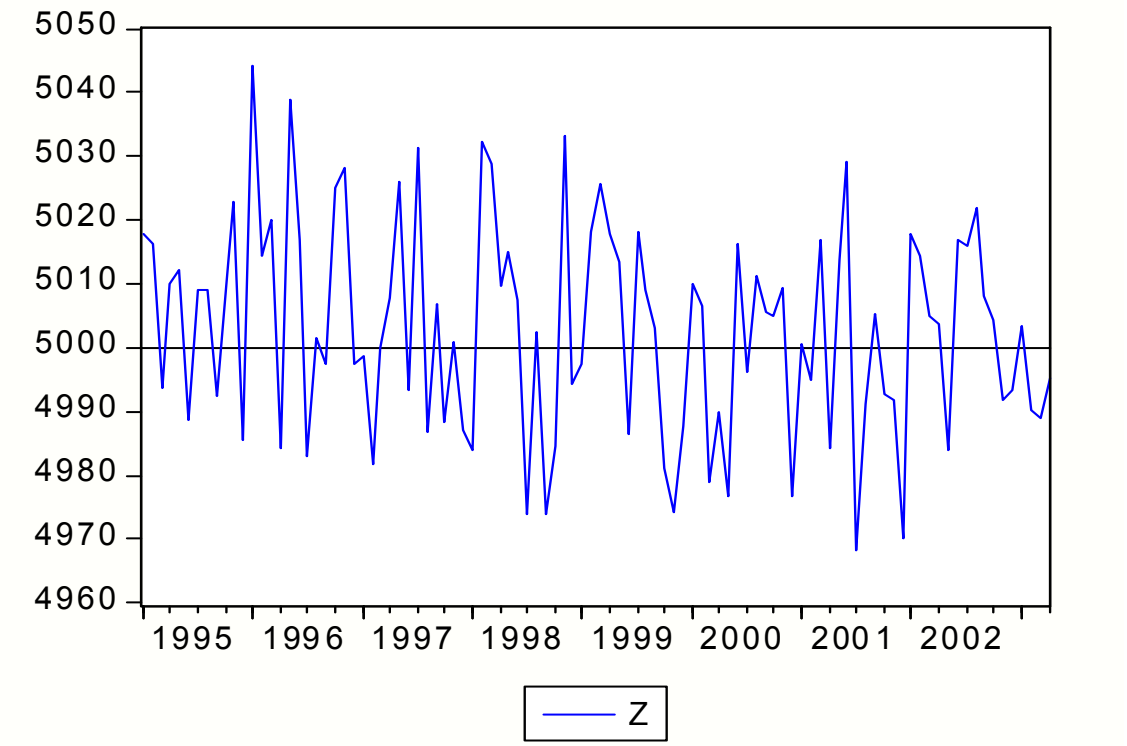
추세변동은 장기적으로 나타나는 추세경향이지만 순환변동은 대체로 2 ~ 3 년 정도의 일정한 기간을 주기로 순환적으로 나타난다. 시간의 흐름에 따라 상하로 반복되는 변동으로 추세선을 따라 변화하는 것이 순환변동이다. 경기변동 곡선은 불황과 경기회복, 호황과 불경기로 인하여 수년을 주기로 나타나고 있는데 순환변동을 나타내는 좋은 예이다.

(4) 불규칙변동

시계열자료에서 어떤 규칙성이 없이 예측 불가능하게 우연적으로 발생하는 변동을 불규칙변동이라 한다. 그러므로 시계열자료에서 추세변동 순환변동 계절변동 요인을 조정한 후에 나타나는 변동이 불규칙변동이다. 불규칙변동은 천재지변이나 급격한 환경 변화로 발생되는 것이 아니라 [그림 5]와 같이 우연적으로 발생하는 변동이다.

프로그램 예제 5: 불규칙변동요인을 갖는 시계열 산출 및 그래프

```
wfcreate(wf="monthly") m  1995:01 2004:12
rndseed 1234
smpl @first @last-20
series z=5000+20*nrnd
graph g5.line(a) z
        g5.draw(line,left) 5000
```

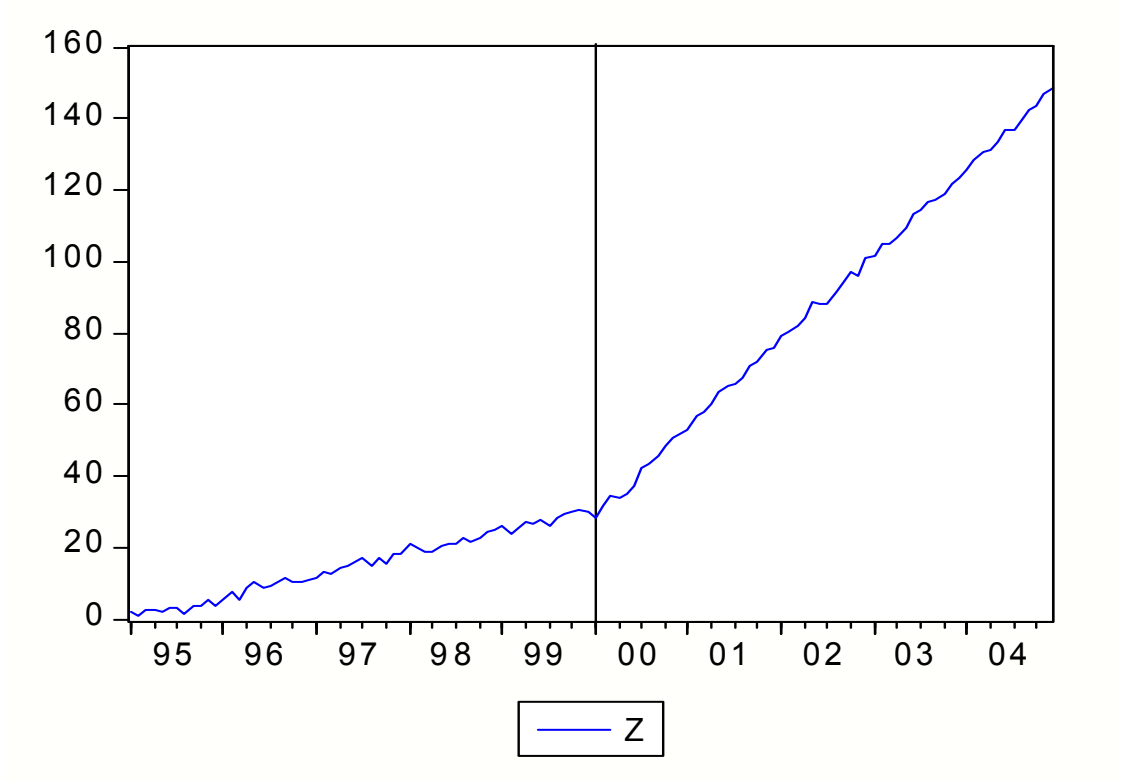


[그림 5] 불규칙변동요인을 갖는 시계열

위와 같이 설명한 형태의 시계열들에서는 그 성질들이 뚜렷하게 나타나고 있으나 실제 자료의 경우에는 그 형태를 구분할 수 없는 경우가 대부분으로 시계열 구성요인(추세, 순환, 계절, 불규칙)들을 정확히 구분한다는 것은 쉽지 않다. 또한 [그림 6]과 같이 동일한 시계열 내에서도 시점에 따라 두 개의 추세(trend)을 보이는 경우도 있다.

프로그램 예제 6: 추세선이 두 개인 시계열 그래프

```
wfcreate(wf="monthly") m  1995:01 2004:12
series tt=@obsnum
rndseed  4321
smpl @first  @last-60
series  x=0.5*tt
smpl @first+60 @last
series  x=2*(tt-46)
smpl @first @last
series z=x+nrnd
graph g5.line z
        g5.draw(shade, bottom)  2000:01
```



[그림 6] 추세선이 두 개인 시계열

2. 시계열 모형 및 분석방법

(1) 분석 목적 및 방법

시계열분석 목적은 미래에 대한 예측(forecast)과 시스템 또는 확률과정을 이해하고 제어(control)하는 두가지로 나눌 수 있다.

예측의 목적을 위한 시계열 분석법에는 추세분석(trend analysis), 평활법(smoothing method), 분해법(decomposition method),자기회귀누적이동평균(ARIMA)모형에 의한 분석법 등이 있으며, 시스템의 이해와 제어의 목적을 위한 시계열분석법에는 스펙트럼 분석(spectral analysis), 개입분석(intervention analysis), 전이함수모형(transfer function model)분석 등이 있다.

또한 시계열의 특성은 시간에 따라 관측되는 관계로 적용분야에 따라 다음의 두가지 접근법이 있다.

(가) 진동수영역(frequency domain)에서의 분석법

스펙트럼 밀도함수(spectral density)를 추정하여 시계열이 주로 어떠한 주기(cycle)를 가지고 움직이는가 하는 시계열의 특성을 이해하는데 중점을 두며 퓨리에분석(Fourier analysis)에 기초한다. 주로 공학분야에서 사용된다.

(나)시간영역(time domain)에서의 분석법

자기상관함수(autocorrelation function) 등을 이용하여 관측값들이 시간에 따른 상관 정도를 파악한 후 이들 관계를 가장 잘 설명해주는 모형을 찾는 방법으로서, 경제 또는 경영학 분야에서 시계열자료의 분석은 주로 이 방법에 의한다.

(2) 분석 모형의 구분

(가) 결정모형(determinstic model)

확률적 요소를 전혀 갖지 않는 모형으로 미래의 진행과정이 해당시점에서 여러 변수들의 값에 의해 완전히 결정되는 모형이다. 공학이나 물리적인 현상을 설명하는 경우에 많이 사용한다.

$$y=f(x_1,x_2,\ldots\ldots,x_p;\beta_1,\beta_2,\ldots\ldots,\beta_m):$$

$f,\beta_1,\ldots\ldots,\beta_m$은 알고있음

(나) 확률모형(stochastic model)

확률오차를 수반하는 모형으로 오차들이 서로 독립이고 동일한 분포를 따르는 확률변수(random variable)이다. 함수식의 형태가 사전에 알려져 있지 않으므로 오차의 확률분포와 자료를 직접 이용하여 모형을 추정한다.

$$y=f(x_1,x_2,\ldots\ldots,x_p;\beta_1,\beta_2,\ldots\ldots,\beta_m)+\text{오차}:$$

$f,\beta_1,\ldots\ldots,\beta_m$은 과거자료들에 의해 결정

(4) 예측모형의 종류

(가) 추세분석모형

추세모형을 이용한 예측법은 다항회귀모형과 유사한 모형을 가정하고 모수의 추정을 통해 예측값을 구하고 있다. 그러나 회귀분석 모형과의 차이점이라면 설명변수로 주로 시간의 함수를 사용한다는 점이다. 이 때 사용되는 함수로는 시간의 일차, 이차 등의 다항식이나 삼각함수 또는 지시함수 등이 사용된다. 주로 시간의 선형함수(linear function)가 사용되나 경우에 따라서는 시간의 비선형함수(nonlinear function)가 사용되기도 한다.

(나) 평활법

현재로 부터 가장 최근에 관측된 자료에는 큰 가중값을 주고, 과거로 갈수록 작은 가중값을 주는 일종의 가중평균을 이용한 예측방법이다. 평활법은 이론적으로 미흡한 점이 많으나, 직관적으로 이해하기 쉽고, 사용이 편리하며 많은 자료의 예측시에 편리한 점 때문에 1970년대 이전까지 많이 이용했었다.

(다) ARIMA모형

ARIMA모형은 현 시점의 관측값을 과거의 관측값들과 백색잡음(white noise)이라고 부르는 오차들의 형태로 표현하는 모형으로서, 1970년대에 들어와서 Box와 Jenkins가 ARIMA모형을 적합시키는 3단계 절차를 제안한 이후로는 박스-젠킨스모형 (Box-Jenkins model)이라는 이름으로 가장 많이 사용되고 있다.

(라) 분해법

가장 오래된 시계열 분석법으로 시계열의 변동을 주요한 요인변동들로 분해하여 각 요인변동들을 추정한 후 이를 이용하여 예측값을 구한다. 이 방법들은 X-11,X-12-ARIMA 방법과 같은 계절조정(seasonal adjustment)의 목적에 주로 이용되고 있다.

(5) 모형적합의 3단계 및 예측시스템

일반적으로 시계열에 적합모형은 3단계를 거쳐서 선택한다.

(i) 모형의 식별(model identification)

(ii) 모형의 추정(model estimation)

(iii) 모형의 진단(model diagnostic checking)

그런데 예측을 목적으로 하는 경우에는 모형수립단계(model building stage)를 거쳐 예측단계(forecasting stage)에서 새로운 자료가 관측될 때마다 예측값을 비교해보아 선택된 모형자체에 변화가 있는지 여부를 통해 안정성을 조사한 후 최종 모형으로 선택해야 한다.

(6) 예측방법의 선택기준

앞에서 언급한 모형수립과 예측의 2단계를 거쳐 예측값을 얻을 수 있다. 그러나 자료의 형태, 예측의 목적 혹은 예측의 성격에 따라 좋은 예측력을 갖는 예측방법들은 달라질 수 있다 . 따라서 바람직한 예측 모형이 갖추어야 할 기준들을 다음과 같이 제시하고 있다 .

① 어느정도의 정확성이 요구되는가? 예측값을 얻기 위해 소요되는 시간 및 노력과 그 예측의 정확성과의 절충(trade off)되어야 한다.

② 예측하고자 하는 기간의 길이(forecast horizon)에 따른 모형의 선택, 즉 단기예측(short-term forecast)과 장기예측(long-term forecast) 중에서 선택한다.

③ 어느 정도로 복잡한가? : 절약성의 원리 (principle of parsimony) 에 따라 간단하면서도 쉽게 이용 가능한 모형을 선택하는 것이 바람직하다. 따라서 설명력이 비슷한 경우에는 간단한 모형, 즉 모수의 개수가 적고 해석이 용이한 모형을 선택하는 것이 좋다.

④ 이용가능한 자료의 종류에 따라 적용 가능한 모형의 선택 : 결측값이 있는 경우에는 이에 적합한 모형을 선택하거나 결측값을 추정한 후에 모형을 적합시킨다. 자료의 개수에 따라, 적용할 수는 있으나 그 결과를 신뢰하기 어려운 모형도 있다. 일반적으로 ARIMA 모형은 50 개 이상의 자료가 있어야만 그 결과를 신뢰할 수 있다고 알려져 있다.

이상의 여러 가지 기준 이외에도 예측력이 좋은지 나쁜지를 판정하기 위하여 다음과 같은 측도들을 사용하기도 한다. 특히 정확한 예측에 목적을 두고 있는 경우에는 어떠한 측도를 사용하느냐에 따라 선택되는 방법이 달라질 수도 있겠으나, 일반적으로 현재까지 관측된 자료, $\{Z_t, t=1,2,3,\ldots.n\}$ 가 있고 시점 t 에서의 1-시차후의 예측값을 $\hat{Z}_t(1)$이라고 한다면 1-시차후의 예측오차(one-step-ahead prediction error) $\hat{e}_t(1)=Z_{t+1}-\hat{Z}_t(1)$을 이용하여 다음과 같은 측도들을 정의하고 있다.

