계절조정방법 이론과
X-12-ARIMA 사용 방법

2006. 9.

서비스업동향과

< 목　　차 >

PART A: 계절조정방법

PART B: X-12-ARIMA 프로그램의 사용방법

머 리 말

원계열에서 순수한 경기적 요인만을 추출하기 위한 방법은 중심화 이동평균을 이용하는 이동평균법과 원계열이 어떠한 시계열 모형을 따른다는 가정하에 계절요인과 추세·순환요인 등의 시계열 구성요인을 분해하여 추출하는 신호추출법(Signal Extraction) 등이 있습니다.

이중 이동평균법은 X-11-ARIMA, X-12-ARIMA 등에 적용하고 있으며, 신호추출법은 TRAMO/SEATS에 적용되고 있습니다.

본 매뉴얼은 TRAMO/SEATS와 함께 현재 세계적으로 가장 널리 사용되고 있는 X-12-ARIMA에 대한 내용을 PART A와 PART B로 나누어 구성하였습니다.

PART A에서는 X-12-ARIMA에 적용되고 있는 RegARIMA 모형, 사전조정을 위한 요일·명절효과·특이치 등의 회귀변수 설정 방법, 이동평균법에 의한 계절조정 절차, 계절조정 결과에 대한 평가 방법, 총합계열의 계절조정 등 계절조정방법에 대한 전반적인 내용을 설명하고자 노력하였습니다.

PART B는 X-12-ARIMA/0.3에서 사용하는 명령어인 ARIMA, AUTOMDL, REGARIMA, X11 등에 대한 옵션 설명을 미국 센서국의 원 매뉴얼 내용에 충실히 따르고자 하였습니다.

끝으로 본 매뉴얼은 1997년 발간되었던 『X-12-ARIMA의 계절조정방법(통계연수원)』을 대폭 수정·보완하였으며, 작성에는 문권순 서비스동향과장, 이대용 사무관, 송금영 사무관, 정미숙, 박진호, 전택련, 오승철, 이은정 등 서비스동향과의 계절조정연구회 회원들의 노력이 들어가 있음을 밝혀둡니다.

본 매뉴얼 내용은 통계청의 공식 입장이 아닌 집필자 사견이며, 발생되는 오류는 집필자의 몫임을 밝혀둡니다. 집필 내용에 대한 문의는 문권순 과장(ksmoon@nso.go.kr)에게 해주시기 바랍니다.

2006. 9.

서비스업동향과장 문 권 순

Ⅰ. 계절조정 개요

1.1 시계열의 구성

시계열은 시계열 자체가 갖고 있는 본래의 특성에 의해서 나타나는 변동과 이외의 변동으로 구성되어 있다고 할 수 있다. 즉, 대부분 경제·사회 시계열의 경우, 어떠한 경제·사회적 특성의 변동 뿐만 아니라 일정한 주기에 의해서 되풀이 되는 기후와 사회관습 등 비경제·사회적 원인의 변동에 의해서 시계열의 주기 및 구조, 경기·사회 전환점 등의 왜곡을 가져올 수 있어 정확한 현상 파악을 하는데 장애요인으로 작용할 수 있다. 즉, 비경제·사회적 원인에 의한 경기·사회변동을 제거하지 않으면 실제 경기·사회변동보다 과대 또는 과소 평가함으로써 전혀 다른 판단을 내릴 수 있다.

시계열 변동요인을 추세요인(Trend components: T), 순환요인(Cycle components: C), 계절요인(Seasonal components: S), 불규칙요인(Irregular components: I)으로 나누며, 원계열(Y)을 다음과 같은 시계열 모드 형태로 표현할 수 있다(Dagum, 1988).

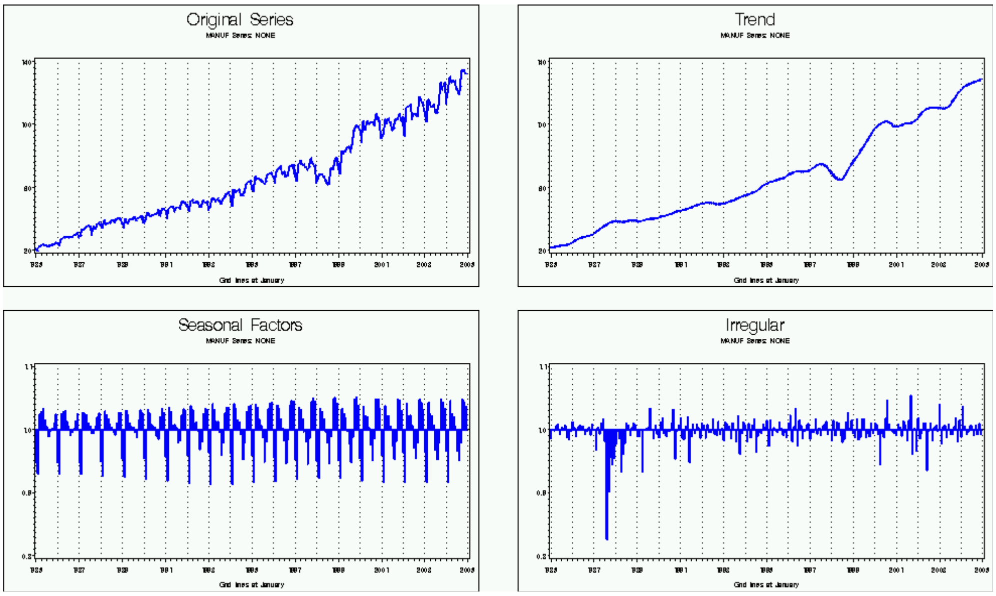
- 승법모형(Multiplicative) : $Y = T \cdot C \times S \times I$
- 가법모형(Additive) : $Y = T \cdot C + S + I$
- Log-가법모형(Log Additive) : $\log(Y) = \log(T \cdot C) + \log(S) + \log(I)$

X-12-AIMA에서는 이들 3개의 모형 이외에도 다음과 같은 pseudo 가법모형을 가정하고 있다.

- pseudo 가법모형(Pseudo Additive) : $Y = T \cdot C \times [S + I\text{-}1]$

여기서 추세·순환요인 중 추세요인은 대체로 10년 이상 동일한 방향으로 상승 또는 하강하는 경향을 나타내는 장기적인 변동을 말한다. 순환요인은 확장과 수축의 현상을 반복하는 주기적인 변동이다. 계절요인은 일정한 기간 동안(일반적으로 1년)에 반복되는 기후, 온도, 생활습관 등에 따라 나타나는 변동이다. 불규칙요인은 위의 세 가지 변동이외의 천재지변, 전쟁, 파업, 급격한 정책변화 등에 의한 변동 뿐만 아니라 명절, 요일의

차이 등에 의한 변동을 말한다. 다음 그림은 원계열(Original), 추세순환(Trend), 계절요인(Seasonal Factors), 불규칙요인(Irregular)이다.



한편, 불규칙요인을 승법모형으로 가정하고 좀 더 자세히 분해하면 다음과 같이 정의할 수 있다.

$$I = TD \times LY \times H \times O \times I'$$

여기서 TD는 월 또는 분기 시계열에서 요일의 횟수차이로 발생되는 변동인 요일효과(Trading day effect), LY는 4년마다 발생하는 2월 윤년에 의한 윤년효과(Leap year effect), H는 음력 명절인 설・추석이 일정한 월에 발생하지 않고 (1월 2월), (9월 10월)에 나타남으로써 발생되는 명절효과(Holiday effect)[1], O는 파업・정책변화 등에 의한 변동인 특이치(Outlier), I′는 나머지 이외의 불규칙요인이다. 이들 TD, LY, H, O는 3.2절에서 자세히 다룬다.

앞에서 논의된 시계열 모드 형태는 다음과 같은 특징을 갖고 있다(U.S. Census Bureau, 2004.a). 승법 계절성을 가진 대부분의 시계열은 계열수준이 증가하고 감소함에 따라 계절변동도 비례적으로 증가·감소하는

1) 서양의 경우, 부활절효과(Easter effect)로 부활절은 3월과 4월에 발생한다.

형태로 나타난다. Pseudo-가법 모형은 어떤 월(분기)에는 방학이나 기후 등에 의해 계절성이 극히 작고, 나머지 월은 승법 계절성을 나타날 때 사용한다. 계절성의 크기가 계열의 수준에 의해서 영향을 받지 않는 것으로 나타나면, 가법 계절성을 가지고 있다고 보고 가법 모드를 사용한다.

승법 분해방법의 대안으로 주로 사용하는 log-가법 모드는 경제분석에서 유용하게 사용할 수 있다. log-가법 계절조정 모드는 먼저, 추세요인을 log 변환한 원계열($\log(O)$)을 가법 분해하여 구한 후, 원계열과 같은 단위를 갖는 추세를 나타내기 위해 이를 지수화(exponential)한다.

1.2 계절조정방법

계절조정(Seasonal Adjustment: SA)이란 원계열이 경기적 요인과 비경기적 요인으로 구성되어 있다고 가정할 때, 비경기적 요인인 계절요인과 예측 가능한 불규칙요인을 원계열에서 제거하는 절차를 말한다. 여기서 예측 가능한 불규칙요인은 주로 명절효과, 요일효과, 특이치 등을 말한다. 승법모형을 가정할 때 다음과 같이 표현할 수 있다.

$$Y = T \cdot C \times \underbrace{S \times TD \times LY \times H \times O}_{\text{제거되는 비경기적 요인}} \times I'$$

계절조정은 크게 계절조정단계와 사전조정(Prior Adjustment)단계로 나눌 수 있다. 계절조정단계는 시계열을 추세·순환요인, 계절요인과 불규칙요인으로 시계열을 분해하는 단계이다. 시계열 요인을 분해하는 방법은 이동평균을 이용하는 linear filtering 방법(Daguem, 1988)과 시계열이 ARIMA 모형을 따른다는 가정 하에서 시계열에 내재된 요인들을 추출하는 신호추출방법(Burman, 1980) 등이 있다. 전자의 linear filtering 방법은 X-11, X-11-ARIMA과 X-12-ARIMA 등에 적용하고 있으며, 신호추출방법은 TRAMO/SEATS(Caporello, Maravall and Sanchez, 2003)에 적용하고 있다. 이외 계절조정방법은 SABL, CPB 등이 있으며(Butter와 Fase, 1991), 최근 X-12-ARIMA와 TRAMO/SEATS 방법을 결합한 DEMETRA가 EUROSTAT(2002)에서 출시되었다.

사전조정단계는 추정 가능한 불규칙요인인 윤년(LY) 및 요일효과(TD),

명절효과(H) 등의 캘린더 효과와 특이치(O)를 계절조정을 하기 전에 조정해 주는 단계이다. 특히, 우리나라의 설(1월, 2월)과 추석(9월, 10월), 서양의 부활절(3월, 4월)은 크리스마스와 달리 명절이 있는 월이 연도에 따라 변한다. 이를 이동 명절월(Moving Holiday)이라 하며, 명절이 있는 월에는 생산, 소비, 고용 등의 경제현상이 명절이 없는 월과 달리 나타난다. 이를 명절효과(Holiday effects)라고 하며, 시계열의 변동이 경제・사회적 원인보다 캘린더 구성에 따라 달리 나타나게 되므로 계절조정 시 감안해 주어야 한다.