① 평균제곱예측오차: MSE(mean square prediction error)

$$MSE=\frac{\sum_{t=1}^{m}\hat{e}_{n-1+t}(1)^2}{m}$$

② 평균절대백분위예측오차: MAPE(mean absolute percentage prediction error)

$$MAPE=\frac{100}{m}\sum_{t=1}^{m}\left|\frac{\hat{e}_{n-a+t}(1)^2}{Z_{n+t}}\right|$$

③ 평균절대예측오차 : MAE(mean absolute prediction error)

$$MAE=\frac{\sum_{t=1}^{m}|\hat{e}_{n-a+t}(1)^2|}{m}$$

제2장 추세분석모형

전통적으로 시계열자료를 분석하기 위해 많이 사용되어 온 방법 중의 하나로서 관측값 Z_t를 시간의 함수로서 표현하는 방법이다. 즉,

$$Z_t = \beta_0 + \beta_1 t + \beta_2 t^2 + \cdots + \beta_p t^p + \varepsilon_t$$ 와 같이

관측값 Z_t 를 시간 t의 다항식으로 나타낸 모형으로 선형다항식추세모형 또는 다항식추세모형이라 한다.

① 상수평균모형(constant mean model) : p=0인 다항추세모형
② 선형추세모형 (linear trend model) : p=1인 다항추세모형
③ 2차 추세모형(quadratic trend model) : p=2인 다항추세모형
④ 계절추세모형(seasonal trend model) : 설명변수로서 삼각함수 혹은 지시함수 사용
⑤ 비선형추세모형(nonlinear trend model) : 비선형함수 사용, $Z_t = \exp(\beta_0 + \beta_1 X_t)_t$
⑥ 자기회귀오차모형(autoregressive error model) : AR 오차모형 사용

다항추세모형에 의해 추세분석함에 있어 중회귀모형에서 독립변수t 의 함수를 사용하는 점이 특징이다. 따라서 다항추세 모형의 분석은 회귀분석의 절차를 똑같이 사용하면 된다. 즉 모형식별단계에서 주어진 자료가 모형을 적합시키기 위해 필요한 등분산성 등의 가정을 만족시키지 못하는 경우에는 로그변환 emdd과 같이 분산안정화변환 등을 통해 자료를 변환한다. 변환된 자료의 시계열그림을 그려 변환 후의 자료에 적절한 모형을 잠정적으로 선택한다. 모형의 추정단계에서는 잠정선택된 모형에 포함된 모수를 주어진 자료를 이용하여 추정한다. 모형의 진단[1] 단계에서는 잔차분석 등을 통해 추정된 모형이 주어진 자료에 잘 적합한지를 진단한다. 최종적으로 확정된 회귀모형은 예측에 사용한다.

(1) 잔차분석

잔차분석(residual analysis)은 잔차들을 분석하여 잠정적으로 설정된 회귀모형이 과연 자료를 잘 설명하는 지를 판단하는 분석단계이다. 잔차는 오차의 추정값으로 생각할 수 있으므로 회귀모형 설정 당시의 오차항에 부가된 가정에 잔차가 잘 부합되는 지를 검토한다. 오차항의 가정에 위반되어 회귀모형에 수용되지 못하고 잔차에 남겨둔 정보가 있다면 이를 밝혀 잠정모형을 갱신해 나간다. 반대로 오차항의 가정들에 위배되지 않는다면 잠정모형을 최종 예측모형으로 사용할 수 있다.

(가) 분산안정화 변환

★ 참고 : 통계소프트웨어를 이용한 통계적 검정에서 대부분의 통계 소프트웨어는 유의확률(significant probability) 혹은 p-값(p-value)을 출력해 준다. 예를들어 p-값=P(T>|t|)으로 계산된다. 일반적으로 컴퓨터 output에서 p-value에 의한 유의수준α의 검정 규칙은

(i) p-값 < α 이면 Ho를 기각한다.
(ii) p-값 ≥ α 이면 Ho를 기각하지 못한다.

시계열자료 Z_t의 분산 $\sigma_t^2 = Var(Z_t)$와 평균수준 $\mu = E(Z_t)$사이에는 보통 함수관계가 있게 된다. 이 함수관계를 f라면 가장 많이 사용되는 함수 f의 형태로는 다음과 같이 두 가지가 있다.

① $\sigma_{z_t}^2 = f(\mu_t) = c\mu_t^2$ ⇒ 자연로그변환 $\{\ln Z_t\}$

② $\sigma_{z_t}^2 = f(\mu_t) = c\mu_t$ ⇒ 제곱근변환 $\{\sqrt{Z_t}\}$

(나) 오차항들의 자기상관관계 검정

① 근사(approximate) 검정법 :

k-시차에서 자기상관관계의 검정. $|\widehat{\rho_k}| > 2n^{-1/2} (k = 1, 2, \ldots)$ 이면 오차항들이 k-시차에서 자기상관관계가 있는 것으로 판단한다. 여기서 $\widehat{\rho_k}$는 잔차들의 k-시차표본자기상관계수로서 오차항들이 자기상관관계를 가지는 지를 검토하기 위해 사용되는 통계량이다.

$$\widehat{\rho_t} = \frac{\sum_{t=k+1}^{n} e_t e_{t-k}}{\sum_{t=1}^{n} e_t^2}, \quad k = 1, 2, \ldots$$

② Durbin-Watson(DW)검정 : 1차 자기상관의 존재여부를 검정하는 것으로 귀무가설 H_0: $\rho_1 = 0$ 를 검정하기 위한 DW 검정통계량은 $D \cong 2(1 - \widehat{\rho_1})$이다. 귀무가설" H_0: ϕ=0"에 대한 DW 검정규칙을 대립가설 H_1 에 따라 정리해보면 다음과 같다.

◆ 대립가설 "H_1: ϕ>0 즉, 양의 자기상관계수"의 검정규칙 :
- ⓐ $D > d_U$이면 유의수준 α에서 H_0를 기각하지 못한다.
- ⓑ $D < d_L$이면 유의수준 α에서 H_0를 기각한다.
- ⓒ $d_L < D < d_U$이면 검정을 유보한다.

◆ 대립가설 "H_1: ϕ<0즉, 음의 자기상관계수"의 검정규칙 :
- ⓐ $4-D > d_U$이면 유의수준 α에서 H_0를 기각하지 못한다.
- ⓑ $4-D < d_L$이면 유의수준 α에서 H_0를 기각한다.
- ⓒ $d_L < 4-D < d_U$이면 검정을 유보한다.

◆ 대립가설 "H_1: ϕ≠0즉, 0이 아닌 자기상관계수"의 검정규칙 :
- ⓐ $D > d_U$ 이고 $4-D > d_U$이면 유의수준 2α에서 H_0를 기각하지 못한다.
- ⓑ $D < d_L$ 이고 $4-D < d_L$이면 유의수준 2α에서 H_0를 기각한다.
- ⓒ ⓐ 혹은 ⓑ가 아니면 검정을 유보한다.

(2) 추세분석모형 사례

(가) 상수평균모형

시계열자료가 [그림5]와 같이 어떤 일정한 수준에 계속 머물면서 불규칙변동요인에 의한 변화만을 보여주는 경우에는 다음의 상수평균모형

$Z_t = \beta_0 + \varepsilon_t$으로 설명 가능하다.

여기서 ε_t는 서로 독립이고 평균 0,분산 σ^2을 갖는 오차항이며 정규분포를 따른다.

(나) 선형추세 또는 2차추세모형

[그림 1]과 같이 직선형의 추세를 갖고 증가하는 경우에는

$$Z_t = \beta_0 + \beta_1 t + \ _t$$ 와 같은

형태의 선형추세모형을 적용할 수 있다. 여기서 $\beta_0 + \beta_1 t$는 선형추세변동요인을 나타낸다. 추세변동요인이 직선이아니고 곡선형태를 따르는 경우에는 2차 추세모형

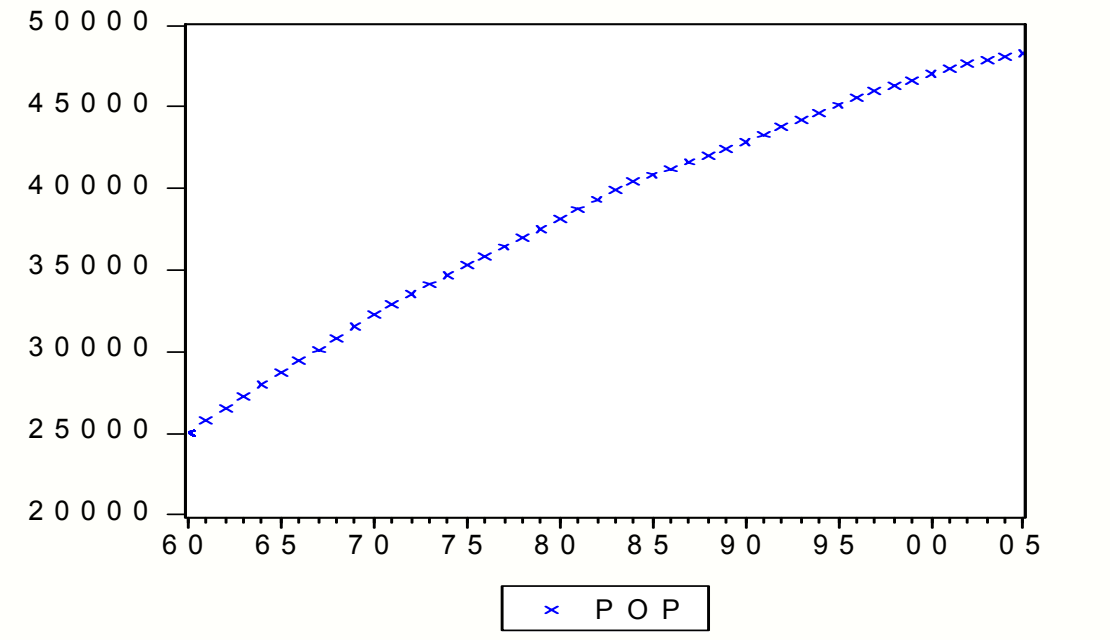
$$Z_t = \beta_0 + \beta_1 t + \beta_2 t^2 + \ _t$$ 이거나 이보다 고차의 다항추세모형

$$Z_t = \beta_0 + \beta_1 t + \beta_2 t^2 + \ldots\ldots\ldots + \beta_p t^p + \ _t$$ 을

이용하여 추세변동요인을 설명할 수 있다. 여기서 ε_t는 서로 독립이고 평균 0,분산 σ^2을 갖는 오차항이며 정규분포를 따른다고 가정한다.

사례 1 : 1960년부터 2005년까지의 우리나라 총인구의 추계자료(단위: 만명)에 대한 추세모형을 설정해보자.

```
wfcreate a 1960 2005
read(t=dat) e:\times\pop2005.dat  pop
genr pop=pop/1000
genr lnpop=log(pop)
genr t=@trend+1
genr t2=T*T

graph popg.scat pop
      popg.setelem(1) symbol(5)
```

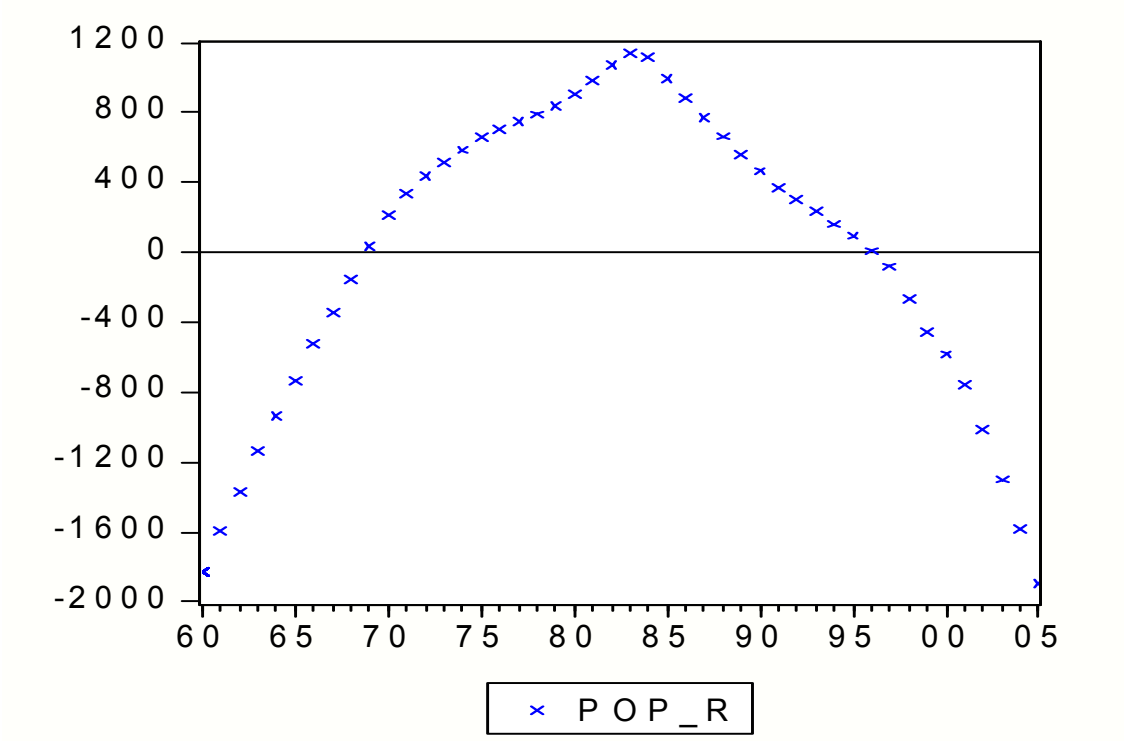


[그림 7] 우리나라 인구 추계자료

시계열 [그림 7]을 보면 증가추세가 있음을 알 수 있다. 증가하는 경향을 볼 때 미세하게 곡선형태를 가지므로 선형추세모형보다는 2차 추세모형을 적합시키는 것이 좋을 것 같으나, 일단 선형추세모형을 적합시켜 보자.

선형추세모형 적합 후의 잔차의 시계열 [그림 8]를 보면 시간의변수의 제곱항이 필요한 것을 알 수 있다.

```
equation eq1.ls pop c t
eq1.makeresids pop_r

graph popg1.scat pop_r
popg1.setelem(1) symbol(5)
popg1.draw(line,left) 0
```

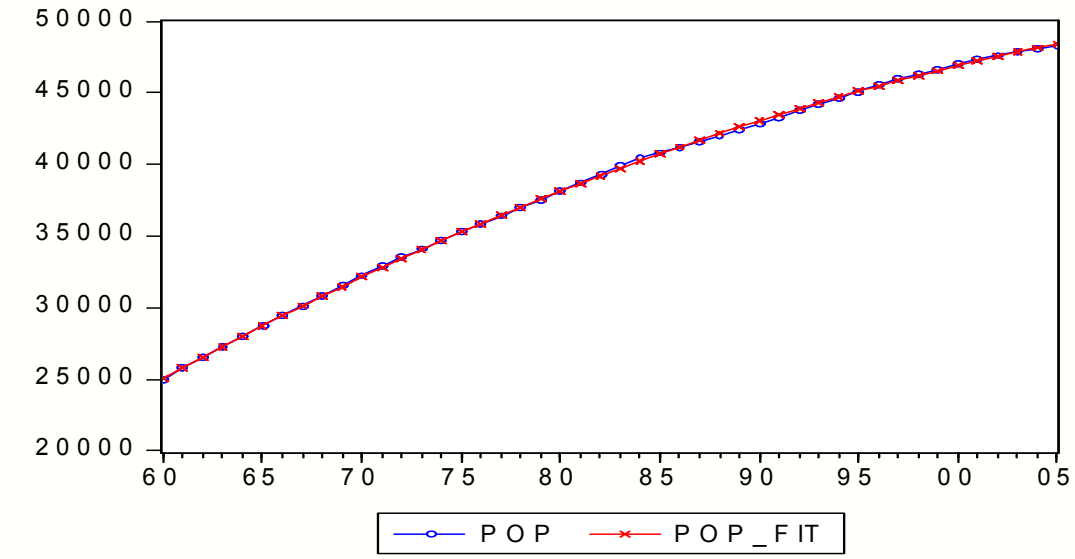


[그림 8] 선형모형 적합후의 잔차 시계열

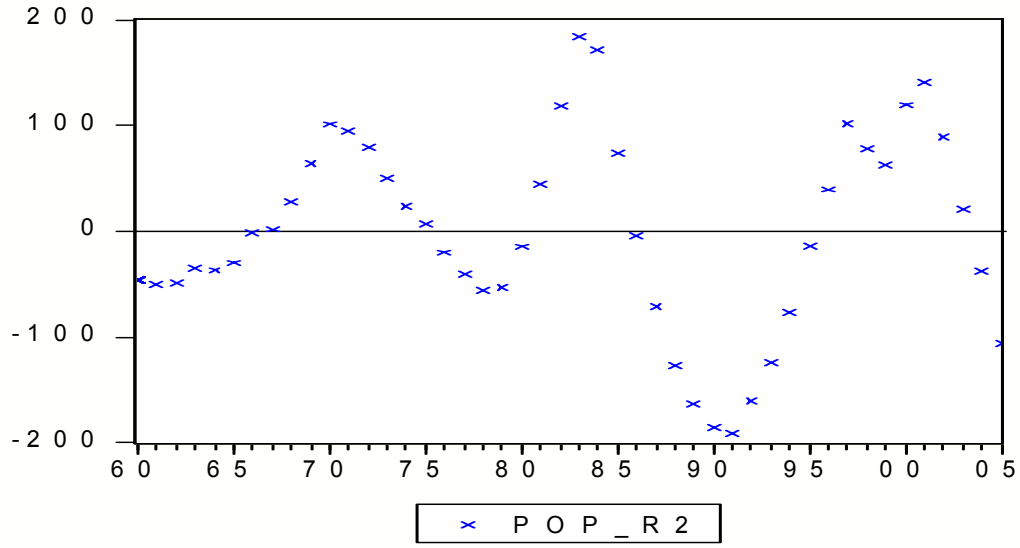
따라서 시간변수의 제곱항을 추가한 2차추세모형의 적합결과 회귀계수들의 유의확률이 0.0001로서 모두 유의하고 결정계수 R^2 도 99.98%로서 설명력도 높으며 관측값과 예측값의 시계열 [그림 9]도 적합이 잘 되었음을 보여주고 있다.

```
equation eq2.ls pop c t t2
eq2.makeresids pop_r2
```

```
eq2.fit pop_fit
graph popg2.line(a)  pop pop_fit
popg2.setelem(1) symbol(1)
popg2.setelem(2)  symbol(5)
graph popg3.scat pop_r2
popg3.setelem(1)  symbol(5)
popg3.draw(line,left) 0
```



[그림 9] 2차 추세모형 적합후의 관측값과 예측값



[그림 10] 2차 추세모형 적합 후의 잔차 시계열

그러나 잔차들의 시계열 [그림 10]을 보면 시간의 흐름에 따라 잔차의 변동 폭이 넓어짐을 qd주고 이다. 오차의 등분산성의 가정이 만족되지 않는 것 같다. 등분산성의 문제를 해결하기 위해서 로그변환을 적용해 보자.

```
equation eq3.ls lnpop c t t2
eq3.makeresids pop_r3
eq3.fit pop_fit

graph popg4.scat pop_r3
popg4.setelem(1)  symbol(5)
popg4.draw(line,left) 0
```

[표 1]은 로그변환된 자료에 대한 2차추세모형 적합결과의 분산분석표와 회귀계수들의 추정값들이다. 회귀계수들의 유의확률들이 모두 0.0001로서 유의하고 결정계수 R^2 도 99.87%로서 설명력도 높다. 여기서 변수 t는 1960년은 1이고 매년 1년씩 증가하는 값을 갖는 시간변수이다. 추정된 모형식은 다음과 같다.

$$\ln(\widehat{pop}_{year}) = 7.81656 + 0.02557 \times t - 0.00024521 \times t^2$$

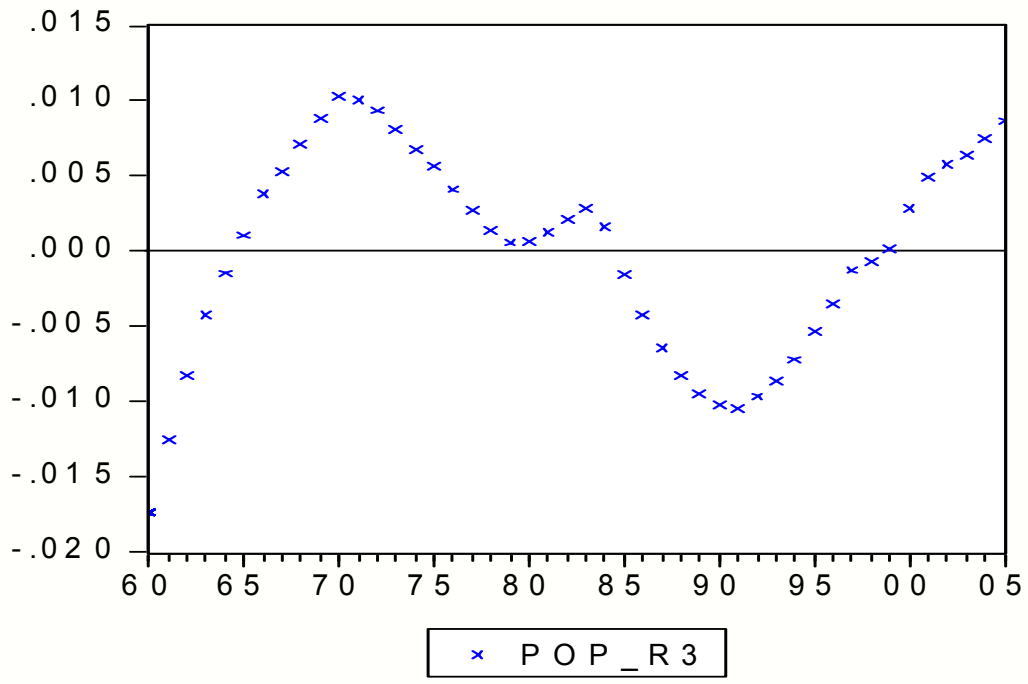
로그변환된 자료에 2차 추세모형을 적합시킨 후 잔차의 시계열 [그림 11]을 보면 [그림 10]과는 달리 시간의 흐름에 따라 변동의 폭이 커지지 않는 것을 볼 수 있어 로그변환으로 분산이 달라지는 문제점은 해결된 것으로 판단된다. 그러나 [그림 11]은 잔차들이 일정기간 동안은 양의 값을 가지다가 다시 일정기간동안은 음의 값을 갖는 패턴을 보여 주어 잔차들이 시간에 따른 상관관계를 가지고 있는 것 처럼 보인다.

[표 1] 우리나라 국내 인구추계의 2차 추세모형 추정 결과

Dependent Variable: LNPOP
Method: Least Squares
Date: 03/30/05 Time: 09:50
Sample: 1960 2005
Included observations: 46

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	7.816562	0.003254	2401.821	0.0000
T	0.025567	0.000319	80.04693	0.0000
T2	-0.000245	6.59E-06	-37.21386	0.0000

R-squared	0.998724	Mean dependent var	8.238763
Adjusted R-squared	0.998664	S.D. dependent var	0.192616
S.E. of regression	0.007040	Akaike info criterion	-7.011433
Sum squared resid	0.002131	Schwarz criterion	-6.892174
Log likelihood	164.2630	F-statistic	16821.78
Durbin-Watson stat	0.076991	Prob(F-statistic)	0.000000



[그림 11] 로그변환된 자료에 대하여 2차추세 모형 적합후의 잔차 시계열

추세모형의 적합시 유의할 점은 자료들이 시간에 따라 관측된 관계로 자기상관관계를 가질 가능성이 높다는 것이다.

오차들이 자기상관이 있는지 여부를 판단하기 위해 Durbin-Watson 검정을 시행하면 출력된 DW통계량의 값 D=0.077를 이용하여 1차 자기상관계수가 0이라는 귀무가설 H_0 : ρ=0를 결정할 수 있다.

독립변수의 개수가 2(p=2)이고 표본의 크기 45일 때 유의 수준 5%에서의 DW 통계량의 신뢰하한과 상한으로 각각 d_L=1.43, d_U=1.62이다. D=0.077<d_L=1.43이므로 H_0 : ρ=0을 기각한다. 즉, 오차들이 서로 독립이 아니라고 판단한다. 실제로 1차자기상관계수의 추정값이 0.873으로서 자기상관이 매우 높음을 확인할 수 있다. 이처럼 오차들이 자기상관을 가지고 있어 독립성의 가정이 만족되지 않는 경우에는 오차들의 공분산을 이용한 일반화 최소제곱법을 이용하거나 자기회귀오차모형(autoregressive error model)으로 적합시키는 방법이 있다.

(라) 계절추세모형

계절변동요만 있는 시계열은 삼각함수 또는 지수함수를 이용하여 계절성분을 설명한다. 예를 들어 주기 s(=12)인 계절 성분이 있는 시계열은 다음과 같은 삼각함수를 이용한다.

$$Z_t = \beta_0 + \sum_{i=1}^{m} A_i \sin\left(\frac{2\pi i}{s} t + \phi_i\right) + \quad _t$$

단, A_i: 진동폭(amplitude)
ϕ_i: 편각(phaseshift)

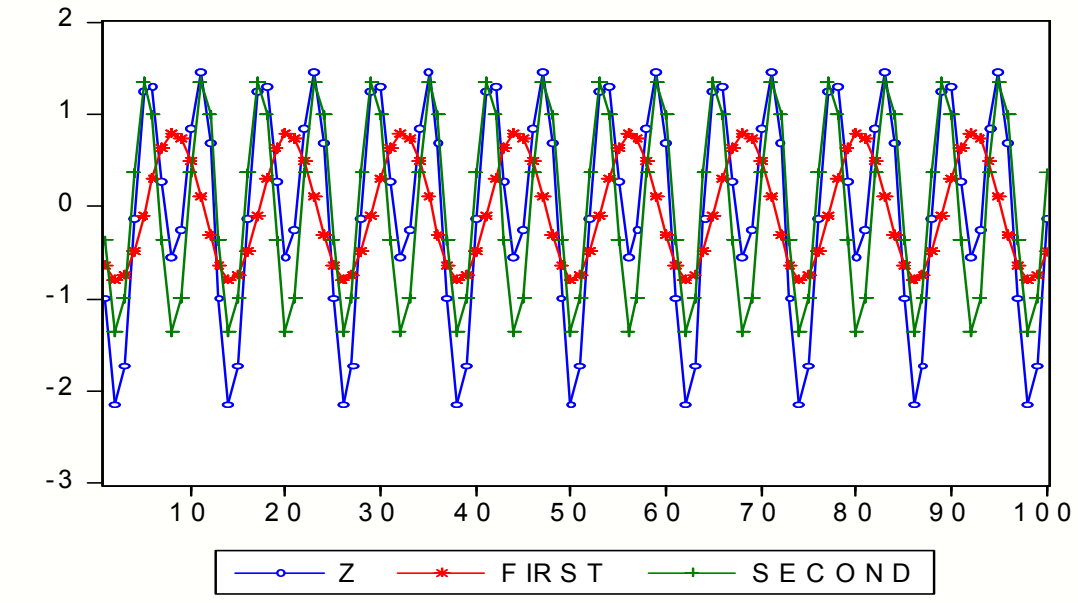
지시함수를 이용하는 경우에는

$$Z_t = \beta_0 + \sum_{i=1}^{s} \beta_i ID_{t,i} + \quad _t$$

단, $ID_{t,i} = \begin{cases} 1, & t = i \\ 0, & \text{그외의경우} \end{cases}$

사례2 : [그림 12]는 몇 개의 주기항들의 합이 어떻게 다양한 계절성분을 가지는 시계열자료를 설명하기 위하여 삼각함수식에서 m=2인 경우 β_0=0, A1=−0.8, A_2=1.4, ϕ_1=π/8, ϕ_2=3π/4를 가지는 E(Z_t)를 그린 것이다.

```
wfcreate u 1 100
!a1=-0.8
!a2=1.4
!pi=3.141592654

genr  phi1=!pi/8
genr  phi2=3*!pi/4

genr  first=!a1*@sin(!pi*(@trend+1)/6+phi1)
genr second=!a2*@sin(!pi*(@trend+1)/3+phi2)
genr  z=first+second

graph grap0.line(a) z first second
        grap0.setelem(1) symbol(1)
        grap0.setelem(2) symbol(4)
        grap0.setelem(3) symbol(6)
```



[그림 12] $E(Z_t)$ =(제1주기항)+(제2주기항)

(마) 선형계절추세모형

계절성을 갖는 대부분의 경제시계열은 선형추세와 계절변동요인을 동시에 포함하고 있다. 이러한 혀애의 시계열은

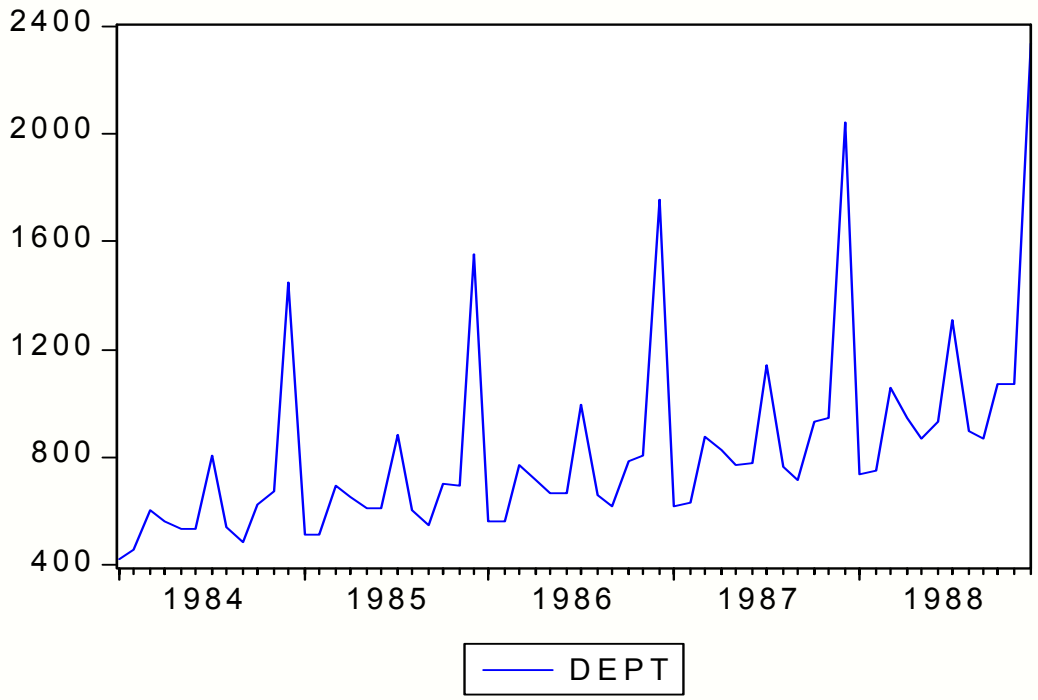
$$Z_t = \beta_0 + \beta_1 t + \beta_2 \sin(\frac{2\pi t}{12}) + \beta_3 \cos(\frac{2\pi t}{12}) + \quad _t$$