사전조정방법은 캐나다의 X-11-ARIMA의 경우, 계절조정에 의해 추정된 초기 불규칙요인을 이용하여 사전조정 요인을 추정하고 있으나, 사용된 ARIMA 모형, 특이치, 계절 filter 등 계절조정방법에 따라 달라지는 단점이 있다. 한편, 미국 Census Bureau의 X-12-ARIMA와 유럽의 TRAMO/SEATS는 원계열과 윤년・요일・명절효과 등의 사전조정변수를 전이함수모형인 RegARIMA[2] 모형에 회귀시켜 사전조정 요인을 추정하는 Bell과 Hillmer(1983) 방법을 적용하고 있다. Bell과 Hillmer의 RegARIMA 방법은 원계열의 정보를 이용하는 한편 모형 선정에 대한 통계적 이론을 강화하였다(사전조정방법은 Ⅲ장에서 다룬다).

1.3 증감률

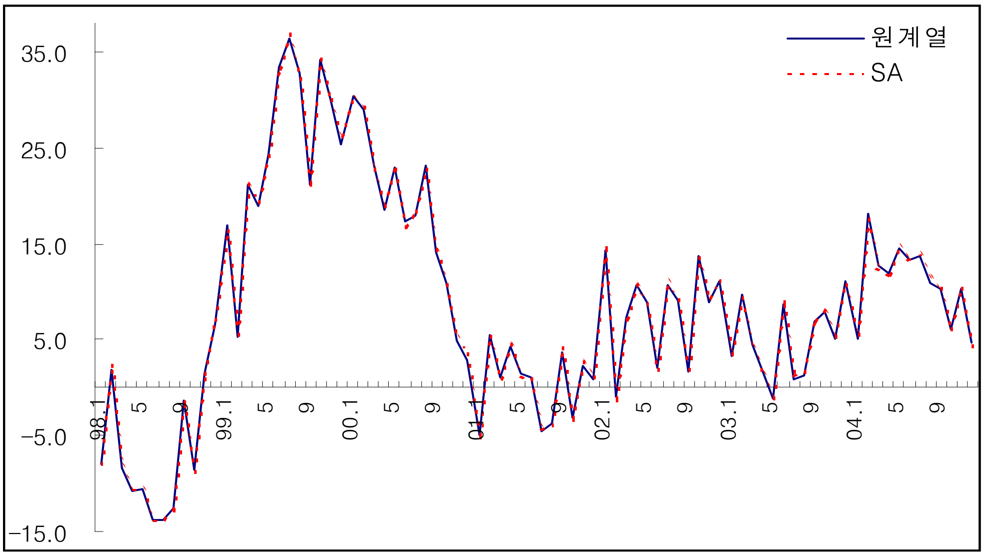
<그림 1.1>은 산업생산지수(1985년1월부터 2004년 12월)에 대한 원계열과 계절조정계열(SA)의 전년동월비이다. 그림을 보면, 원계열과 계절조정계열의 전년동월비 차이는 거의 없는 것으로 보인다. 이는 비경기적 요인인 계절요인이 매년 동일한 월에는 크게 변하지 않기 때문이다. 따라서 전년동기비는 계절조정 프로그램이 없는 경우, 계절변동요인을 제거하는 가장 간단한 방법이라고도 할 수 있다.

그러나 전년동기비를 사용할 때는 전년의 시계열 수준이 당해연도 증감률 변동에 영향을 미치므로, 전년도의 시계열 수준과 경기(사회)동향의

2) 최근 Census Bureau의 X-12-ARIMA/0.3에 x11regression 명령을 추가함으로써, 불규칙요인에 요일효과, 명절효과 등 사전조정변수를 회귀하여 사전조정요인을 추정할 수 있도록 하고 있다. 본 매뉴얼에서는 PART B, x11regression에서 간단히 다루도록 한다.

움직임을 충분히 파악하고 사용해야 한다. 또한 전년동기비는 경기 전환점(turning point)이 전월비에 비해 늦게 나타나므로 잘못된 정보를 제공할 가능성이 있음을 주의해야 한다(통계분석과, 1997).

<그림 1.1> 전년동월비



<그림 1.2>는 산업생산지수(1985년1월부터 2004년 12월)에 대한 원계열과 계절조정계열(SA)의 전월비이다. 그림을 보면, 원계열을 사용할 때, 계절적 원인으로 인하여 주기적으로 큰 폭의 증가와 감소가 있음을 알 수 있다. 따라서 전월비를 이용 시에는 계절요인이 제거된 계절조정계열을 사용하여야 함을 알 수 있다. 또한 계절조정 전월비 역시 (1월, 2월)과 (9월, 10월)에는 증감이 다른 월보다 크게 나타나 명절효과의 조정이 필요함을 보여주고 있다.

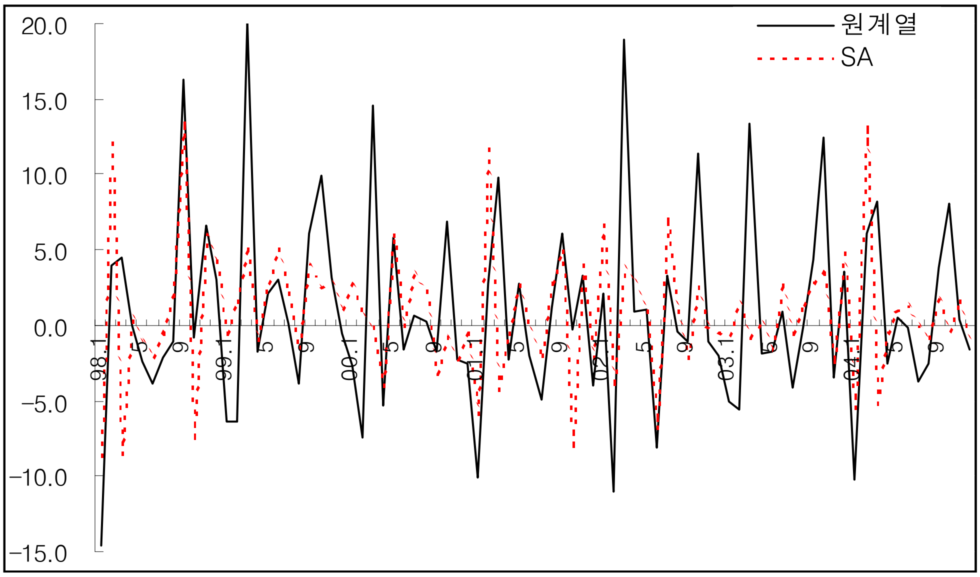
전월비는 전년동기비보다 불규칙요인 등에 의해 영향을 크게 받기 때문에 증감률의 변동성이 크게 나타나는(증감 방향이 자주 바뀌는) 경향이 있어 쉽게 증감 추세를 파악하기 어려운 단점이 있다. 따라서 호주, 캐나다 등에서는 계절조정에 의해 추정되는 추세요인을 함께 사용하고 있다[3].

전월비는 전월에 대비한 단기 경기 동향을 파악하므로 연간 성장을 보

3) 추세요인을 사용하고 있는 관련 통계의 월별 보도 자료는 다음과 같다. 호주: Labor Force, 캐나다: Monthly Survey of Manufacturing, Wholesale trade, Retail trade

기 어려운 단점이 있다. 따라서 해당기의 성장(또는 감소) 추세가 계속 유지된다는 가정 하에 1년 동안의 성장을 나타내는 연율(Annual rate)에 의하여 연간 성장을 구한다.

<그림 1.2> 전월비 증감률



$$\text{연율} = \left(\left(\frac{SA_t}{SA_{t-1}}\right)^s - 1\right) \times 100 \quad (1.1)$$

여기서 s는 월자료인 경우 12, 분기인 경우 4이다.

그러나 연율은 해당기의 증감에 따라 1년간의 성장이 추정되어 당해연도의 누적기간(월, 분기)의 실적은 무시된다. 따라서 당해연도의 실적을 고려한 기하평균을 이용한 연간 성장률(연율)에 의해 1년 동안의 성장을 다음과 같이 구해 볼 수 있다.

j=0,1,2,⋯,(s-1)에 대해서,

$$\text{연간 성장률} = \left\{\left(\prod_{i=0}^{j} \frac{SA_{t-i}}{SA_{t-1-i}}\right)^{s/(j+1)} - 1\right\} \times 100 \quad (1.2)$$

이때 j=0이면, $(SA_t/SA_{t-1})^s - 1$이므로 t기의 연율을 나타내며, j=(s-1)이면 $(SA_t/SA_{t-s}) - 1$이므로 전년대비 증감률을 나타낸다.

계절조정지수에 의해 연율과 연간 성장률을 구해보면, 2003년과 2004

년 12월의 연간 성장률은 전년비와 일치한다. 그러나 연율은 해당 월의 증감에 의해 1년간의 성장률을 구하므로 전년비와 일치하지 않는다. 또한 1년 동안의 증감 추이를 본다면, 연율의 경우 변동성이 매우 큼을 알 수 있으나, 연간 성장률은 변동성이 작고 안정적으로 움직이고 있다.

<월별자료의 연율과 연간 성장률>

	SA_t	K_t	전월비	연 율	전년비	PK_t	연간 성장률
03.12월	122.8	1.0478	4.78	75.22	10.76	1.11	10.76
04. 1월	116.0	0.9453	-5.47	-49.11		0.95	-49.11
2월	131.3	1.1317	13.17	341.29		1.07	49.85
3월	124.6	0.9484	-5.16	-47.01		1.02	5.96
4월	123.9	0.9948	-0.52	-6.07		1.01	2.82
5월	125.1	1.0096	0.96	12.16		1.02	4.62
6월	126.8	1.0133	1.33	17.24		1.03	6.63
7월	126.8	0.9999	-0.01	-0.07		1.03	5.64
8월	125.7	0.9916	-0.84	-9.62		1.02	3.60
9월	127.8	1.0167	1.67	21.98		1.04	5.50
10월	127.2	0.9952	-0.48	-5.56		1.04	4.34
11월	129.2	1.0158	1.58	20.68		1.05	5.73
12월	128.0	0.9909	-0.91	-10.43	4.28	1.04	4.28

주) $K_t = SA_t / SA_{t-1}$, $PK_t = \prod K_{t-i}$

그러나 일반적으로 연간 성장률을 보기 위한 연율은 (식 1.1)에 의해 구하고 있다. 특히, 미국 상무성 경제분석국(BEA: Bureau of Economic Analysis)에서는 계절조정 GDP 통계치를 연율에 의한 계절조정계열 ($= (SA_t)^4$)을 발표함으로써 전분기비가 바로 연율이 되도록 하고 있다.

1.4 경상작업을 위한 계절조정

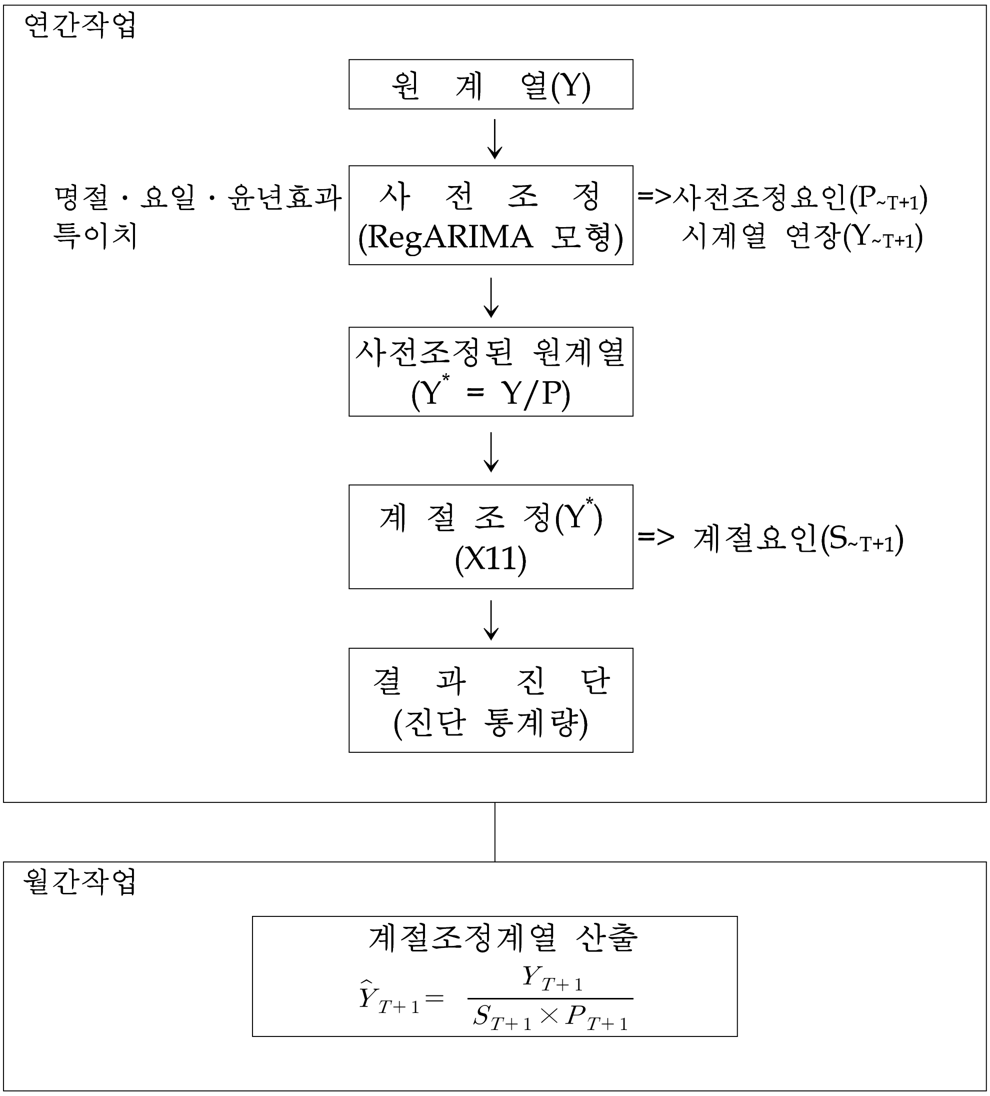
계절조정 결과는 새로운 시계열이 추가 될 때마다 변하므로, 새로운 시계열이 추가될 때마다 계절조정을 재실시하여야 한다. 그러나 시계열이 추가될 때마다 시계열 연장을 위한 ARIMA 모형과 명절효과, 요일효과, 특이치 등 사전조정변수를 고려한 RegARIMA 모형을 매번 재설정하여 계절조정을 실시하는 것은 쉬운 일이 아니다. 이처럼 시계열이 추가될 때

마다 계절조정을 재실시하는 것을 동시조정법이라 한다.

이와 대비되는 개념으로 요인 예측법에 의한 계절조정법이 있다. 요인 예측법은 일정한 기간(대부분 1년 단위)마다 계절조정을 하여 사전에 계절요인과 사전조정요인 등을 예측한 후, 새로 추가된 시계열에 대해서 사전조정요인과 계절요인을 제거하여 계절조정계열을 구하는 방법이다.

X-12-ARIMA에서 시계열 $(Y_1, Y_2, \ldots,\ Y_T)$에 대한 경상작업을 위한 계절조정 수행 절차는 <그림 1.3>과 같다.

<그림 1.3> X-12-ARIMA 계절조정 수행 절차



주) $Y_{\sim T+1} = (Y_1, Y_2, \ldots,\ Y_T,\ Y_{T+1})$

먼저, 스펙트럼 분석 등을 통해 원계열에 계절성이 존재하는지 확인한다. 이때 계절성이 없다면, 계절조정절차를 걸치지 않아도 된다.

계절성이 있다면, 명절・요일・윤년 등 캘런더 효과와 특이치가 존재하는지 확인하고 이에 따라 적절한 사전조정변수를 갖는 RegARIMA 모형을 설정한다(진단 방법은 Ⅲ장 참고). 이때 설정된 RegARIMA 모형의 회귀계수에 의해 사전조정요인(P)이 추정된다.

계절조정을 실시하기 전에 원계열에서 RegARIMA 모형에 의해 추정된 사전조정요인을 제거하며, 이를 사전조정된 원계열이라 한다. 계절조정절차에서는 사전조정된 원계열에 대하여 이동평균법인 X11 방법에 의해 계절조정을 하게 된다(이동평균법은 4.2절 참고). 이때 계절요인(S)이 추정되고 계절조정계열(SA)이 산출된다.

최종 계절조정 결과가 산출되면 X-12-ARIMA에서는 Q-통계량, 스펙트럼분석, 이동기간분석 등 여러 가지 진단 통계량을 제공해 준다. 이러한 진단 통계량을 이용하여 계절조정 결과에 대한 적정성을 판단한다.

계절조정 결과가 적정하다고 판단되면, 추정된 사전조정요인(P)과 계절요인(S)을 이용하여 새로 추가된 시계열에서 이들 요인들을 제거함으로써 계절조정계열을 산출한다.

Ⅱ. X-12-ARIMA의 실행

X-12-ARIMA을 실행하기 위해서는 먼저 입력화일인 *.spc 화일을 먼저 만들어야 하며, 결과는 *.out, 에러 메시지는 *.err라는 화일에 자동으로 저장된다. 이때 Windows의 스크린 상에는 spc 화일과 output 화일만 보여진다.

2.1 INPUT과 OUTPUT

X-12-ARIMA를 실행하기 위해서는 먼저 입력명령문인 spc 화일을 만들어야 한다. spc 화일은 다음과 같은 기본 명령어로 구성된다.

series { 자료의 입력 }
regression { 사전조정 내용 }
automdl { ARIMA 모형의 자동선정 } 또는
arima { ARIMA 모형의 지정 }
x11 { 계절조정 내용 }

이외에 계절조정 결과에 대한 진단을 위한 history 및 slidingspans 등의 명령어를 사용할 수 있다. 이에 대한 상세한 내용은 PART B의 “X-12-ARIMA 프로그램의 사용방법”에서 자세히 다루도록 한다.

실행결과는 *.out에 인쇄된다. 이때 후속 분석을 위해서 계절요인(D10), 계절조정계열(D11), 불규칙요인(D13), 추세순환요인(D12) 등을 저장하기 위한 save 명령을 이용할 수 있다. save 명령에 의해 저장되는 화일은 spc 화일명에 D10 또는 D11 등의 확장자를 갖는다.

output을 다른 화일명으로 저장하고자 하는 경우에는 다음과 같이 DOS 프롬프트상에서 다음과 같이 실행한다.

```
path\x12a   path\filename   path\outname
```

여기서 filename은 filename.spc이며 outname은 output 화일의 이름으로 결과에는 outname.out으로 저장된다. 다음은 DOS 프롬프트상에서의 예이

다. 여기서 X-12-ARIMA를 실행하기 위한 실행화일은 x12a.exe이다.

(예 2.1) 입력 spc 화일은 test.spc, 결과는 test.out으로 출력하고자 하는 경우, 실행방법은 다음과 같다.

x12a test

(예 2.2) 입력 spc 화일은 test.spc, 결과는 result.out으로 출력하고자 하는 경우, 다음과 같이 입력과 출력 화일명을 지정하고 실행한다.

x12a test result

분석하고자 하는 시계열이 하나 이상인 경우, X-12-ARIMA에서는 2가지 방법을 이용하여 실행할 수 있다. 즉, multi-spc mode와 single-spc mode가 있으며, 이들을 실행하기 위해서는 먼저 metafile을 생성해야 한다.

- multi-spc mode: 여러 개의 시계열에 대해서 각각 spc 화일 작성
- single-spc mode: 여러 개의 시계열에 대해서 한 개의 spc 화일 작성

multi-spc의 경우, 각 시계열에 대하여 계절조정을 하기 위한 개별 spc 화일을 작성한 후, 일괄처리를 하기 위한 입력 spc metafile인 *.mta라는 확장자를 갖는 화일을 작성한다.

(예 2.3: multi-spc) 시계열자료 xuu1, xuu2, xuu3에 대해서 각각의 계절조정 옵션을 기술한 spc 화일을 각각 xuu1.spc xuu2.spc xuu3.spc로 작성한 후, 이들 개별 spc 화일을 test.mta에 의해서 일괄 실행한다. 이때 test.mta의 metafile은 *.spc의 확장자 없이 다음과 같이 작성한다.

```
<test.mta>
        xuu1
        xuu2
        xuu3
```

실행은 X-12-ARIMA Windows 버전의 경우, window상에서 test.mta를 더블 click하면 된다. 그러나 DOS 버전은 다음과 같이 한다.

x12a -m test

이때 X-12-ARIMA를 실행하면 다음과 같이 각 *.spc 화일에 대한 output이 생성된다.

xuu1.out, xuu2.out, xuu3.out

Single-spc mode의 경우, 하나의 입력 spc 화일에 의해서 여러 개의 시계열을 계절조정하기 위한 방법이다. Single-spc mode를 이용하기 위해서는 먼저 *.dta라는 확장자를 갖는 data metafile을 생성시켜야 한다. 이때 data는 같은 format 형태로 입력되어 있어야 한다(입력형태에 관한 내용은 PART B: SERIES 명령 참조). data metafile에는 분석하고자 하는 시계열의 data filename과 결과 filename을 쓸 수 있다. 결과 화일이름을 지정하지 않으면 data filename이 결과 화일이름이 된다. data 화일이 현재의 경로(path)에 있지 않으면, 경로명을 반드시 지정해야 한다. data 화일은 data metafile에 나타나는 순서대로 실행된다.

Single-spc mode는 계절조정을 위한 spc 파일을 하나를 작성하여 여러 시계열에 적용하므로 정확한 계절조정보다는 계절성, 사전조정 등의 확인을 위한 초기 검진으로 이용하는 것이 유리하다.

(예 2.4: Single-spc) 시계열 xuu1, xuu2, xuu3를 계절조정한다면, test.dta라는 data metafile의 구성은 다음과 같이 한다.

<test.dta>

xuu1.dat
xuu2.dat
xuu3.dat

X-12-ARIMA의 실행은 Windows 및 DOS 상에서 할 수 있다. Windows에서는 생성한 data metafile인 *.dta를 더블 click하면, spc 파일을 선택하는 스크린이 나오며 이때 spc 파일을 선택하면 된다.