또는

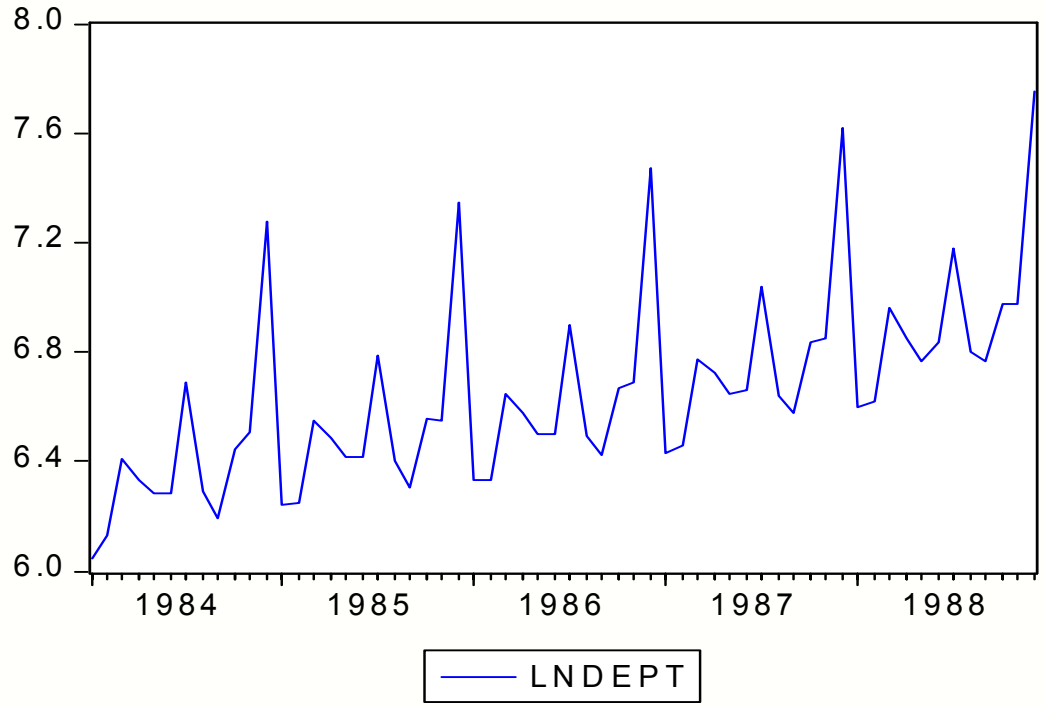
$$Z_t = \beta_0 + \beta_1 t + \sum_{i=1}^{s} \beta_{s_i} ID_{t,i} + \quad _t$$

사례 3: 1084년 1월부터 1988년 12월까지의 A백화점의 매출액(단위:십만원) 자료에 대한 추세요인과 계절요인을 동시에 가지는 모형을 설정해 보자. 시계열 그림을 그려보면 [그림 13]와 같이 추세와 함께 계정성이 있으며 시간의 흐름에 따라 자료의 변동 폭이 커짐을 알 수 있다. 이러한 자료의 분석을 위해서는 우선 자료의 적절한 변환을 통하여 변동폭을 균일하게 한다. 경제시계열의 대분분은 시간의 흐름에 따라 변동의 폭이 점차로 커지므로 로그변환 등을 적용하는 것이 일반적이다. 로그변환을 시행한 후의 시계열 [그림 14]을 보면 [그림 13]와는 달리 변동의 폭이 어느 정도 균일하여 졌음을 볼 수 있다.

```
wfcreate(wf="monthly") m  1984:01 1988:12
read(t=dat,norect)  c:\times\depart1.dat dept
genr lndept=log(dept)
genr t=@trend(1983:12)
for !n=1 to 12
series m{!n}=@seas({!n})
next

graph g1.line(a) dept
graph g2.line(a) lndept

equation  eq1.ls lndept  t m1 m2 m3 m4 m5 m6 m7 m8 m9 m10 m11 m12
eq1.makeresid  lndept_r

graph g3.line(a) lndept_r
g3.draw(line,left) 0
g3.setelem(1) symbol(5)
```



[그림 13] 백화점 월별 매출액의 시계열



[그림 14] 로그변환한 백화점 월별 매출액의 시계열

로그변환된 자료에 계절형 지시함수모형을 적합시켜보자. 이를 위하여 먼저 1984년 1월을 t=1로 시작하여 매월 1씩 증가하는 값을 가지는 시간변수 t를 독립변수로 지정하고 계절성을 파악하기 위해서 12개(주기가 12이므로)의 지시함수 I_1, I_2 ,.... I_{12}를 다음과 같이 지정

한다.

$$I_j = \begin{cases} 1, & \text{관측값이 j번째 달에 관측된 경우} \\ 0, & \text{그렇지 않은 경우} \end{cases}$$

모수들 사이의 제약 조건으로는 절편이 없는 모형을 적합시킨다. t와 I_1, I_2 ,.... I_{12}를 독립변수로 하는 선형계절추세모형을 적합시킨 결과 변수 t와 각 계절에 행당되는 지지함수들이 모두 유의 하므로 적합된 모형은 다음과 같다.

$$\widehat{\ln(DEPART_t)} = 0.010660 \times t + 6.064190 \times I_{t,1} + 6.080800 \times I_{t,2} + \ldots\ldots + 7.110482 \times I_{t,12}$$

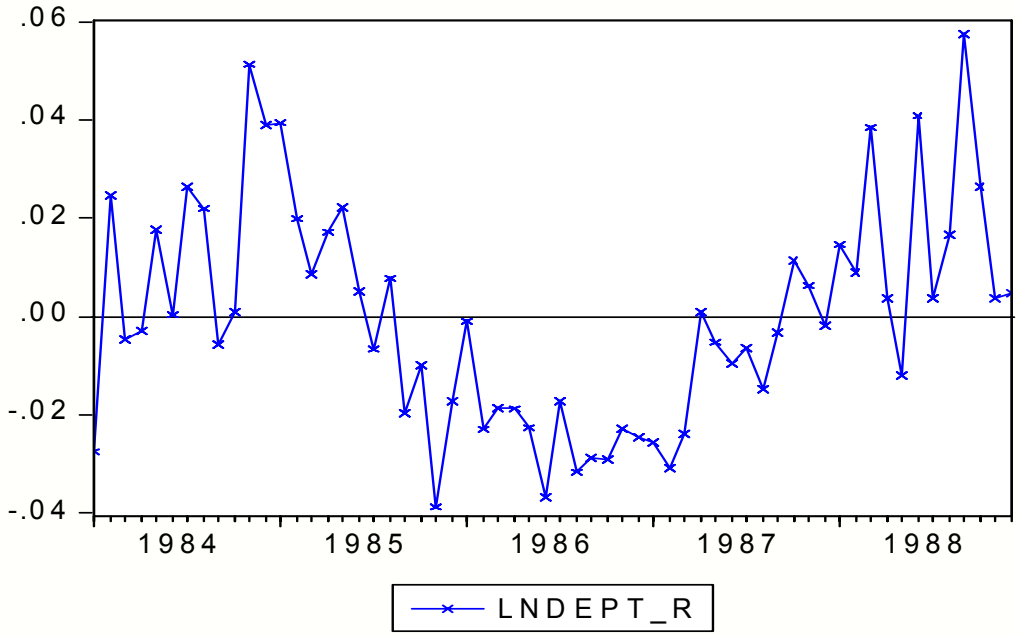
모형의 유의성검정에 관한 분산분석표에서 모형은 매우 유의하나, 잔차의 시계열 [그림 15]을 보면 일정기간 동안 동일한 부호를 갖는 잔차가 계속됨을 알 수 있다.

[표 2] 백화점 월별 매출액의 선형계절추세모형 추정 결과

Dependent Variable: LNDEPT
Method: Least Squares
Date: 03/30/05 Time: 11:06
Sample: 1984M01 1988M12
Included observations: 60

Variable	Coefficient	Std. Error	t-Statistic	Prob.	
@TREND(1983:12)		0.010660	0.000192	55.38796	0.0000
M1	6.064190	0.012295	493.2152	0.0000	
M2	6.080800	0.012372	491.5044	0.0000	
M3	6.381118	0.012451	512.5025	0.0000	
M4	6.295346	0.012532	502.3236	0.0000	
M5	6.213239	0.012616	492.4729	0.0000	
M6	6.219777	0.012703	489.6410	0.0000	
M7	6.588507	0.012791	515.0747	0.0000	
M8	6.184283	0.012882	480.0619	0.0000	
M9	6.100115	0.012975	470.1300	0.0000	
M10	6.333451	0.013071	484.5545	0.0000	
M11	6.341712	0.013168	481.5974	0.0000	
M12	7.110482	0.013268	535.9296	0.0000	

R-squared	0.995890	Mean dependent var	6.651223
Adjusted R-squared	0.994841	S.D. dependent var	0.352251
S.E. of regression	0.025300	Akaike info criterion	-4.326868
Sum squared resid	0.030085	Schwarz criterion	-3.873093
Log likelihood	142.8060	Durbin-Watson stat	0.826420



[그림 15] 선형계절추세모형 적합후의 잔차 시계열

또한 잔차들의 각 시차에 대한 자기상관계수가 자기상관계수 표준오차의 2배수인 $2/\sqrt{\text{표본크기}} = 2/\sqrt{60} = 0.2582$ 보다 크다면 그 시차에서 잔차들 간에 자기상관이 존재하는 것으로 판단된다. 잔차들의 자기상관들이 시차 1, 2, 3, 4, 5, 6에서 비교적 큰 자기상관관계가 존재함을 알 수 있다. 따라서 이 사례에서도 오차들이 서로 독립이라는 가정 하의 일반적인 추세모형 만으로는 설명이 되지 않음을 알 수 있다. 이와 같이 시간에 따라 관측된 시계열 자료의 경우에는 오차들이 서로 독립이라는 가정이 만족되기 어려우므로 자기회귀오차모형이나 ARIMA모형을 이용하는 것이 바람직하다.

[표3] 잔차의 자기상관계수

Date: 03/30/05 Time: 11:37

Sample: 1984M01 1988M12
Included observations: 60

Autocorrelation	Partial Correlation		AC	PAC	Q-Stat	Prob
|**** |	|**** |	1	0.574	0.574	20.763	0.000
|**** |	|** |	2	0.499	0.253	36.741	0.000
|**** |	|*** |	3	0.579	0.350	58.591	0.000
|*** |	| |	4	0.424	-0.045	70.555	0.000
|*** |	|* |	5	0.451	0.148	84.340	0.000
|*** |	| |	6	0.418	-0.013	96.366	0.000
|** |	*| |	7	0.258	-0.150	101.04	0.000
|** |	*| |	8	0.223	-0.123	104.59	0.000
|** |	| |	9	0.212	-0.010	107.87	0.000
| |	**| |	10	0.016	-0.262	107.89	0.000
| |	| |	11	0.037	0.004	108.00	0.000
*| |	**| |	12	-0.088	-0.243	108.60	0.000
*| |	|* |	13	-0.138	0.067	110.11	0.000
*| |	| |	14	-0.116	-0.046	111.20	0.000
.*| .	. |*.	15	-0.187	0.089	114.10	0.000

(바) 자기회귀오차모형

회귀모형을 이용하여 자료를 분석할 때 일반적으로 오차항이 서로 독립이라는 가정을 전제로한다. 오차의 독립성 가정이 지켜지지 않고 가지상관관계가 존재할 경우 최소제곱법을 이용한 회귀분석은 모수추정의 효율이 떨어지고 편기(bias)를 가지거나 예측값의 신뢰구간 및 유의성검정 등에 오류가 있을 수 있고 잔차들에 규칙적인 상관관계의 정보가 존재하여 예측값을 더욱 개선할 여지가 남아 있다.

시계열 자료를 회귀모형에 적합시킬 때 오차들이 시간에 따른 자기상관관계를 갖게 되는 경우를 흔히 접하게 된다. 특히 오차항이 자기회귀과정(autoregressive process)을 따르는 경우의 k 차 자기회귀오차모형(autoregressive error model)은 다음과 같다.

$$Z_t = \beta_0 + \beta_1 X_{t1} + \beta_1 X_{t2} + \ldots\ldots + \beta_1 X_{tp} + \quad_t, \quad t = 1,2,\ldots\ldots,n$$

여기서

$$\quad_t = \phi_1 \quad_{t-1} + \phi_1 \quad_{t-2} + \ldots\ldots + \phi_1 \quad_{t-k} + a_t$$

이고 a_t 들은 서로 독립이다.

자기회귀오차모형이 일반회귀모형과 다른점은 오차항인 ε_t 들이 회귀모형에서처럼 서로 독립이지 않고 자기상관관계를 가지며 특히 k차 자기회귀과정 AR(k)를 따른다는 점이다.

오차들의 자기상관 여부는 DW검정을 사용한다.

$$D_j = \frac{\sum_{t=j+1}^{n} (\widehat{a_t} - \widehat{a_{t-j}})^2}{\sum_{t=1}^{n} \widehat{a}_t^{\,2}}$$

사례 4 : 사례 3에서 로그변환된 백화점 매출액 자료의 잔차 시계열 [그림15]과 잔차의 자기상관계수에 대한 근사검정의 결과로부터 오차항들이 서로 독립이지 않고 자기상관관계가 존재하고 있다. 이 시계열 자료에 자기회귀오차모형을 적합시키기 위해 자연로그변환된 자료에 절편없는 선형계절추세모형을 기본모형으로 하고 오차항에 대해서는 적절한 시차의 AR모형을 적합시킨다. 시차를 몇차까지 고려하는냐 하는 것은 자료에 따라 다르지만 이 자료인 경우에는 오차항의 자기상관계수를 15차까지의 유의성을 보면 5시차까지 유의한 것으로 보인다.([표 3] 참고).

```
equation  eq2.ls lndept  t m1 m2 m3 m4 m5 m6 m7 m8 m9 m10 m11 m12 ar(1) ar(2) ar(3)
ar(4) ar(5)
equation  eq2.ls lndept  t m1 m2 m3 m4 m5 m6 m7 m8 m9 m10 m11 m12 ar(1)  ar(3)
eq2.makeresid  lndept_r1

graph g4.line(a) lndept_r1
g4.draw(line,left) 0
g4.setelem(1) symbol(5)
```

잔차들의 자기회귀계수가 시차 2,4,5차에서는 유의하지 않다고 판정되고 최종적으로 시차 1과 3만이 유의한 시차로 선택되었다. 따라서 다음의 자기회귀오차모형이 적합된다.

$$\widehat{\ln(DEPART_t)} = 0.01\times t + 6.06\times I_{t,1} + 6.08\times I_{t,2} + \cdots\cdots + 7.11\times I_{t,12} + \varepsilon_t$$

$$\varepsilon_t = 0.38\,\varepsilon_{t-1} + 0.39\,\varepsilon_{t-3}$$

$$\widehat{Var}(\varepsilon_t) = 0.000356$$

단 EViews에서는 $\varepsilon_t = \phi_1\,\varepsilon_{t-1} + \phi_1\,\varepsilon_{t-2} + \cdots\cdots + \phi_1\,\varepsilon_{t-k} + a_t$의 모형식을 사용하므로 추정된 $\hat{\phi}_i$들의 부호가 반대로 정의되어 있음을 유의한다. 이때 DW검정의 유의확률은 0.4963이므로 오차항들 간에 더 이상의 체계적인 상관관계가 존재한다고 볼 수 없다. 이것은 잔차 시계열 [그림 16]에서도 확인된다.