DOS에서는 *.spc, *.dta 등의 확장자 없이 다음과 같이 실행한다.

```
x12a specfile  -d  test
```

여기서 -d는 data metafile이 있음을 나타내며, specfile은 계절조정을 수행하기 위한 입력명령 화일인 specfile.spc 화일을 나타낸다. 결과는 test.dta에서 기술되어 있는 data 파일 이름에 따라 xuu1.out, xuu2.out, xuu3.out으로 생성된다. 결과 화일을 특별히 다른 경로명과 파일명을 지정하고자 하는 경우에는 경로명과 출력 화일명을 다음과 같이 data metafile에서 지정할 수 있다.

```
<test.dta>
      c:\data\xuu1.dat   c:\output\xuu1_out
      c:\data\xuu2.dat   c:\output\xuu2_out
      c:\data\xuu3.dat   c:\output\xuu3_out
```

(예 2.5: Mutli-spc mode의 경우) 1980년 1월부터 2004년 12월까지 광업, 제조업, 전기・가스・수도업의 생산지수를 계절조정하고자 한다. data 화일은 “12f8.0” 형태이며 이름은 각각 i102c.dat, i102d.dat, i102e.dat, 입력 spc 화일을 i102c.spc, i102d.spc, i102e.spc이라고 하자<data file의 입력형태는 Part B: SERIES를 참조>.

・(1단계) 개별 시계열에 대한 계절조정 입력 명령문 작성(예: i012c.spc)

```
series { title="Mineral Production Index"
         name="MPI"
         start=1980.01
         period=12
         span=(1980.01, 2004.12)
         file="i102c.dat"
         format="(12f8.1)"    }
regression {  …………  }
x11 {  …………  }
```

・(2단계) 개별 입력 명령문인 *.spc에 대한 metafile인 test.mta 작성

```
i102c
i102d
i102e
```

· (3단계) 실행

```
x12a -m test
```

실행 결과는 i102c.out, i102d.out, i102e.out에 각각 저장된다.

(예 2.6: Single-spc mode의 경우) 1980년 1월부터 2004년 12월까지 광업, 제조업, 전기·가스·수도업의 생산지수를 계절조정하고자 한다. data 화일이름은 각각 i102c.dat, i102d.dat, i102e.dat이며 “tramo” 형식으로 입력되어 있다. 이때 결과는 temp라는 경로에 out1, out2, out3로 저장하고자 한다<data file의 입력형태는 Part B: SERIES를 참조>.

· (1단계) data metafile 작성: meta.dta

```
i102c.dat c:\temp\out1.dat
i102d.dat c:\temp\out2.dat
i102e.dat c:\temp\out3.dat
```

· (2단계) 입력 spc 화일 작성: test.spc

```
series { title="Industrial Production Index"
        start=1980.01
        period=12
        span=(1980.01 2004.12)
        format="tramo"   }
regression {  …………  }
x11 {  …………  }
```

· (3단계) 실행

```
x12a test -d meta
```

2.2 Log 파일

X-12-ARIMA를 실행할 때마다 log 화일에 시계열과 spc 화일의 모형, 계절조정 진단 통계량 요약 등이 저장된다. X-12-ARIMA가 muli-spc 혹은

single-spc mode을 실행할 때, log 화일은 metafile이나 data metafile과 같은 이름과 경로에 저장된다. 예를 들어 다음과 같은muli-spc mode의 경우,

x12a -m exports

exports.mta에 저장된 spc 화일의 각각이 실행되며, 사용자가 선정한 진단 통계량이 exports.log 화일에 저장된다.

사용자는 **series, composite, transform, x11, x11regression, regression, automdl, estimate, check, slidingspans**와 **history** 명령에서 savelog 옵션에 의해 진단통계량을 저장할 수 있다. 저장할 수 있는 진단 통계량은 <Part B>의 각각의 명령에서 다루기로 한다.

2.3 Flags(표시)

Flags는 DOS 상태에서 X-12-ARIMA를 실행할 때 이용한다.

일반형식: path\x12a arg1 arg2 … argN
여기서 arg1 arg2 … argN은 flag이거나 파일이름이다.

다음은 입력화일이 test1이고 출력화일은 test2인 경우 X-12-ARIMA의 실행방법은 다음과 같다.

```
x12a  test1  test2
x12a  -i  test1  -o  test2
x12a  -o  test2  -i  test1
```

flag의 내용을 좀더 상술하면 다음과 같다.

-n: X-12-ARIMA 결과의 출력 시 표가 많이 출력되기 때문에 출력 화일에 나타나는 표 번호를 지정하기 위해 사용한다.

-s: 계절조정의 가장 중요한 진단결과를 출력 화일(.out)과 분리하여 .xdg 화일로 저장한다.

예1) x12a test -s

출력결과는 test.out과 test.xdg에 저장되며 test.xdg에는 계절조정의 주요한 진단결과가 저장된다.

예2) x12a test -s -o testout

출력결과는 testout.out과 testout.xdg에 저장된다.

-c: 입력 metafile (-m)로 실행되는 총합계절조정 시 사용한다. 총합계절조정에 있어서 X-12-ARIMA는 개별 시계열을 계절조정한다. 입력 명령문에는 각 개별시계열에 대한 내용이 필요하다. -c 옵션을 사용할 때, 개별 시계열에 대한 입력 명령문의 계절조정과 모형옵션은 무시되며, 개별 시계열은 시계열을 구성하기 위해서만 사용된다. 이 옵션은 시계열을 구성하기 위한 REGARIMA모형을 사용할때 유용하다.

X-12-ARIMA에서 사용되는 flag

flag	flag 내용
-i filename	입력 화일의 화일 이름
-o filename	프로그램이 실행되는 동안에 생성되는 모든 출력 화일의 화일 이름
-m filename	입력 metafile의 파일 이름
-d filename	data metafile의 파일 이름
-n	입력 명령문에서 요구한 표만 인쇄
-w	출력 화일의 한 줄이 132자로 인쇄
-p	출력 화일의 page표시를 없앰.
-s	계절조정의 진단결과를 화일 속에 저장
-c	총합 계절조정의 경우, 개별 요인의 합을 계산
-v	입력 명령문의 error check

2.4 일반적인 입력사항

자료의 값은 항상 같은 형태(format)로 입력하며 월(분기)별 자료의 날

자는 “연도.월(연도.분기)”로 표시한다. 연도를 표시할 때 “67”은 1967년으로 인식한다. 월은 1~12의 정수 또는 영문 3자를 사용할 수 있다. (즉, jan feb mar apr may jun jul aug sep oct nov dec). 따라서 1967.12, 67.12, 67.dec와 1967.dec는 같은 의미이다. 날자는 시계열의 시작시점과 분석하고자 하는 시계열의 기간 및 특이치에 대한 회귀변수를 정의할 때 사용한다. 예를 들어 1978년 4월의 가법특이치와 1982년 9월을 시작으로 하는 수준변동에 대한 회귀변수는 다음과 같이 사용한다.

```
regression { variables=(ao1978.apr ls1982.sep) }
```

“ ”안에 있는 문자를 제외한 문자, 숫자, 공란, = . , { } ( ) [ ] + - # 등은 입력문자로서 사용할 수 있다.

{ }는 spc 화일의 옵션을 지정할 때 사용하며, ( )는 옵션에서 여러 개의 목록을 사용할 때, [ ]는 다음 두 가지 형태로 사용할 수 있다. 첫째, 부활절 명절 변수와 같이 특별한 옵션에서 정의되는 값을 지정할 때, 둘째, 결측 시차(missing lags)를 갖는 ARIMA 모형을 사용할 때이다.

```
regression { variable =(td Easter[14])}.
arima { model=(0 1 [1, 3])}
```

주석(comment)은 # 이후에 하도록 하며, 여러 개의 줄을 사용하는 경우에는 각 줄의 첫 번째 column에 #을 해야 한다. 시계열 이름을 나타내는 title은 한 줄에 79자 이내에 다음과 같이 한다.

```
series{ Title=" Index of Wholesale Trade" }
```

spc 화일의 한 줄은 132자로 제한되어 있으며, 132자가 넘는 경우 이후의 문자는 무시된다. 그러나 **file** 옵션에 의해 data를 읽혀들이는 경우에는 이 제한이 적용되지 않는다.

outlier{ }와 같은 null list를 사용할 수 있으며, 이때 적용되는 값은 default 값이 적용된다. 숫자는 지수 형태(즉, 400, 400.0, 4.e+2)를 포함할 수 있으며, free format으로 지정할 수 있다. 정수를 요구하는 경우에는 소수 또는 지수형은 사용할 수 없으며 정수형을 사용하여야 한다.

blank, space, tab, blank line를 분리 연산자로 사용할 수 있다. 여러 개의 옵션 값이 있는 경우, 컴머에 의해서 각 옵션 값을 분리할 수 있다. data=(0, 1, 2, 3, 4, 5) 컴머는 default 값이 대체되는 결측 옵션 값을 나타내기도 한다. span=(1967.4,)의 경우, 1967.4월 이후의 컴머는 두 번째 기간이 결측 옵션 값으로 자료의 마지막 관측치를 취하게 된다.

spc 화일에는 series 명령문이 항상 맨 처음에 있어야 하며, spc 화일 내에서 다른 명령문의 순서는 제약이 없다.

2.5 X-12-ARIMA 프로그램의 한계

X-12-ARIMA에서는 시계열의 최대길이, 모형의 최대 회귀변수 개수 등에 대한 제한을 두고 있다. 그러나 이러한 제한은 FOTRAN source 화일의 model.prn을 수정한 후 다시 compile하면 해결할 수 있다.

현재, X-12-ARIMA/0.3에서 입력시계열의 최대 길이는 600개 이다. 전방과 후방을 예측하여 연장한 시계열의 최대 길이는 60년이다. 모형에 사용되는 회귀변수의 최대 개수는 80개이며, 사용자 정의 회귀변수의 최대 개수는 52개이다. AR과 MA 모수에 대한 최대 시차는 36, ARIMA에 대한 일반 또는 계절 차분의 최대 차수는 3이다.

Ⅲ. X-12-ARIMA의 REGARIMA 모형

계절조정은 비경제적 요인을 원계열에서 분리하는 과정으로 계절조정단계와 사전조정단계로 나눌 수 있다. 계절조정단계는 이동평균을 이용하는 linear filtering 방법(Dagum, 1988)과 시계열이 ARIMA 모형을 따른다는 가정 하에서 시계열에 내재된 요인들을 분해하여 추출하는 신호추출방법(Burman, 1980) 등이 적용된다. 전자의 linear filtering 방법은 X-11, X-11-ARIMA와 X-12-ARIMA 등에 사용되고 있으며, 신호추출방법은 TRAMO/SEATS(Caporello, Maravall 등, 2003)에 사용되고 있다. 사전조정단계는 윤년 및 요일효과, 명절효과 등의 캘린더 효과와 함께 특이치를 계절조정하기 전에 조정해 주는 단계이다. 사전조정단계에서는 계절조정에 의해 추정된 초기 불규칙요인을 이용하는 방법과 원계열에 사전조정변수를 회귀하는 Box-Jenkins의 전이함수(transfer function)를 이용하는 RegARIMA(Bell과 Hillmer, 1983) 모형을 이용하는 방법이 있다.

RegARIMA 모형을 이용하는 방법은 X-12-ARIMA와 TRAMO/SEATS 등에 적용되고 있다. 불규칙요인을 이용하는 방법은 X-11-ARIMA 등에 적용되고 있으며 자세한 내용은 통계청(문권순과 윤은경, 1995)을 참고할 수 있다. 최근 X-12-ARIMA/0.3은 원계열을 이용하는 RegARIMA 모형뿐만 아니라 원계열 대신 불규칙요인에 사전조정변수를 회귀하여 사전조정요인을 추정하는 방법을 소개하고 있다. 이때 사용되는 명령은 X11REGRESSION이다.

RegARIMA 모형을 설정하기 전에 시계열을 시간변화에 따라 어떻게 변하고 있는지를 볼 수 있는 time-plotting을 이용하면 계절성의 형태, 특이치, 시계열의 비정상성에 대한 정보를 얻을 수 있다. 또한 이러한 그림에 의해서 시계열의 변환 가능성이나 시계열의 차분 등을 검토할 수 있다. 그러나 X-12-ARIMA에서는 time-Plotting 기능이 미약하므로 SAS 등의 다른 software를 사용해야 한다.

3.1 REGARIMA 모형

정상성 조건을 만족하는 시계열 z_t에 대한 Box-Jenkins(1976)의 승법계절 ARIMA(AutoRegressive Integrated Moving Average process)은 다음

과 같다.

$$\Phi(B)\Phi(B^s)(1-B)^d(1-B^s)^D z_t = \Theta(B)\Theta(B^s)a_t \qquad (3.1)$$

여기서 B는 후행연산자($Bz_t = z_{t-1}$), s는 계절주기의 기간, $(1-B)^d(1-B^s)^D$의 d는 비계절 차분, D는 계절 차분을 나타낸다. 만일 차분을 하지 않은 (d=D=0) 경우, 식(3.1)은 (z_t - μ)가 된다. 이때 μ=E(z_t)이다. $\phi(B)$는 비계절 자기회귀 AR(p) 모형으로 $\phi(B)=(1-\phi_1 B-...-\phi_p B^p)$ 이며, $\Phi(B^s)$는 계절 AR(P) 모형인 $\Phi(B^s)=(1-\Phi_1 B^s-...-\Phi_P B^{Ps})$ 이다. $\theta(B)$는 비계절이동평균 MA(q) 모형으로 $\theta(B)=(1-\theta_1 B-...-\theta_q B^q)$ 이며, $\Theta(B^s)$는 계절이동평균 MA(Q) 모형인 $\Theta(B^s)=(1-\Theta_1 B^s-...-\Theta_Q B^{Qs})$ 이다. 한편, a_t는 평균이 0, 분산이 σ^2인 i.i.d, 즉 백색잡음과정이다.

RegARIMA 모형은 Bell과 Hillmer(1983)에 의해 제안되어 윤년, 요일, 명절효과 등의 캘린더 효과, 특이치 등 사전조정 요인을 원계열에 회귀함으로써 추정할 뿐 아니라 시계열의 전향예측(forecasting) 및 후향예측(backcasting)을 위한 모형으로 X-12-ARIMA 및 TRAMO/SEATS에 적용되고 있다. RegARIMA 모형은 기본적으로 Box와 Jenkins(1976)의 전이함수(transfer function)를 기본으로 하고 있다.

x_{it}를 윤년, 요일, 명절 등의 캘린더 효과와 특이치 등을 나타내는 회귀변수라면 원계열 y_t에 대해서 다음과 같은 회귀모형을 설정할 수 있다.

$$y_t = \sum_i \beta_i x_{it} + z_t \qquad (3.2)$$

여기서 x_{it}는 y_t와 동시에 관찰되는 회귀변수이며, β_i는 회귀계수이다.
이때 $z_t = y_t - \sum_i \beta_i x_{it}$는 회귀잔차계열이며 ARIMA 모형을 따른다. 전통적인 회귀분석모형에서는 z_t가 독립임을 가정하고 있으나, 시계열 자료의 경우 대개는 자기상관관계를 갖게 되어, 회귀모형의 가정에서와 같이 z_t가 서로 상관이 되어 있지 않다라는 가정은 불명확한 결과를 유도할 수 있다.

X-12-ARIMA에서는 식(3.1)과 (3.2)에 의해서 다음과 같이 ARIMA 모형과 회귀모형을 결합한 RegARIMA 모형을 만들 수 있다.

$$\phi(B)\Phi(B^s)(1-B)^d(1-B^s)^D(y_t-\sum_i \beta_i x_{it}) = \theta(B)\Theta(B^s)a_t \qquad (3.3)$$

식(3.3)의 RegARIMA 모형은 회귀방정식($\sum\beta_i x_{it}$)에 순수한 ARIMA 모형 식(3.1)을 적용하여 일반화하거나, 회귀모형 식(3.2)의 z_t에 ARIMA모형을 적용하여 일반화한 것으로 생각할 수 있다.

RegARIMA 모형은 먼저 회귀효과($z_t = y_t - \sum\beta_i x$)의 평균이 0인 계열을 만들기 위하여 차분을 하며 이를 w_t라면 w_t는 다음과 같은 ARIMA모형을 따른다.

$$\phi(B)\Phi(B^s)w_t = \theta(B)\Theta(B^s)a_t$$

따라서 RegARIMA 모형 식(3.3)을 다음과 같이 표현할 수 있다.

$$(1-B)^d(1-B^s)^D y_t = \sum_i \beta_i(1-B)^d(1-B^s)^D x + w_t \qquad (3.4)$$

여기서 w_t는 정상계열의 ARIMA 모형이다.

RegARIMA 모형 식(3.3)의 회귀계수 x_{it}는 동일시점의 y_t에 영향을 미친다. 시차의 효과는 사용자 정의에 의해서 적절한 시차회귀변수를 X-12-ARIMA 프로그램에서 사용함으로써 구할 수 있다. X-12-ARIMA 프로그램은 RegARIMA의 ARIMA 부분에 두 개이상의 가볍 ARIMA 요인, AR · MA 다항식의 missing 시차, 모형을 추정할 때 사용자 정의의 AR · MA 모수, 추세상수(trend constant) 등을 포함할 수 있다.

3.2 사전조정을 위한 회귀변수

RegARIMA 모형은 윤년 · 요일 · 명절효과 및 특이치 등의 회귀변수를 이용한 회귀모형 설정과 함께 회귀잔차(z_t)에 대한 ARIMA 모형의 설정을 필요로 한다. 회귀모형은 PART B의 regression 명령을 이용하며, ARIMA 모형은 PART B의 arima 또는 automdl 명령을 이용하여 설정할 수 있다.

계절성을 갖는 시계열의 모형화에 사용되는 회귀변수는 X-12-ARIMA

프로그램에서 제공되는 변수를 사용하거나 사용자가 설정하여 사용할 수 있다. X-12-ARIMA에서 제공되는 회귀변수는 <표 3.1>과 같다.

REGARIMA 모형에 의하여 추정된 개별회귀계수에 대한 통계적 유의성을 평가하기 위하여 표준화된 t-통계량이 주어지며, 요일효과(trading-day effect)와 같은 특별한 효과에 대한 회귀계수들의 유의성을 평가하기 위하여 χ^2-통계량이 주어진다.

회귀모형에서 가장 기본적인 회귀계수는 상수항이다. ARIMA 모형에서 차분을 하지 않고 모형 속에 다른 회귀변수가 없다면, 시계열은 평균을 나타내는 절편회귀모형이다. ARIMA 모형이 차분을 포함한다면, X-12-ARIMA는 회귀변수를 사용한다. 이때 대응되는 모수는 모형의 차분차수와 같은 계차를 갖는 추세 다항식을 만들므로 추세상수라고 한다. 예를 들어, 비계절 차분(d>0)을 하였지만 계절 차분(D=0)은 하지 않았다면, 추세상수 변수는 t^d에 비례한다. $0 \le j < d$에 대해서 t^j인 경우, 낮은 차수의 다항항(lower order polynomial terms)은 $(1-B)^d$에 의해서 0으로 차분되어 회귀계수가 추정되지 않으므로 회귀변수 속에 포함되지 않는다. 만일 계절차분 (D>0)이 요구되면, 차분하지 않은 추세상수에 대한 회귀변수의 특성은 추세상수가 (d+D)의 다항식 계차를 따르지만 모형은 더욱 더 복잡해진다. 추세상수가 없는 경우 모형 식(3.3)은 d+D-1의 다항식 계차를 따른다.

<표 3.1> X-12-ARIMA의 회귀변수

회귀효과 및 변수	변수정의
추세상수(Trend Constant) *const*	$(1-B)^{-d}(1-B^s)^{-D} I(t \ge 1)$, 여기서 $I(t \ge 1) = \begin{cases} 1 & t \ge 1 \\ 0 & t < 1 \end{cases}$
고정된 계절성(Fixed Seasonal) *seasonal*	$M_{1,t} = \begin{cases} 1 & 1\text{월} \\ -1 & 12\text{월} \\ 0 & \text{이외} \end{cases}$, $\cdots$, $M_{11,t} = \begin{cases} 1 & 11\text{월} \\ -1 & 12\text{월} \\ 0 & \text{이외} \end{cases}$
sincos[ ]	계절빈도 $w_j = 2\pi j/s$, $1 \leq j \leq s/2$에 대해서 삼각함수의 회귀변수값 $\sin(w_j t)$, $\cos(w_j t)$ 이용 월자료 자료의 경우, sincos[1,2,3,4,5,6]
윤년효과(Leap Year) *lpyear*	2월(1/4분기) 윤년인 경우 0.75, 윤년이 아닌 경우 -0.25, 이외의 월은 1

<표 3.1> X-12-ARIMA의 회귀변수

회귀효과 및 변수	변수정의
월길이(Length of Month) (월 flow) *lom*	$m_t - \overline{m}$, 여기서 $m_t = t$월의 일자 수, $\overline{m} = 30.4375$
분기길이(Length of Quarter) *log*	$q_t - \bar{q}$, 여기서 $q_t = t$ 분기의 일자수, $\bar{q} = 91.3125$
요일효과(Trading Day) (월 또 분기 flow)	요일효과를 추정하기 위한 회귀변수는 6개의 회귀변수와 1개 회귀변수를 이용하는 방법이 있다.
tdnolpyear	6개의 회귀변수 $(D_{1t} - D_{7t}),\ldots,(D_{6t} - D_{7t})$ 이용
td1nolpyear	1개의 회귀변수 $(\sum_{j=1}^{5} D_{jt} - \frac{5}{2}\sum_{j=6}^{7} D_{jt})$ 이용
td	6개의 회귀변수(tdnolpyear)와 윤년효과변수(lpyear)
td1coef	1개의 회귀변수(td1nolpyear)와 윤년효과변수(lpyear)
Stock 요일효과(stock Trading Day) (월 stock) *tdstock[ω]*	$D_{1t} = \begin{cases} 1 & t\text{월의 } \tilde{w}\text{일이 월요일} \\ -1 & t\text{월의 } \tilde{w}\text{일이 일요일} \\ 0 & \text{이외} \end{cases}$ $\vdots$ $D_{6t} = \begin{cases} 1 & t\text{월의 } \tilde{w}\text{일이 토요일} \\ -1 & t\text{월의 } \tilde{w}\text{일이 일요일} \\ 0 & \text{이외} \end{cases}$ 여기서 t월의 길이인 $\tilde{w}$는 ω보다 작다. 월말은 ω=31로 한다. 즉 tdstock[31]
특이치(Additive Outlier) *aodate*	$AO_t^{(t_0)} = \begin{cases} 1 & t = t_0 \\ 0 & t \neq t_0 \end{cases}$ (date는 t_0 시점이다)
수준변화(Level shift) *lsdate*	$LS_t^{(t_0)} = \begin{cases} -1 & t < t_0 \\ 0 & t \geq t_0 \end{cases}$
일시적 변화(Temparary Change) *tcdate*	$TC_t^{(t_0)} = \begin{cases} 0 & \text{for } t < t_0 \\ \alpha^{t-t_0} & \text{for } t \geq t_0 \end{cases}$
임시램프(Temparary Ramp) *rpdateo-date1*	$RP_t^{(t_0, t_1)} = \begin{cases} -1 & t = t_0 \\ \frac{(t-t_0)}{(t_1-t_0)} - 1 & t_0 < t < t_1 \\ 0 & t \geq t_1 \end{cases}$

3.2.1. 고정된 계절효과

월간자료의 고정된 계절효과는 각 월에 대하여 12개의 indicator 변수를 이용하여 모형화할 수 있다. 그러나 이들 12개 변수를 합하면 1이 되기 때문에 전체 수준효과는 알 수 없게 된다. 따라서 12개의 indicator변수를 적절한 재모수화(reparameterization)를 통하여 11개의 대조변수(contrast)로 변환하여 사용한다. 또 다른 재모수화는 고정된 월별의 Fourier 삼각(trigonometric) 시계열 표현으로부터 11개의 변수를 취하는 것이다. 이들 두가지 모수화에 사용되는 변수는 <표 3.1>에 있다.
X-12-ARIMA는 이들 옵션에 따르며, 또한 선택된 빈도(frequencies)에 의한 삼각항(trigonometric term)의 지정에 따르기도 한다.