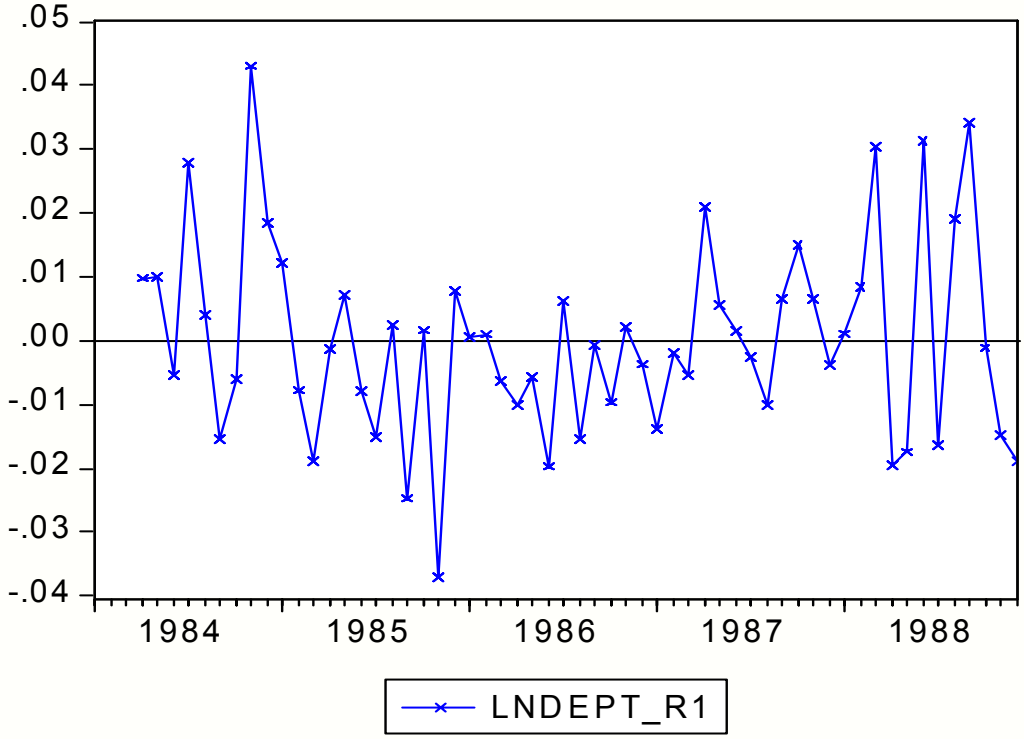
[표 4] 자기회귀오차모형 추정결과

Variable	Coefficient	Std. Error	t-Statistic	Prob.
AR(1)	0.457477	0.164932	2.773726	0.0086
AR(2)	-0.086711	0.180325	-0.480861	0.6334
AR(3)	0.359609	0.163608	2.197986	0.0343
AR(4)	-0.050862	0.187304	-0.271546	0.7875
AR(5)	0.198636	0.167915	1.182959	0.2444

Dependent Variable: LNDEPT
Method: Least Squares
Date: 03/30/05 Time: 14:05
Sample (adjusted): 1984M04 1988M12
Included observations: 57 after adjustments
Convergence achieved after 3 iterations

Variable	Coefficient	Std. Error	t-Statistic	Prob.
T	0.010764	0.000835	12.89054	0.0000
M1	6.071008	0.037679	161.1260	0.0000
M2	6.073136	0.037791	160.7021	0.0000
M3	6.380111	0.038088	167.5087	0.0000
M4	6.296770	0.037073	169.8489	0.0000
M5	6.210315	0.036932	168.1534	0.0000
M6	6.217374	0.037075	167.6986	0.0000
M7	6.587230	0.036910	178.4669	0.0000
M8	6.181866	0.036951	167.3003	0.0000
M9	6.097362	0.037161	164.0797	0.0000
M10	6.330950	0.037323	169.6278	0.0000
M11	6.338880	0.037558	168.7748	0.0000
M12	7.107356	0.037883	187.6116	0.0000
AR(1)	0.449131	0.124658	3.602904	0.0008
AR(3)	0.374460	0.123050	3.043155	0.0040

R-squared	0.997942	Mean dependent var	6.675274
Adjusted R-squared	0.997256	S.D. dependent var	0.343025
S.E. of regression	0.017968	Akaike info criterion	-4.979543
Sum squared resid	0.013559	Schwarz criterion	-4.441897
Log likelihood	156.9170	Durbin-Watson stat	1.938166
Inverted AR Roots	.91	-.23-.60i -.23+.60i	



[그림 16]자기회귀오차모형 적합 후의 잔차 시계열

제3장 평활법

시계열자료가 갖는 변수는 변화하는 정도가 심한 경우가 보통이다. 순환변동과 계절변동은 물론 불규칙변동까지 작용하기 때문이다. 이러한 변화가 심한 시계열 자료로부터 변수가 갖는 추세변동과 순환변동 및 계절변동 요인을 정확하게 파악하기란 어려운 일이다. 왜냐하면 본래 시계열 자료는 변동이 심하여 들쭉날쭉한 변수값들이 존재하기 때문이다. 그러므로 변화 폭이 심한 시계열자료를 평탄하고 변화가 완만한 값으로 변환시키는 것을 평활(smoothing) 이라 한다. 즉 평활이란 들쭉날쭉한 시계열자료 값을 평탄한 값으로 변환하는 것이다.

많은 경제 및 사회 시계열자료들은 변동폭이 큰 계절성분을 포함하거나 설명할 수 없는 불규칙변동을 포함하고 있기 때문에 추세를 정확하게 파악하기 어렵다. 실제로 정부통계 작성기관이나 기업 등에서는 정확한 추세파악을 위하여 계절조정된 시계열자료를 필요로 하는 경우가 많다.

따라서 시계열분석에서 불규칙변동을 통제하거나 제거할 수 있다면 각 변동요인에 대한 추정은 신뢰성을 갖게 된다. 시계열자료에서 불규칙변동요인을 통제하고 추세, 순환, 계절변동요인을 파악하기 위하여 본래 자료값을 변환시키는 방법이 평활법이다. 그런데 대표적인 평활법으로 이동평균법 (moving average method) 과 지수평활법 (exponential smoothing method) 가 있다.

관측된 시계열들의 평균수준이 시간대에 관계없이 거의 변하지 않는, 즉 불규칙변동만을 포함하는 시계열인 수평적계열(horizontal series)에 대해서는 단순이동평균법과 단순지수평활법이, 추세성을 갖는 시계열에 대해서는 선형이동평균법과 선형 및 이중지수평활법이, 그리고 추세 및 계절성을 갖는 시계열에 대해서는 계절지수평활법이 적용된다.

또한 지수평활법이 예측의 목적으로 주로 사용되는 반면에 이동평균법은 예측의 목적보다는 분해법에서 계절조정하는데 주로 이용된다.

1. 이동평균법(moving average method)

이동평균법은 시계열자료들의 평활에 의해 계절변동 또는 불규칙변동을 제거하여 전반적인 추세를 뚜렷하게 파악할 수 있도록 해준다. 이동평균법은 표본평균처럼 모든 관측값에 동일한 가중치를 주는 대신에 일부 자료에만 동일한 가중치를 준다.

시계열자료로부터 일정한 기간에 해당하는 자료 값을 묶어 반복해서 산술평균을 구하면 새로운 시계열자료를 얻을 수 있다. 이러한 새로운 시게열자료를 구하는 방법을 이동평균법이라 한다. 원래 시계열자료는 변화가 많아 들쭉날쭉한 자료값을 갖는데 이동평균은 원 자료에 숨겨진 어떤 추세 경향을 개괄적으로 파악할 수 있게 한다.

이제 시계열 변수 Z_t 가 있다고 한다. 그리고 t 시점에서 m 기간 범위로 구한 이동평균을 $\overline{Z_m}$ 이라 하자. 그러면 $\overline{Z_m}$ 은 m 이동평균이라 부르고 다음 식으로 구할 수 있다.

m 이동평균 : $\overline{Z_m}=(Z_t+Z_{t+1}+\ldots..+Z_{t+(m-1)})/m, \quad t=1,2,\ldots.,(t-m)$

2. 지수평활법

지수평활법은 최근의 자료에 가중치를 크게 주고 과거 자료일수록 점차적(지수적 : exponentially)으로 가중치를 줄여나가는 방법이다. 따라서 최근의 자료를 주로 이용하여 미래를 예측해주므로 시계열의 구조에 변화가 있을 경우 이 변화에 쉽게 대처할 수 있으며 그 계산법이 쉽고 많은 자료의 저장이 필요없다는 등의 장점이 있어 많은 사람들이 사용하여 왔다. 많이 사용하는 지수평활법으로는 단순지수평활법, 이중지수평활법, 삼중지수평활법과 Winters의 계절지수 평활법 등이 있다.

시계열자료가 다음과 같은 모형을 따른다고 생각해 보자

$$Z_t = \beta_0 + \ _t$$

여기서 $_t$은 서로 독립이고 평균 0, 분산 σ^2를 가지며 β_0는 시간에 따라 변화하는 미지의 모수이다. 따라서 시계열이 추가될 때마다 추가 자료를 포함하여 β_0의 추정값을 수정해야 한다.

단순지수평활법은 최근의 자료에 더 큰 가중치를 부여하는 다음과 같은 모형을 설정한다.

$$F_{n+1} = \alpha Z_n + \alpha(1-\alpha) Z_{n-1} + \alpha(1-\alpha)^2 Z_{n-2} + \ldots\ldots = SM_n$$

여기서 α는 평활상수라고 하고 0과 1사의 값을 갖으며, SM_n은 시점 n에서의 평활값이라고 한다. 따라서 단순지수평활법은 시계열 자료들에 대한 가중평균인 평활값으로 미래를 예측하는 방법이라고 할 수 있다.

이 식을 축소하면

$$F_{n+1} = \alpha Z_n + (1-\alpha) F_{n-1}$$
$$\text{또는} \quad SM_n = \alpha Z_n + (1-\alpha) SM_{n-1}$$

여기서 $F_{(n-1)+1}$는 시전 n-1에서 예측 시점 n의 예측값이고 시전 n에서의 예측오차는 $E_n = Z_n - F_n$이 된다. 따라서 α, Z_n, 그리고 F_n이 주어지면 $F-n+1$이 계산할 수 있다. 단순지수평활 예측을 하기 위해서는 초기값 F_1이 주어져야 한다. 일반적으로 초기값 F_1은 표본평균 $\bar{Z}$를 이용하거나, 첫 번째 관측치 Z_1를 주로 이용한다.

지수평활법에 의한 예측에서는 관측값들에 대한 가중치의 역할을 하는 평활상수 α의 결정이 매우 중요한 문제가 된다. 일반적으로 평활상수 α의 값이 작으면 평활의 효과가 커서 예측값 F_n은 시계열의 최근 변화에 대해 둔감하고 서서히 반응을 보인다. 반면에 α의 값이 크면 평활의 효과가 작아 예측값 F_n은 최근의 관측값에 의해 크게 영향을 받으므로 시계열의 최근 변화에 민감하게 반응을 보인다. 따라서 시계열의 수준 변화가 심한 경우는 α의 값이 1에 근사하기 때문에 초기 평활값으로 Z_1을 선택하는 것이 바람직하고, 시계열자료가 안정적이고 변화가 완만한 경우는 α가 0에 근접하므로 표본평균 $\bar{Z}$를 선택하는 것이 적절할 것이다.

적절한 α값의 결정은 다음과 같은 예측오차의 제곱합을 최소로 하는 α값을 결정하면 된다.

$$MSE(\alpha) = \frac{1}{n} \sum_{1}^{n} (Z_t - F_t)^2$$

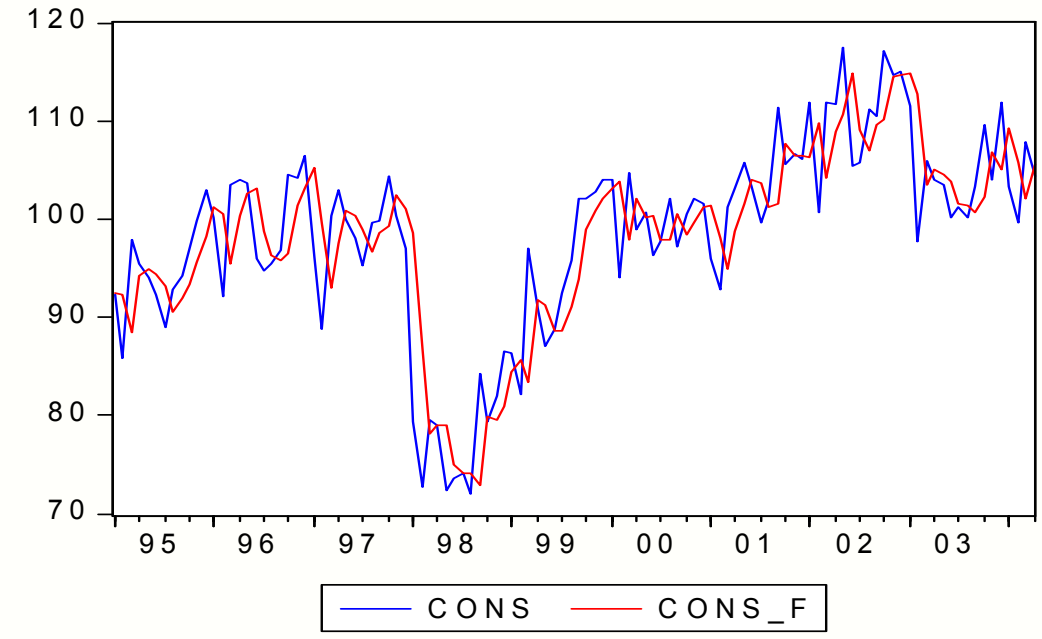
예제 3-1) 1995년 1월부터 2004년 12월까지의 내수용소비재출하지수이다.(자료출저 통계청) 이 자료의 시계열그림인 [그림 17]에서 실선으로 그려진 원자료를 보면 평균수준이 시간대별로 다르다고 판단되어 단순지수평활법을 적용하여 보기로 하자.