3.2.2. 요일효과(Trading-day effect)와 윤년효과(Leap year effect)

요일효과는 각 월 또는 다른 연도의 같은 월에 대한 구성 요일의 회수가 다르기 때문에 생기는 효과이다. 이는 월요일에서 일요일까지의 모든 요일에서 동일하게 경제활동이 이루어진다고 볼 수 없기 때문에 어떠한 월에 특정한 요일이 많고 적음에 따라 월의 경제활동은 달라질 수 있다.

특히, 매일 매일의 실적을 측정하여 월(분기)별 실적을 내는 flow 계열이 아닌 해당 월의 말일(월말) 값을 측정하는 stock 계열의 경우 측정된 값이 어떠한 요일이였느냐에 따라 측정값은 민감하게 변할 수 있다. 또한 2월의 월 길이가 매년 일정하지 않음으로써 2월의 계절요인에는 4년마다 반복되는 윤년효과가 포함되어 있다. 따라서 X-12-ARIMA에서는 요일의 회수가 다름으로써 나타나는 flow와 stock 계열의 요일효과 뿐만 아니라 월 길이효과인 윤년효과까지 조정할 수 있도록 했다.

Young(1965)은 추정된 초기 불규칙요인과 월별 요일수를 다중회귀모형으로 설정하여 요일효과를 추정하고자 했다. 한편, Bell과 Hillmer(1983)는 원계열에 월별 요일 회수를 회귀시키는 RegARIMA 모형에 의해서 요일효과를 추정하고자 하였다.

X-12-ARIMA는 요일효과를 6개 또는 1개의 요일효과변수를 다음과 같이 정의하여 Bell과 Hillmer (1983)의 RegARIMA 모형에 적용하고 있다.

6개(tdnolpyear): (월요일 수 - 일요일 수), …, (토요일 수 - 일요일 수)
1개(td1nolpyear): (월~금요일 수) - (토, 일요일 수)×5/2

월 길이 효과를 반영하는 방법은 Y_t를 $\overline{m}Y_t/m_t$로 재조정하는 방법이다. 여기서 Y_t는 변환하기 전의 원계열이며, m_t는 t월의 월 길이(일수), $\overline{m}=30.4375$로 평균 월 길이(일수)이다. 여기서 $\overline{m}$는 윤년을 고려하여 4년간의 평균 월 길이이다.

요일효과를 추정하기 위한 모형을 설정하기 위하여 다음과 같이 정의하자.

α_j : j 요일효과, j=1(월), 2(화), 3(수), 4(목), 5(금), 6(토), 7(일)

D_{jt} : t월의 j 요일 회수

$m_t = \sum_{j=1}^{7} D_{jt}$: t월의 월 길이

이때 t월의 누적요일효과는 $\sum_{j=1}^{7}\alpha_j D_{jt}$이며 $\overline{\alpha}=\sum \alpha_j/7$이다.

t월의 누적효과는 다음과 같이 월 길이효과와 요일별 순효과 $(\alpha_j-\overline{\alpha})$로 분해할 수 있다.

$$\begin{aligned}\sum_{j=1}^{7}\alpha_j D_{jt} &= \overline{\alpha}m_t + \sum_{j=1}^{7}(\alpha_j - \overline{\alpha})D_{jt} \\ &= \overline{\alpha}m_t + \sum_{j=1}^{6}(\alpha_j - \overline{\alpha})(D_{jt} - D_{7t}) \\ &= \overline{\alpha}m_t + \sum_{j=1}^{6}\beta_j(D_{jt} - D_{7t})\end{aligned} \tag{3.5}$$

한편, $m_{t+48} = m_t$이므로

$$\begin{aligned} m_t^* &= \frac{1}{4}\sum_{k=1}^{4} m_{t+12k} \\ &= \begin{cases} 25 & t=2\text{월인 경우} \\ m_t & \text{이외} \end{cases}\end{aligned}$$

따라서 윤년효과변수(LY)와 월 길이효과변수(LOM)를 다음과 같이 승법모형으로 정의하자.

$$LY_t = \frac{m_t^*}{m_t} = \begin{cases} 28.25/28 \approx 1.009 & t = 2\text{월 윤년이 아닌 경우} \\ 28.25/29 \approx 0.974 & t = 2\text{월 윤년인 경우} \\ 1.0 & \end{cases}$$

$$LOM_t = \frac{\overline{m}}{m_t} = \frac{30.4375}{m_t}$$

이때 6개 요일효과를 추정하기 위한 모형은 다음과 같이 설정할 수 있다.

명 령 어	수 식
TD	: $\log(\frac{m_t^*}{m_t} y_t) = \sum_{j=1}^{6} \beta_j (D_{jt} - D_{7t}) + e_t$
TDNOLPYEAR	: $\log y_t = \sum_{j=1}^{6} \beta_j (D_{jt} - D_{7t}) + e_t$
TDNOLPYEAR + LY	: $\log y_t = \beta_0 LY + \sum_{j=1}^{6} \beta_j (D_{jt} - D_{7t}) + e_t$
TDNOLPYEAR + LOM	: $\log y_t = \beta_0 LOM + \sum_{j=1}^{6} \beta_j (D_{jt} - D_{7t}) + e_t$

주) TD는 원계열에 대해서 log 변환을 하는 경우임

위의 식에서 6개 요일효과변수(TDNOLPYEAR) 대신 1개 요일효과변수(TD1NOLPYEAR)를 다음과 같이 설정할 수 있다.

$$\log y_t = \beta(\sum_{j=1}^{5} D_{jt} - \frac{5}{2} \sum_{j=6}^{7} D_{jt}) + e_t$$

또한 요일효과와 윤년효과가 혼합된 TD는 다음과 같은 형태를 갖는 TD1COEF를 사용할 수 있다.

$$\log(\frac{m_t^*}{m_t} y_t) = \beta(\sum_{j=1}^{5} D_{jt} - \frac{5}{2} \sum_{j=6}^{7} D_{7t}) + e_t$$

TD 요일효과변수를 이용하여 승법분해 요일효과모형을 설정하기 위하여 RegARIMA에 대입하면 다음과 같다. 모형의 추정은 (3.4)절에서 설명한다.

$$(1-B)^d (1-B^{12})^D \left\{ \log(\frac{m_t^*}{m_t}) y_t - \sum^{6} \beta_j (D_{jt} - D_{7t}) - \sum_{j=7}^{r} \beta_j x \right\} = w_t \qquad (3.6)$$

여기서 $\sum_{j=7}^{r} \beta_j x$는 요일효과이외의 효과이며, 이외의 효과가 없으면 0이다. 모형 식(3.6)으로부터 요일효과요인을 구하면 다음과 같다.

TD	:	$\frac{m_t^*}{m_t} \exp\left\{ \sum_{j=1}^{6} \beta_j (D_{jt} - D_{7t}) \right\}$
TDNOLPYEAR	:	$\exp\left\{ \sum_{j=1}^{6} \beta_j (D_{jt} - D_{7t}) \right\}$
TDNOLPYEAR + LY	:	$\exp\left\{ \beta_0 LY + \sum_{j=1}^{6} \beta_j (D_{jt} - D_{7t}) \right\}$
TDNOLPYEAR + LOM	:	$\exp\left\{ \beta_0 LOM + \sum_{j=1}^{6} \beta_j (D_{jt} - D_{7t}) \right\}$

여기서 $|x|$가 작으면 $e^x \approx 1 + x$임을 이용하면

$$\exp\left\{ \sum_{j=1}^{6} \beta_j (D_{jt} - D_{7t}) \right\} \approx 1 + \sum \beta_j (D_{jt} - D_{7t})$$가 된다.

가정된 TD, TDNOLPYEAR, TDNOLPYEAR+LY, TDNOLPYEAR+ LOM 중에서 어느 모형을 사용할 것인가는 다음의 Akaike 정보량(Akaike Information Criterion: AIC)이 최소가 되는 모형을 선정한다(Findley, Monsell 등, 1997).

$$AIC_T = -2L_T + 2n_p \tag{3.7}$$

여기서 L_T는 우도함수(Likelihood function), T는 일반 및 계절차분을 한 후의 관측치 수, n_p는 추정된 모수의 수이다. U.S. Census Bureau(2004)는 AIC 통계량은 차분, 특이치 등에 따라 관측치 수가 변동되므로 이를 보완해 줄 수 있는 수정된 $AICC_T$ 통계량을 사용할 것을 권장한다.

$$AICC_T = -2L_T + 2n_p \frac{T}{T - n_p - 1} \tag{3.8}$$

한편 X-12-ARIMA에서는 X-11-ARIMA와 달리 6개 요일회귀변수의 유의성을 검정하기 위하여 요일별 회귀계수에 대한 t-검정 통계량 뿐만 아니

라 χ^2-통계량도 제공해 주고 있다. χ^2-검정을 위한 귀무가설은 “모든 요일의 효과가 같다($H_0 : \beta_1 = \ ... \ = \beta_7$)”이다. 따라서 p-값이 너무 작은 경우 귀무가설을 기각하며, 이때 RegARIMA 모형에 요일효과를 반영할 필요가 있다고 하겠다.

일정한 일자를 기준으로 기준일의 상태를 측정하는 stock 계열의 요일효과는 관찰된 값의 요일에 대한 7개 indicator 변수를 이용하여 모형화할 수 있다. 이들 변수의 합($\sum \alpha_j$)은 항상 1이기 때문에 특이점(singularity) 문제를 야기시키므로 6개의 대조변수를 대신 사용한다.

1≤j≤7에 대해서 다음과 같이 정의하자.

$$I_{jt} = \begin{cases} 1 & t\text{월의 관측치가 } j\text{요일인 경우} \\ 0 & \text{이외} \end{cases}$$

이때 stock 계열의 모형과 요인은 다음과 같다.

$$\text{모형: } \log y_t = \sum_{j=1}^{6} \beta_j (I_{jt} - I_{7t}) + w_t$$

$$\text{요인: } \exp\left\{ \sum_{j=1}^{6} \beta_j (I_{jt} - I_{7t}) \right\}$$

여기서 $\sum_{j=1}^{7} \beta_j = 0$이므로 $\sum_{j=1}^{7} \beta_j I_{jt} = \sum_{j=1}^{6} \beta_j (I_{jt} - I_{7t})$이다.