EViews의 smooth를 이용하면 [표 5]와 같이 예측오차의 제곱합 $SSE(\alpha)=\sum_{1}^{n}(Z_t-F_t)^2$이 최소인 평활상수 α=0.614값과 오차제곱의 합, 오차제곱평균(MSE)의 제곱근이 산출된다.

```
wfcreate m 1995:01 2004:04
read(t=dat,norect) c:\times\consumer.dat cons

smooth(s) cons cons_f
graph   g1.line cons cons_f
            g1.setelem(1) lpat(1)
            g1.setelem(2) lpat(2)
```

[표 5] α=0.614인 경우의 단순지수평활한 결과

Date: 03/31/05 Time: 16:21		
Sample: 1995M01 2004M04		
Included observations: 112		
Method: Single Exponential		
Original Series: CONS		
Forecast Series: CONS_F		
Parameters:	Alpha	0.6140
Sum of Squared Residuals		3337.727
Root Mean Squared Error		5.459042
End of Period Levels: Mean		104.8114

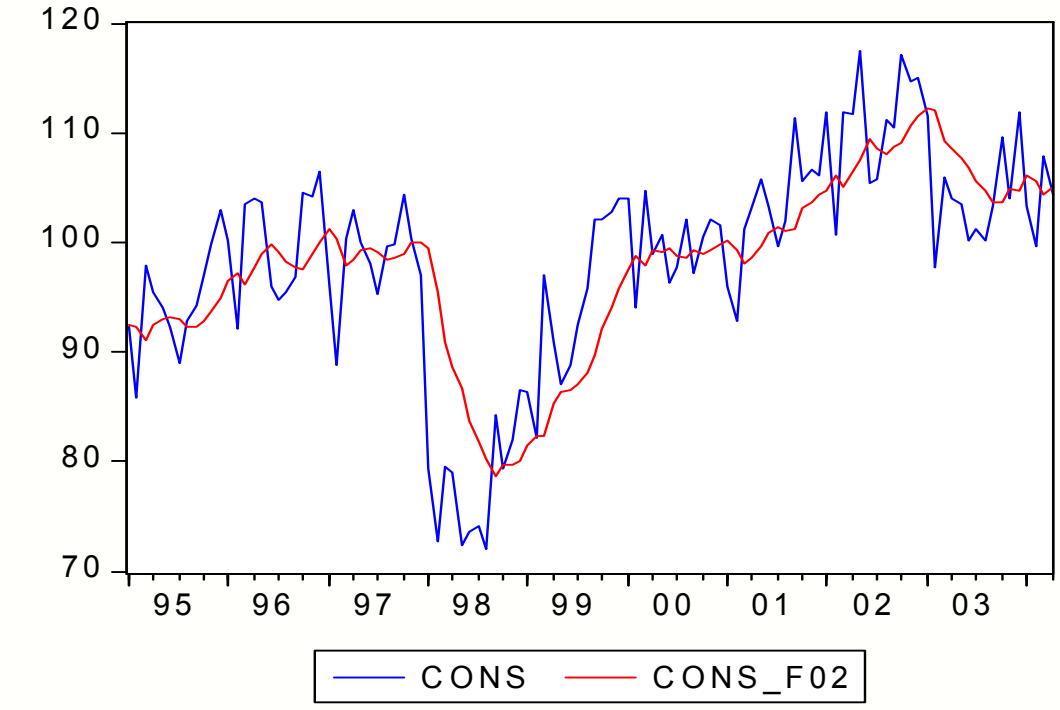


[그림 17] 내수소비재출하지수와 단순지수평활값(α=0.614)

```
smooth(s,0.2) cons cons_f2
graph   g2.line cons cons_f2
        g2.setelem(1) lpat(1)
        g2.setelem(2) lpat(2)
genr er6=cons-cons_f1
genr er2=cons-cons_f2
graph g3.line er6
g3.draw(line,left)   0
graph g4.line er2
g4.draw(line,left) 0
```

[표 6] α=0.2인 경우의 단순지수평활한 결과

Date: 03/31/05 Time: 16:51		
Sample: 1995M01 2004M04		
Included observations: 112		
Method: Single Exponential		
Original Series: CONS		
Forecast Series: CONS_F02		
Parameters:	Alpha	0.2000
Sum of Squared Residuals		4497.468
Root Mean Squared Error		6.336873
End of Period Levels: Mean		104.9522



[그림 18] 내수소비재출하지수와 단순지수평활값(α=0.2)

[표 7] 단순지수평활법에 의한 1시차 후 예측값과 예측오차

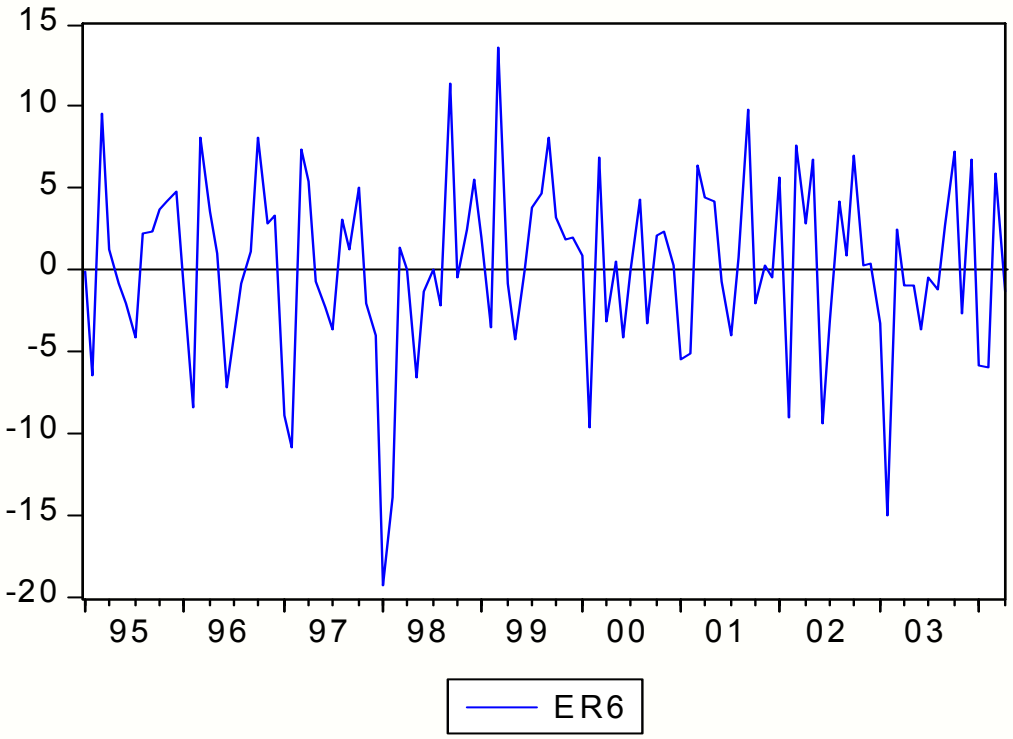
			$\alpha=0.614$		$\alpha=0.2$	
T	DATE	$\hat{Z}_t$	$\widehat{Z_{t-1}}(1)$	$\widehat{\ }_{t-1}(1)$	$\widehat{Z_{t-1}}(1)$	$\widehat{\ }_{t-1}(1)$
1	1995M01	92.33000	92.38686	-0.056857	92.38686	-0.056857
2	1995M02	85.90300	92.35195	-6.448947	92.37549	-6.472486
3	1995M03	97.90500	88.39228	9.512725	91.08099	6.824011
4	1995M04	95.44800	94.23311	1.214885	92.44579	3.002209
5	1995M05	94.12500	94.97906	-0.854058	93.04623	1.078767
109	2004M01	103.4000	109.2910	-5.891031	106.1487	-2.748655
110	2004M02	99.70000	105.6739	-5.973921	105.5989	-5.898924
111	2004M03	107.9000	102.0059	5.894083	104.4191	3.480861
112	2004M04	104.3000	105.6249	-1.324900	105.1153	-0.815311

그러나 α=0.614는 일반적으로 사용되도록 권해지는 값 0.01에서 0.3과는 차이가 있다. 따라서 α=0.2를 사용하여 지수평활한 결과인 [그림 18], [표 6] 그리고 [표 7]을 함께 살펴보자. [표 7]에는 α=0.614와 α=0.2인 경우의 1 시차후 예측값 $\widehat{Z_{t-1}}(1)$과 예측오차 $\widehat{\ }_{t-1}(1)\ Z_t(1)-\widehat{Z_{t-1}}(1)$이 계산되어 있다. 따라서 평활상수 α=0.614를 사용하는 편이 더 예측력이 좋은 것을 알 수 있다.

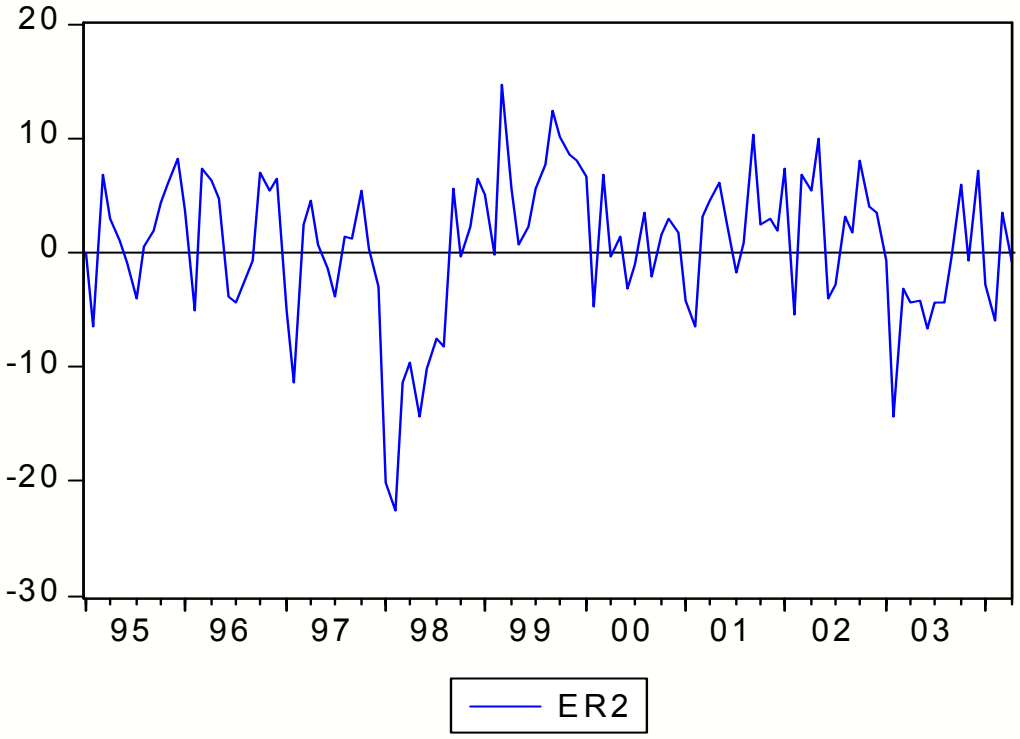
추가로 예측오차들에 어떤 규칙적인 상관관계가 존재하는지를 검사해야할 것이다. 예측오차들이 0에 관해 랜덤한 형태를 보이지 않는다면 예측오차에 예측의 정보가 남아 있는 것으로 판단되고 이것은 예측값이 더 개선될 여지가 있음을 의미하는 것이다.

[표 8] 예측오차의 자기상관계수

시차	$\alpha=0.614$	$\alpha=0.2$
1	0.049	0.506
2	-0.084	0.315
3	-0.121	0.204
4	-0.034	0.190
5	0.064	0.203
6	0.121	0.188
7	0.078	0.103
8	-0.104	-0.054
9	-0.152	-0.129
10	-0.216	-0.163
11	0.033	-0.016
12	0.470	0.188
Mean	0.18067	0.56095
Std. Dev.	5.48057	6.34036
Probability	0.7278	0.3511



[그림 19] 평활상수 α=0.614인 경우의 예측오차



[그림 20] 평활상수 α=0.2인 경우의 예측오차

[표 8]에는 예측오차의 자기상관계수와 평균 및 표준편차 예측오차의 평균이 0인지에 관한 검정의 유의확률 등이 계산되어 있다.
각 시차에서 예측오차의 상관계수가 자기상관계수 표준오차의 2배수인

$$2 \cdot \frac{1}{\sqrt{\text{표본크기}}} = 2 \cdot \frac{1}{\sqrt{112}} = 0.19 \text{ 보다}$$

크다면 그 시차에서 예측오차간의 자기상관계수가 존재하는 것으로 판단된다. [표 8]으로부터 α=0.2의 경우에는 시차 1, 2, 3, 5에서 비교적 큰 자기상관관계가 존재한다. 물론 α=0.614의 경우에도 시차 10, 12에서 자기상관이 존재하므로 예측법을 더 개선할 필요가 있다. 예측오차의 시계열 그림 [그림 19], [그림 20]을 보도라도 α=0.2의 경우가 α=0.614의 경우 보다 더 체계적인 상관관계가 있음을 알 수 있다.
귀무가설 "H_0 : 예측오차의 평균이 0이다."에 대한 대립가설 "H_1 : 예측오차의 평균이 0이 아니다"의 검정에서 α=0.614의 경우 유의확률(0.7278)이 0.10보다 커서 귀무가설을 기각할 수 없고 α=0.2의 경우도 유의확률(0.3511)이 0.10보다 커서 귀무가설이 기각을 기각할 수 없다. 그러나 α=0.614 경우의 예측오차들이 α=0.2 경우보다도 0에 더 랜덤하다고 판단된다.

3. 계절지수평활법

단순지수평활법은 시계열이 일정한 주기성을 보이지 않을 때 사용되는 방법으로서, 시계열이 계절변동과 같이 일정한 주기를 가질때는 적합하지 않은 방법이다. 계절변동을 포함하고 있는 시계열들도 그 변동의 모양에 따라 두가지 모형으로 구분할 수 있다. 즉 시계열의 평균 수준이 시간의 흐름에 따라 변화하지만 그 변동의 폭 즉 분산이 시간의 흐름에 관계없이 일정할 경우(Homogeneous)에 사용되는 가법계절모형(additive seasonal model)과 분산이 시간의 흐름에 따라 점차로 커지는 경우에 사용되는 승법계절모형(multiplicative seasonal model)로 나눌 수 있다.

시계열 변동의 폭과 계절주기의 폭이 추세에 비례하여 변화할 때 사용되는 원터스(Winters) 승법계절모형은 추세변동과 계절변동의 곱에 의해 시계열을 나타낸다. 원터스의 승법계절모형은 시점 n+l에서 다음과 같다

$$\begin{aligned} Z_{n+1} &= T_{n+l} \cdot S_{n+l} + I_{n+l} \\ &= (T_{n+l} + \beta_{1,n} l) S_{n+l} + I_{n+l} \end{aligned}$$

단, T_{n+l}은 추세변동, S_{n+l}은 계절주기 s를 가지는 계절변동 그리고 I_{n+l}은 오차항으로 불규칙변동이다.

원터스의 승법계절지수평활법은 추세변동과 계절변동의 곱에 불규칙변동이 더해져서 시계열이 구성되어 있다고 가정하여 각 변동요인들을 평활법에 의해 추정한 후 이를 이용하여 예측값을 구한다. 일반적으로 추세변동은 선형추세로 고려하지만 시계열의 특성에 따라 고차추세모형도 고려할 수 있다.
승법 계절지수평활에 의한 예측은 다음과 같은 식에 의해 예측값이 생성된다. 즉 n+l점에서

$$\hat{T}_{n+l} = \alpha_1 \left(\frac{Z_{n+1}}{\hat{S}_{n+1-s}} \right) + (1-\alpha_1)(\hat{T}_n + \hat{\beta}_{1,n})$$

$$\hat{\beta}_{1,n+1}=\alpha_2(\hat{T}_{n+1}-\hat{T}_n)+(1-\alpha_2)\hat{\beta}_{1,n}$$

$$\hat{S}_{n+1}=\alpha_3(\frac{Z_{n+1}}{\hat{T}_{n+1}})+(1-\alpha_3)\hat{S}_{n+1-s}$$

단 평활상수 $\alpha_1, \alpha_2, \alpha_3$는 모두 0과 1사이의 값으로서 각 변동요인들의 변화 정도가 급하게 진행될수록 1에 가가운 값을, 변화 정도가 아주 서서히 이루어질수록 0에 가까운 값을 가질 것이다. 과거의 자료를 이용하여 여러 평활상수 집합($\alpha_1, \alpha_2, \alpha_3$)에 대하여 1시차후 예측오차제곱합 SSE($\alpha_1, \alpha_2, \alpha_3$)을 계산하여 이를 최소로 해주는 ($\alpha_1, \alpha_2, \alpha_3$)값을 평활상수로 채택하면된다.

초기평활값 그리고 $\hat{S}_{j-s}(j=1,2,\ldots,s)$를 구하기 위해먼저 매 계절주기 간격내의 관측값들의 평균을 구한다.

$$\bar{Z}_1=\frac{1}{s}\sum_{i=1}^{s}Z_i, \quad \bar{Z}_2=\frac{1}{s}\sum_{i=s+1}^{2s}Z_i, \ldots, \bar{Z}_k=\frac{1}{s}\sum_{i=(k-1)s+1}^{s}Z_i,$$

이것을 이용하여 다음과 같이 초기평활값 그리고 $\hat{S}_{j-s}$를 구할 수 있다.

$$\hat{\beta}_{1,0}=\frac{\bar{Z}_k-\bar{Z}_1}{(k-1)s}$$

$$\hat{T}_0=\hat{Z}_1-\frac{s}{2}\hat{\beta}_{1,0}$$

$$\hat{S}_{j-s}=\frac{s\hat{R}_j}{\sum_{i=1}^{s}\hat{R}_i} \quad i=1,2,\ldots,s$$

단,

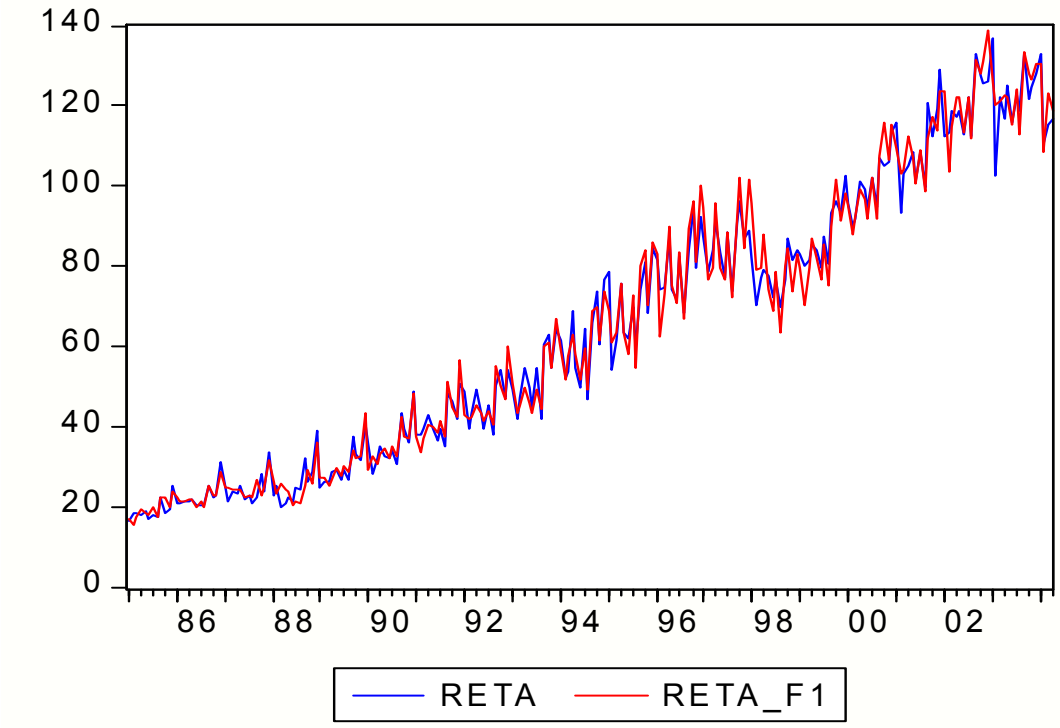
예제 3-2) 1985년 1월부터 2004년 12월까지의 종합소매업판매액지수이다.(자료출저 통계청) 이 자료의 시계열그림인 [그림 21]에서 실선으로 그려진 원자료를 보면 평균수준이 시간대별로 다르다고 계절변동요인이 존재하는 것으로 판단되어 승법계절지수평활법을 적용하여 보기로 하자.

EViews의 smooth를 이용하면 [표 9]와 같이 예측오차의 제곱합 $SSE(\alpha)=\sum_{1}^{n}(Z_t-F_t)^2$이 최소인 평활상수 α_1=0.25, α_2=0.14, α_3=0.63값과 오차제곱의 합, 오차제곱평균(MSE)의 제곱근이 산출된다.

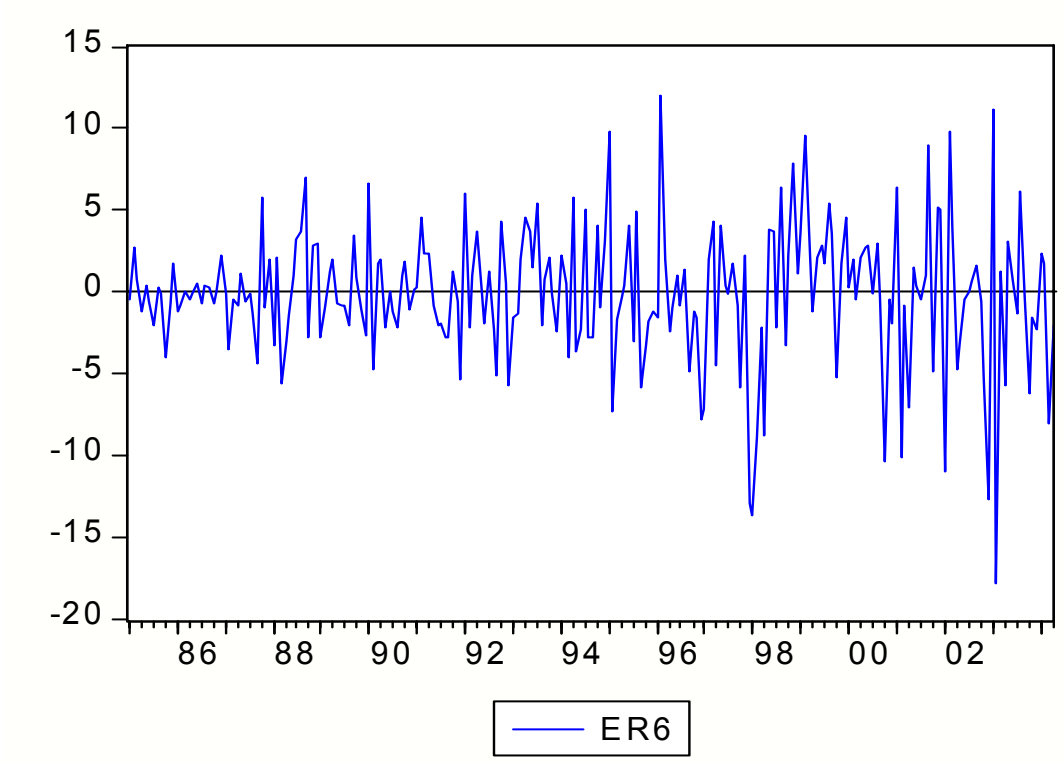
```
wfcreate m 1985:01 2004:04
read(t=dat,norect) c:\times\retail.dat reta

smooth(m) reta reta_f1
graph g1.line reta reta_f1
g1.setelem(1) lpat(1)
g1.setelem(2) lpat(2)
```

[표 9] 평활상수 α_1=0.25, α_2=0.14, α_3=0.63에 의한 승법계절지수평활법 결과

Date: 04/01/05 Time: 13:44			
Sample: 1985M01 2004M04			
Included observations: 232			
Method: Holt-Winters Multiplicative Seasonal			
Original Series: RETA			
Forecast Series: RETA_F1			
Parameters:	Alpha	0.2500	
	Beta	0.1400	
	Gamma	0.6301	
Sum of Squared Residuals			4013.861
Root Mean Squared Error			4.159462
End of Period Levels:	Mean	120.9532	
	Trend	-0.244490	
Seasonals:	2003M05	1.019563	
	2003M06	0.947042	
	2003M07	1.008637	
	2003M08	0.947072	
	2003M09	1.074123	
	2003M10	1.005305	
	2003M11	1.021653	
	2003M12	1.054544	
	2004M01	1.064986	
	2004M02	0.881771	
	2004M03	0.997016	
	2004M04	0.978288	



[그림 21] 종합소매업판매액지수와 계절지수평활법의 예측치



[그림 22] 계절지수평활법의 예측오차

“예측오차의 평균이 0이다”의 검정에서 유의확률이 0.5049로 0.10보다 커서 오차들의 평균이 0를 중심으로 랜덤하다고 볼 수 있다.그러나 [그림 22]을 보면 시간의 흐름에 따라 예측오차의 크기가 커져 예측의 정확도가 떨어짐을 알 수 있다.

제4장 계절조정방법

경제시계열과 관련된 계절변동이란 대부분이 1년 즉 12개월을 기준으로 규칙적으로 반복되는 변동을 의미한다. 이러한 변동은 자연적인 현상에 기인하는 것으로 1년을 주기로 하는 관습적인 인간의 생활이 다양한 경제 및 사회활동에 영향을 미친다. 그러나 대부분의 계절변동이 경제의 실질적인 움직임과는 무관하게 매년 일정한 형태을 가지고 반복되며, 이러한 변동의 폭이 매우커서 장기적인 추세나 경기순환의 움직임을 정확하게 파악하기 어렵게 하는 경우가 많다. 따라서 관습적으로 반복되는 움직임을 원 시계열에서 제거할 수 있다면 다른 변동요인들의 움직임을 좀 더 정확하게 분석·예측할 수 있을 것이라는 것이 계절조정의 의미다.

계절조정을 위한 프로그램으로 Shiskin(1967) 등에 의해 미국 센서스국에서 개발된 X-11 방법이 널리 사용되어 왔다. 그러나 이동평균과정에서 시계열의 양끝에서 결항(결측)값이 생기는 문제를 해결하기 위하여 캐나다 통계청의 Dagum(1975)이 X-11-ARIMA를 개발하였다.

X-11-ARIMA방법은 미국을 비롯하여 서구 여러나라의 경제시계열에는 적절할 수 있으나 우리나라와 같이 설과 추석과 같이 음력을 이용한 명절이 있는 경우에는 한계를 가지고 있다. 즉 경제활동일수를 사전조정하는 과정에서 미국 또는 캐나다와 우리의 문화의 차이를 고려하지 않아 명절효과와 조업일수(영업일수)효과, 이상치 제거가 제대로 반영되지 않는다는 한계를 지니고 있다. 이러한 문제점을 고려하여 미국 상무부에서는 1995년말부터 X-12-ARIMA를 개발하여 지속적으로 프로그램을 UPdate하고 있으며 계절조정에도 활용하고 있다.

X-12-ARIMA는 X-11-ARIMA방법의 한계를 해결할 수 있도록 RegARIMA라는 사전조정모형을 채택하고 있다. 그러나 세계 각국에서는 경제시계열의 계절조정을 할 때 그 나라의 실정에 적합한 방법을 선택하여 사용하고 있는 실정이다. 스페인 중앙은행에서는 1997년에 TRAMO/SEATS를 개발하여 사용하고 있다.

X-12-ARIMA에서는 X-11ARIMA에서 다루었던 시계열의 구조를 가법, 승법 및 로그가법모형 이외에도 추가적으로 의가가법모형(pseudo-additive model)을 적합시킬 수 있다. 의사가법모형은 영국중앙통계청에서 개발되어 사용한 방법이다. 일반적인 시계열 변동요인들은 다음과 같은 형태으로 이루어진다.

$$\begin{aligned} \text{승법:} \quad & O_t = T_t \times C_t \times S_t \times I_t \\ \text{가법:} \quad & O_t = T_t + C_t + S_t + I_t \\ \text{로그가법:} \quad & \log(O_t) = T_t + C_t + S_t + I_t \\ \text{의사가법:} \quad & O_t = T_t \times C_t(S_t + I_t - 1) \end{aligned}$$

여기서, T_t:추세변동, C_t:순환변동, S_t:계절변동, I_t:불규칙변동,

여기서 의사가법 모형은 매년 일정한 월 또는 분기에 0에 가까운 값을 갖는 시계열의 계절조정에 적합하다. 즉 농산물의 생산량과 같이 매년 일정한 시기에 0에 가까운 값을 갖는다면 해당 월 또는 분기의 계절변동 값이 0에 가까워질 것이다. 이러한 시계열에 승법모형을 가정한다면 원계열을 계절변동요인으로 나누어 계절조정계열을 구하는 과정에서 0에 가까운 값으로 나누게 되므로 문제가 발생한다.

그러나 X-12-ARIMA방법에서는 시계열의 변동요인을 좀 더 구체적으로 다음과 같이 분해하고 있다.

승법: $O_t = T_t \times C_t \times S_t \times D_t \times H_t \times E_t \times I_t$

여기서, D_t:요일변동, H_t:명절효과, E_t:이상치(특이치)

요일변동은 월별로 경제활동일수가 달라서 발생하는 변동 효과이다. 일요일과 공휴일이 어느해 어느달에 더 많이 들어 있는가에 따라 그 달에 실제로 조업(영업)일수가 다른데서 파생되어 시계열에 영향을 주는 변동효과이다. 또한 명절효과는 설과 추석이 1월과 2월, 9월과 10월에 각각 이동함으로써 파생되는 경제활동에 영향을 주는 효과이고, 이상치는 사전에 탐지할 수 없는 이상점에 의한 효과를 나타내는 변동요인이다.

X-12-ARIMA에 의한 계절조정과정은 3단계의 계산과정, 즉 RegARIMA모형에 의한 시계열 확장 및 사전조정과정, 이동평균방법에 의한 시계열구성요인 분해과정 그리고 계절조정결과에 대한 사후진단과정으로 이루어진다.

1. RegARIMA모형을 이용한 사전조정단계

이 과정은 시계열의 특성을 고려하여 좀더 순수한 계절변동요인을 산출하기 위한 사전조정과정이다. 사전에 파악이 가능한 요일변동(조업일수), 윤년, 명절효과 및 파업효과 등의 사전조정요인과 AO(additive outlier), LS(level shift), TC(temporary change) 등과 같이 사전에 파악되지 않는 이상점(outlier)을 설명변수로 하는 RegARIMA모형을 원시계열(원계열) O_t에 적합시킨다. X-12-ARIMA는 이상점의 존재 여부를 자동적으로 탐지하는 기능을 포함하고 있으므로 탐지된 이상점을 사전조정요인과 함께 설명변수로 하는 회귀모형을 적합시킨다. 원 시계열에서 이들 변동요인을 미리 제거하여 안정적인 원계열을 산출하며, 이동평균을 적용하는 과정에서 생기는 결측값을 보정하기 위하여 시계열의 양끝의 값을 적합된 ARIMA모형으로 1년씩 예측하여 확장하는 과정이다.