분기 flow 계열의 경우, X-12-ARIMA의 요일효과 명령은 월별 명령과 동일하다. 그러나 분기자료는 월별자료처럼 요일에 대해서 심하게 변하지 않기 때문에 분기의 요일효과는 월별의 경우보다 상대적으로 작게 나타난다. 월 또는 분기이외의 계절주기를 갖는 flow 계열이나 월별 이외의 stock 계열 자료는 요일효과변수를 사용할 수 없다.

X-12-ARIMA에서는 월말 뿐만 아니라 월초 또는 중순과 같은 월의 중간 일에 관측된 stock 계열에 대해서도 회귀변수 명령을 사용할 수 있다.

X-12-ARIMA에서 td, td1coef는 log 변환 시, 윤년효과를 반영하여 변수변환을 하므로 lom이나 lpyear를 추가로 사용할 수 없다. 요일효과변수인 tdnolpyear와 td1nolpyear 사용 시, 윤년효과 또는 월 길이효과에 대한

반영은 regression 명령이나 transform 명령에 의해서 할 수 있다.

regression 명령에 의해 하는 경우, lom 또는 lpyear에 대한 회귀계수가 추정된다. transform 명령의 adjust=lom(또는 lpyear) 옵션에 의해서 수행이 되는 경우, 변수변환과 함께 월 길이 또는 윤년효과가 조정이 된다. 이때 효과를 조정하는 요인은 결과표의 A2표에서 볼 수 있다.

(예 3.1) 다음은 산업생산지수('85.1월~2004.12월)를 이용한 요일효과 분석의 예이다. 요일효과를 분석하기 위한 회귀변수명은 td, tdnolpyear, td1coef, td1nolpyear를 사용할 수 있다. 예에서는 6개 요일효과변수를 이용하는 tdnolpyear를 사용하는 경우의 X-12-ARIMA 프로그램과 주요결과는 다음과 같다.

```
series{ Title=" MANUF Series"
        start=1980.01
        period=12      decimals=1
        span=(1985.01,2004.12)
        file="c:\data\manuf.txt"
        format="tramo"              }
transform{function=auto}
regression{  variables=tdnolpyear    }
automdl {     }
x11{   }
```

이때 X-12-ARIMA의 추정된 요일효과의 회귀계수와 t-값, χ^2-값, AIC-통계량은 다음과 같다.

<표 3.2>의 개별 회귀계수에 대한 t-값을 보면 유의수준 0.05(임계값: 1.96)에서 유의적이지 않음을 보여주고 있다. 따라서 6개 회귀변수에의해 추정된 요일효과는 의미가 없다하겠다. 이는 추정된 회귀계수를 살펴보아도 같은 결과를 얻게 된다. 회귀계수는 $\beta_j = \alpha_j - \overline{\alpha}$이므로 j요일의 생산활동이 평균적 생산활동에서 얼마나 차이가 있는지를 나타낸다고 볼 수 있다. 즉, 월·토·일요일은 평균적 생산활동보다 적게 생산되며, 화·수·목·금요일에는 많이 생산됨을 알 수 있다. 그러나 이들 계수를 자세히 보면 목요일이 너무 작게(0.0003) 생산에 영향을 미치고 있다. 또한, 토요

일(-0.0067)이 일요일(-0.0007)보다 적게 생산이 되고 있어 일반적인 생각과 차이가 있음을 알 수 있다.

<표 3.2> 회귀분석 결과

변 수	모수추정	표준편차	t-값
Mon	-0.0050	0.00447	-1.11
Tue	0.0025	0.00451	0.56
Wed	0.0071	0.00456	1.55
Thu	0.0003	0.00455	0.07
Fri	0.0024	0.00458	0.52
Sat	-0.0067	0.00455	-1.46
Sun(derived)	-0.0007	0.00444	-0.16
상수	-0.0003	0.00031	-1.03

* For full trading-day and stable seasonal effects, the derived parameter estimate is obtained indirectly as minus the sum of the directly estimated parameters that define the effect.

다음 <표 3.3>은 요일효과를 추정하기 위한 회귀변수인 tdnolpyear, td1nolpyear, td, td1coef 중에 어느 변수가 적절한지 분석하기 위한 표이다. Findley와 Monsell(1997)의 최소 AICC기준에 따라 td1nolpyear와 td1coef 모형을 고려할 수 있다. 이때 두 모형에 의해서 추정된 요일효과 회귀계수는 0.0023으로 t-값은 4.30으로 의미가 있다고 하겠다.

<표 3.3> 요일효과 검정을 통계량

회귀변수	AICC-통계량	χ^2-통계량(p-값)/ 회귀계수(t-값)
tdnolpyear	1045.014	18.93(0.00)
td1nolpyear	1035.701	./0.0023(4.30)
td	1041.544	24.48(0.00)
td1coef	1035.082	./0.0023(4.30)

요일효과는 6개 요일별 효과보다는 주말과 주중에 의한 요일효과가 통계적으로 의미있는 것으로 해석된다. 한편, 원계열을 log변환할 때 윤년효

과도 함께 반영한 td1coef와 윤년효과를 반영하지 않은 td1nolpyear의 결과와는 거의 유사하여 윤년효과는 큰 영향을 미치지 않는 것으로 보인다. 이때 lpyear를 회귀변수로 하여 회귀계수(t-값)를 추정하면 0.0193(1.35)로 유의적이지 않은 결과를 얻는다.

3.2.3. 명절효과(Holiday effect)

일반적으로 경제활동은 크리스마스, 명절(부활절, 설・추석 등) 등 특별한 날에 생산이나 소비(매출)가 증가 또는 감소하는 경향이 있다. 그러나 크리스마스는 매년 12월 25일에 반복해서 나타나므로 계절요인에 포함되나, 서양의 부활절이나 우리 나라의 설・추석과 같은 명절은 연도에 따라 명절 월이 변동하는 이동명절월 효과를 갖는다. 서양의 부활절(Easter day)은 3월 22일부터 4월 25일, 설(Lunar New Year)은 1월 22일부터 2월 20일, 추석(Mid-Autumn Holiday)은 9월 8일부터 10월 8일 사이에 나타난다. 추석의 경우 1900년부터 2050년 사이에 10월에 추석이 있는 경우는 추석의 약 25%를 차지하고 있어 9월에 추석이 있는 경우보다 매우 낮은 빈도를 보여주고 있다. 따라서 시계열의 기간이 짧은 경우는 10월 추석이 특이치로 식별될 수 있으므로 시계열의 기간을 충분히 길게 하여야 한다.

가. 명절효과의 검정

X-12-ARIMA에서는 명절효과가 있는지를 검정하기 위하여 명절효과 변수가 포함된 RegARIMA 모형의 AICC-통계량($AICC_{with}$)과 명절효과 변수가 포함되지 않은 RegARIMA 모형의 AICC-통계량($AICC_{without}$)을 비교하여 AICC-통계량이 작은 RegARIMA 모형을 최적 모형으로 선택한다. 즉, $AICC_{\triangle}$가 음 또는 0일 때, $AICC_{with} - AICC_{without} < AICC_{\triangle}$이면 회귀변수가 포함된 RegARIMA 모형이 선호된다. TRAMO/SEATS는 AICC-통계량 대신 BIC-통계량을 이용한다.

그러나 AICC-통계량에 의한 명절효과의 사전검정은 다음과 같은 단점을 갖고 있다. 우리나라의 명절과 같이 2개 이상의 명절변수 중 어느 한

명절변수가 선호되지 않는 경우, 어느 명절변수가 선호되지 않았는지 판단할 수 없다. 또한 AICC-사전검정은 유의성 검정이 아니므로 어느 정도의 유의수준을 갖는지 알 수 없다. AICC-통계량은 ARIMA모형의 차분과 특이치 등이 같은 모형에서 비교하여야 하는 등 사용상의 제약점이 있다 (U.S. Census Bureau, 2004). 따라서 개별 명절효과 변수들에 대한 유의성 검정 대신 다음과 같은 t-검정 통계량(문권순, 2005)을 사용할 수 있다.

명절효과 변수를 포함하지 않은 RegARIMA 모형의 잔차 a_t는 백색잡음과정으로 평균 0, 분산 σ^2를 갖는 정규분포를 따르며, 이들 잔차에는 RegARIMA 모형에 의해 제거되지 않은 명절효과 성분이 남아 있게 된다. 따라서 명절효과를 검정하기 위하여, 이동명절의 잔차 (a_t, a_{t+1})를 다음과 같이 명절 있는 월의 잔차(a_H)와 명절 없는 월의 잔차(a_{NH})로 분류한 후, 두 집단에 대한 평균차이($\mu_H = \mu_{NH}$) 검정인 t-검정을 실시한다.

$$\begin{aligned} a_H &= \{a_t \mid t \in \text{명절 있는 월}\} \\ a_{NH} &= \{a_t \mid t \in \text{명절 없는 월}\} \end{aligned} \tag{3.9}$$

두 집단의 분산이 동일한 경우, 자유도 $n_H + n_{NH} - 2$를 갖는 두 집단 t-검정 통계량을 이용할 수 있다.

$$t = \frac{\bar{a}_H - \bar{a}_{NH}}{\sqrt{S^2\left(\frac{1}{n_H} + \frac{1}{n_{NH}}\right)}} \tag{3.10}$$

여기서, pooled 분산 $S^2 = \dfrac{(n_H - 1)S_H^2 + (n_{NH} - 1)S_{NH}^2}{(n_H + n_{NH} - 2)}$ 이다. 두 집단의 분산이 동일하지 않은 경우는 Satterthwaite의 자유도(df)를 갖는 t-분포를 이용한다.

$$df = \frac{(S_H^2/n_H + S_{NH}^2/n_{NH})^2}{(S_H^2/(n_H - 1))^2 + (S_{NH}^2/(n_{NH} - 1))^2}$$

그래프적 방법에 의한 명절효과 유무의 검증은 다음 (예 3.2)와 같이

상자수염그림(Box-and-Whisker Plot)을 이용하여 (예 3.2)와 같이 분석할 수 있다. 상자수염그림을 작성하기 위하여 설·추석명절인 (1월, 2월), (9월, 10월)의 RegARIMA 잔차를 (식 3.9)와 같이 명절 있는 월과 명절 없는 월로 분류한다. 분류된 RegARIMA 잔차와 이외 월의 RegARIMA 잔차에 대한 각 군별 평균과 4사분위수를 구한 후 상자수염그림을 그린다.

(예 3.2) 다음은 (예 3.1)과 같이 산업생산지수('85.1월~2004.12월)의 경우, 명절효과가 존재하는지 보기 위하여 RegARIMA 잔차의 t-검정(문권순, 2005)을 실시하고 상자수염그림을 분석한 예이다. RegARIMA 모형에는 특이치만 적용하였다.

이때 <표 3.4>를 보면 설의 경우, 명절 있는 월의 잔차 평균(표준편차)은 -0.037(0.0304), 명절 없는 월의 평균(표준편차)은 0.0306(0.0232)로 t-검정 통계량은 -7.01로 매우 커 명절 있는 월과 명절 없는 월간 평균 차이가 있다고 하겠다. 추석역시 t-검정 통계량이 -4.66로 크게 나타나 월간 평균 차이가 있다고 하겠다. 따라서 설과 추석효과가 있다고 생각할 수 있다.