X-11-ARIMA에 비해 가장 보완된 단계로서 조정하고자 하는 사전효과들을 설명변수로 모형에 포함시켜 추정이 가능하도록 하였다. 특히 사전에 파악이 되지 않는 이상점들을 탐지할 수 있는 기능을 보완하였다는 점에서 X-11-ARIMA와 큰 차이점이라 할 수 있다.

X-12-ARIMA에서도 X-11-ARIMA의 경우와 같이 승법모형과 가법모형에 대해 각각 5개씩의 표준 ARIMA모형을 설정하여 제공하고 있다. 이들 모형을 시계열에 적용하여 잔차들의 Ljung-Box 검정결과 유의확률이 5%이상이고, 최근 3년동안의 평균절대예측오차(MAE)가 15%보다도 작으며, 과대차분(overdifferencing)의 징후가 없는 모형을 선택한다. 만약 표준모형이 모두 기준에 적합하지 않으면 사용자가 모형을 지정해야 한다.

여기서 5개의 표준모형은 (0 1 1)(0 1 1)s, (0 1 2)(0 1 1)s, (2 1 0)(0 1 1)s,
(0 2 2)(0 1 1)s, (2 1 2)(0 1 1)s이다.

2. 시계열 구성요인 분해과정

이 과정은 사전조정된 시계열을 추세, 순환, 계절 및 불규칙변동요인으로 분해하여 최종적으로 계절조정계열을 구하는 계산과정이다. 여러 가지 이동평균방법을 적용하여 각 변동요인을 다음과 같은 계산과정으로 반복하여 산출한다.

① 원계열 O_t의 중심화 12개월 이동평균을 취함으로써 계절변동성분 S_t와 불규칙변동성분 I_t를 제거하여 잠정적인 추세순환변동성분 TC_t를 산출한다.

② O_t로부터 잠정적인 TC_t를 제거하여 잠정적인 S_tXI_t를 산출한다.
[$O_t/TC_t = S_tXI_t$]

③ 잠정적인 S_tXI_t를 동일 월의 연도별 계열로 분류하고 각각에 대해 수년치를 가중 이동평균함으로써 I_t를 제거하여 잠정적인 S_t를 산출한다.

· I/S비율에 따라 MSR, (3x3)항, (3x5)항, (3x9)항,(3x15)항 등의 이동평균방법을 이용한다.

④ O_t로부터 잠정적인 S_t를 제거하여 잠정적인 계절조정계열 TC_tXI_t를 산출한다. [$O_t/S_t=TC_tXI_t$]

⑤ 잠정적인 계절조정계열 TC_t+I_t에 적당한 Henderson 가중이동평균을 실시함으로써 I_t를 제거하여 수정된 TC_t를 산출한다.

· 5항, 7항, 9항, 13항, 23항 등의 Henderson의 가중이동평균방법을 이용한다.

⑥ 수정된 TC_t를 사용하여 ②~④의 과정을 반복함으로써 최종적인 3가지 성분(TC_t, S_t, I_t)을 산출한다.

또한 이 계산과정에는 특이치(극단값)이 변동요인의 산출결과를 좌우하는 것을 최소화하기 위하여 불규칙변동요인의 이동 5개년 표준편차 σ를 계산하여 2.5σ를 벗어나는 특이치를 제거한 후 다시 이동 5개년 표준편차 σ를 계산하는 과정이 포함되어 있다.

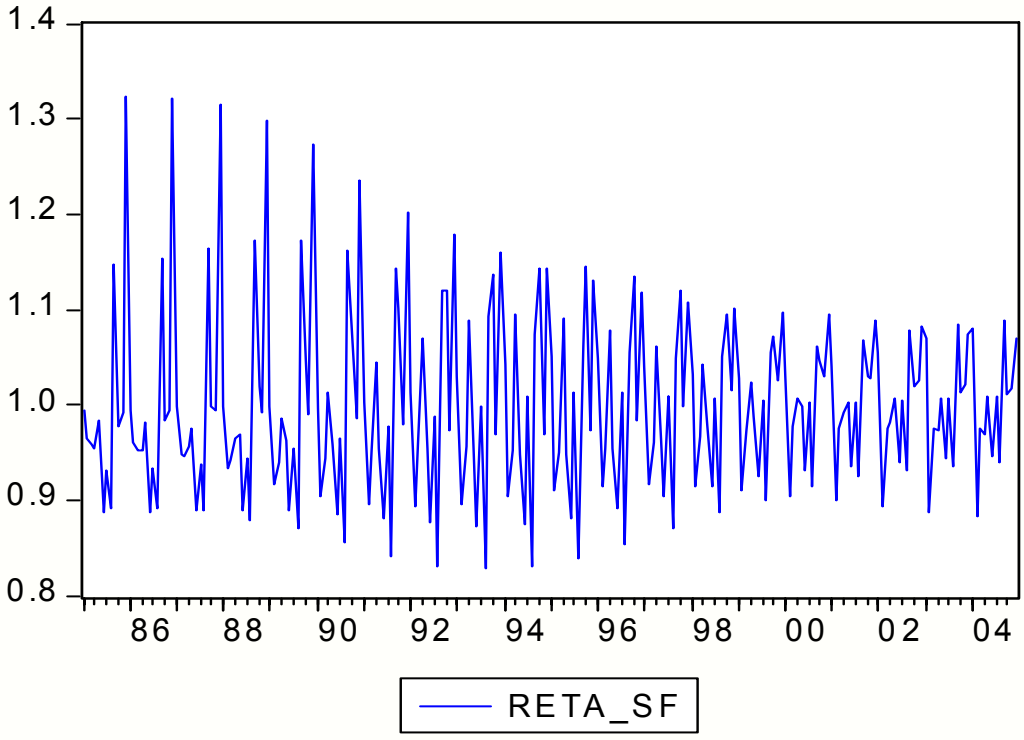
3. 통계적 평가과정

이 과정에서는 사전 및 이동평균 계산과정에서 산출된 시계열 구성요인들이 얼마나 잘 산출되었는지에 관한 여러 가지 통계적평가 및 진단을 결과를 산출한다. 평가 통계값으로 Q-통계량, M1-M11 품질관리통계량들 이외에도 계절변동요인과 요일변동효과가 남아 있는지를 파악하기 위한 스펙트럼추정량과 계절조정된 계열의 안정성을 진단하기 위한 sliding span 과 revision history diagnostics 통계값이 산출된다.

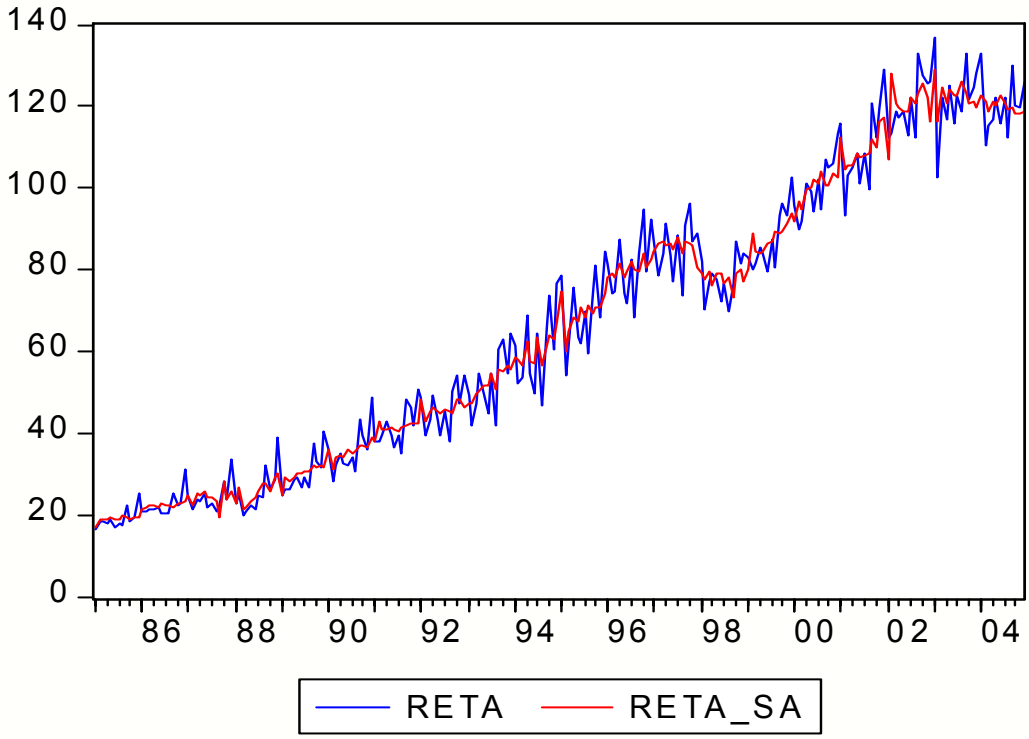
예제 4) 종합소매업판매액지수를 계절조정하기 위하여 Eviews를 이용하여 최종적인 계절변동요인(D10), 계절조정계열(D11), 추세순환변동요인(D12), 불규칙변동요인(D13)을 구해보자.

```
wfcreate m 1985:01 2004:04
read(t=dat,norect) c:\times\retail.dat reta
reta.x12(tf=0,amdl=b,reg="td1coef",save="d10 d11 d12")
```

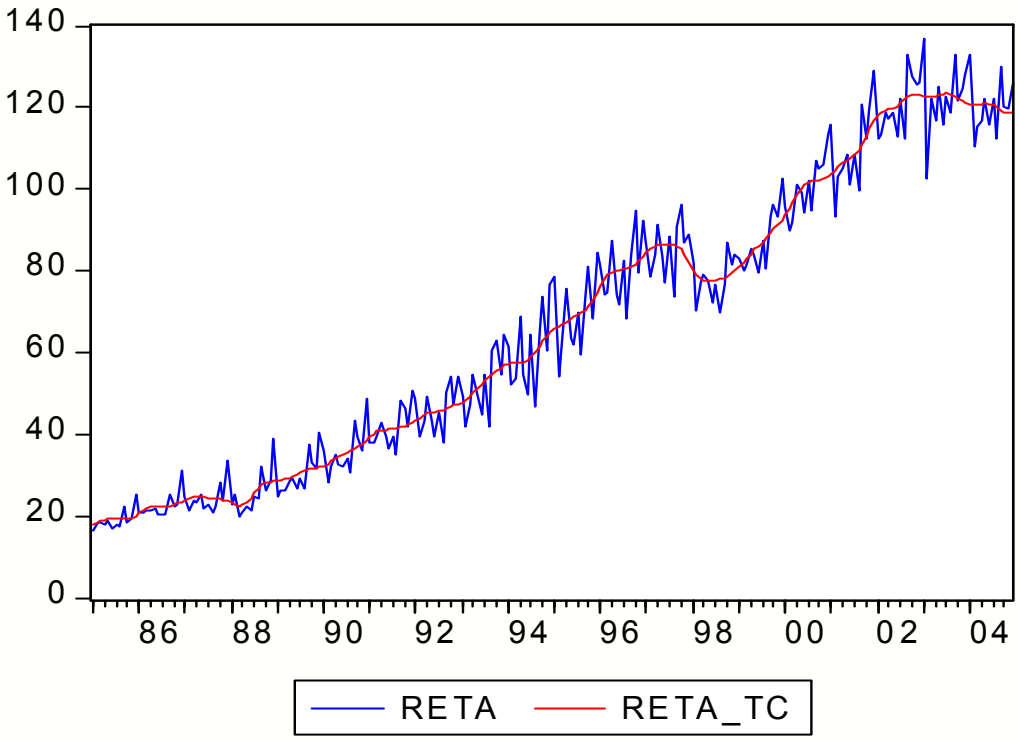
T	DATE	원계열 O_t	계절요인 (D10)	계절조정계열 (D11)	추세요인 (D12)	불규칙요인 (D13)
1	1985M01	16.6	0.994374	16.79058	18.04185	0.930646
2	1985M02	18.2	0.964148	19.04531	18.35650	1.037524
3	1985M03	18.2	0.959188	18.82887	18.68933	1.007466
4	1985M04	18.1	0.954053	19.04486	18.97644	1.003605
5	1985M05	18.9	0.983042	19.33736	19.19010	1.007674
6	1985M06	17.1	0.886699	19.10032	19.29459	0.989931
234	2004M06	115.8	0.946814	122.7767	120.8667	1.015802
235	2004M07	122.0	1.009264	120.7639	120.5186	1.002035
236	2004M08	112.2	0.940513	119.1818	120.0077	0.993118
237	2004M09	130.0	1.089167	119.8176	119.4155	1.003368
238	2004M10	120.4	1.010868	118.1921	118.9486	0.993640
239	2004M11	119.6	1.016361	118.1286	118.7206	0.995014
240	2004M12	126.1	1.069799	118.5552	118.6621	0.999099



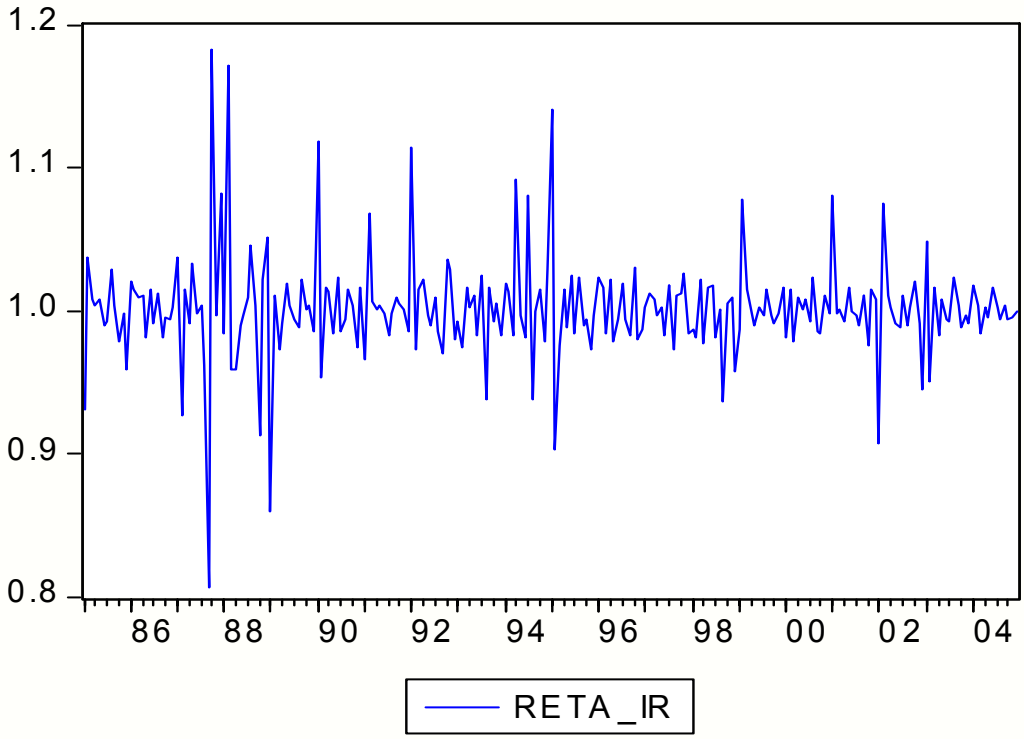
[그림 23] 종합소매업판매액지수의 계절변동요인



[그림 24] 종합소매업판매액지수의 계절조정계열



[그림 25] 종합소매업판매액지수의 추세순환계열



[그림 26] 종합소매업판매액지수의 불규칙변동요인