<표 3.4> 명절있는 월과 없는 월의 t-검정: 명절효과 미반영

	명절있는 월	명절없는 월	t-검정(p-값)
설	-0.037(0.0304)	0.0306(0.0232)	-7.01(0.00)
추석	-0.015(0.0306)	0.0343(0.0361)	-4.66(0.00)

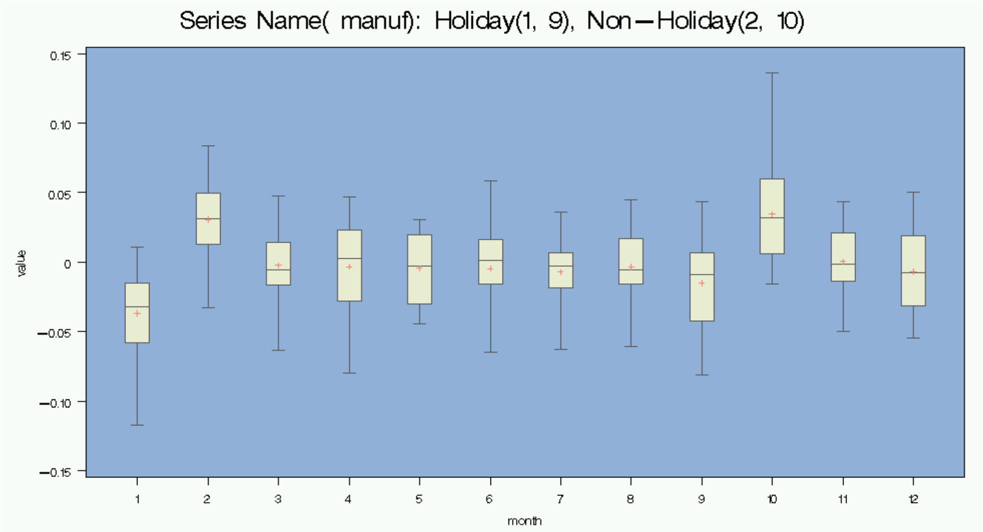
주) 평균(표준편차)

<그림 3.1>은 명절효과를 반영하지 않은 RegARIMA 모형의 잔차에 대한 상자수염그림으로 명절 있는 월(1, 9)과 명절 없는 월(2, 10)로 분류하여 그린 그림이다. 설과 추석 명절이 있는 월인 1과 9의 평균("+")은 다른 월보다 낮은 수준을 나타내고 있으나, 명절 없는 월인 2와 10의 평균은 다른 월보다 높게 나타나고 있어 명절효과가 있음을 보여주고 있다.

<표 3.5>는 명절효과 길이가 4인 Bell-Hillmer의 명절효과변수를 RegARIMA 모형에 적용하여 명절효과를 제거한 RegARIMA 모형의 잔차 평균(표준편차)이다. 이때 설과 추석에 대한 t-검정(p-값) 각 -2.08(0.05), -2.08(0.04)로 평균간 차이가 있다고 하기 어렵다. 한편 상자수염그림 (그림 3.2)를 보면, 명절 있는 월(1, 9)과 명절 없는 월(2, 10)간의 차이가 크

게 줄어들었음을 알 수 있다.

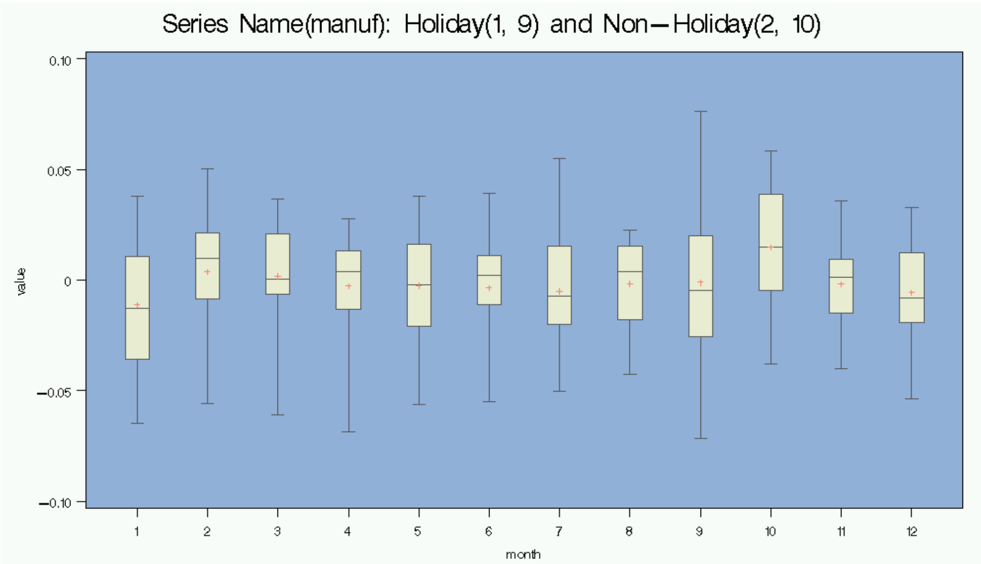
<그림 3.1> 제조업생산지수의 RegARIMA 잔차(명절효과 미반영)



<표 3.5> 명절있는 월과 없는 월의 t-검정: 명절효과 반영

	명절있는 월	명절없는 월	t-검정
설	-0.012(0.028)	0.0063(0.0287)	2.04(0.05)
추석	-0.004(0.0383)	0.0188(0.0291)	-2.08(0.04)

<그림 3.2> 제조업생산지수의 RegARIMA 잔차(명절효과 반영)



나. 명절효과의 추정

월별 flow 계열의 명절효과(holiday effect)는 다음 경우에 발생한다고 할 수 있다.

- 명절을 전후하여 규칙적으로 시계열이 증가 또는 감소하는 경우
- 명절이 있는 월(2개월 또는 그 이상)이 다른 월의 계열과는 다른 형태를 보이는 경우

명절 있는 월(명절월)이 (1월, 2월), (9월, 10월)로 바뀌면서 나타나는 설과 추석의 경우, 명절이 어느 월에 있느냐에 따라 생산, 출하, 판매, 소비 등에 크게 영향을 미칠 수 있다. X-12-ARIMA에서는 부활절 효과를 조정해 주기 위하여 회귀변수 easter[w]를 이용하여 명절효과를 선택적으로 반영하도록 하고 있다<표 3.1>. 노동절(9월 첫 번째 월요일)도 명절효과로 영향을 미친다고 볼 수 있으나 부활절보다는 덜 일반적이다. 노동절 효과를 추정하기 위한 회귀변수는 labor[w]이다. 한편 추수감사절(11월 네 번째 목요일)을 추정하기 위한 변수는 thank[w]이다.

대부분의 명절효과는 명절이 발생한 날짜 이전에 생산, 출하, 판매 등을 앞당기므로 명절의 영향은 명절일 이전에 나타난다. 따라서 명절이 (t+1)월 초에 있다면 명절의 영향은 t월말에도 미치게 된다. 이러한 영향을 명절효과 기간(ω)이라고 한다.

명절효과 요인의 추정방법은 시계열의 이용 측면에서 불규칙요인을 이용하는 방법과 원계열을 이용하는 방법으로 나눌 수 있다. 불규칙요인을 이용하는 방법은 Pfefferman과 Fisher(1981)와 X-11-ARIMA/88에 적용된 Dagum(1988, 1992) 방법 등이 있다(문권순 · 윤은경, 1995 참조).

X-11-ARIMA/88에서는 부활절 요인 E_i를 이동명절월인 3월과 4월의 · 초기 불규칙요인(IC_{ij}) 차이를 이용하여 추정하였다. ω를 부활절이 영향을 미치는 기간인 부활절효과 기간이라 하면, i=1, 2, ⋯, T년에 대해서, i년 j월의 부활절(LE_i)에 대한 가중치 step function $f(LE_i)$를 다음과 같이 정의할 수 있다. 여기서 LE_i는 i년의 부활절 일자와 최초의 부활절(3월 22일)간의 차이 일수이다. 예를 들어, 3월 31일이면, LE_i=10이다.

$$f(LE_i) = \begin{cases} 1 & LE_i \le 10\text{인 경우} \\ \dfrac{\omega + 10 - LE_i}{\omega} & 10 < LE_i < 10 + \omega\text{인 경우} \\ 0 & LE_i \ge 10 + \omega\text{인 경우} \end{cases}$$

이때 j를 이동명절 월(3월, 4월)이라 하면, i년의 부활절 요인 E_i는 다음과 같이 추정할 수 있다.

$$E_i = \frac{1}{2} f(LE_i) \left[\frac{\sum_{i \in M} (IC_{i,j+1} - IC_{i,j})}{n_M} - \frac{\sum_{i \in A} (IC_{i,j+1} - IC_{i,j})}{n_A} \right]$$

명절효과 추정 시, 원계열을 이용하는 방법은 Bell과 Hillmer (1983)의 RegARIMA 모형을 이용하는 방법으로 X-12-ARIMA와 TRAMO/SEATS에서 적용하고 있다. Bell과 Hillmer의 RegARIMA 모형은 부활절 일자와 부활절이 영향을 미치는 부활절효과 기간(ω)에 의해 결정되는 식(3.11)의 부활절효과 변수 $H(\omega, t)$를 식(3.3)의 RegARIMA 모형에 적용(원계열에 회귀)하여 부활절 요인 E_i를 추정한다.

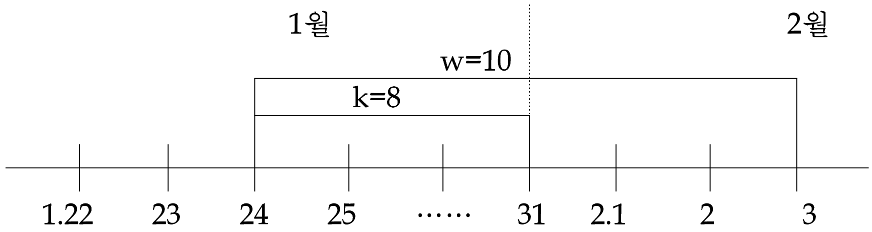
X-12-ARIMA에서는 다음 식(3.11)과 같이 Bell과 Hillmer의 명절효과 변수 $H(\omega, t)$를 RegARIMA 모형에 적용하고 있다.

$$\begin{aligned} H(\omega,t) &= \frac{1}{\omega} \sum_{d=1}^{\omega} h(d,t) \\ &= \frac{1}{\omega} [t\text{월이 명절효과 기간}(\omega)\text{내에 있는 일수}] \\ &= \frac{k}{\omega} \end{aligned} \tag{3.10}$$

여기서

$$h(d,t) = \begin{cases} 1 & t\text{월 } d\text{일이 명절효과 기간}(\omega)\text{내에 있는 경우} \\ 0 & \text{이외} \end{cases}$$

설을 예로 들면, ω=10이고 설이 2월 3일이라면, 1월에는 8일간 명절에 의해서 영향을 받는다. 따라서 H(ω, 1) = 8/10, H(ω, 2) = 2/10이 된다.



d_h를 명절 일자라고 하면, Findley와 Monsell(1996)은 명절 일과 명절 영향 기간에 따라 다음과 같이 H(ω,t)를 정의한다.

· 명절이 (t-1)월에 있는 경우

$$H(\omega,t-1) = 1, \qquad H(\omega,t) = 0$$

· 명절이 t월, $d_h > \omega$이면

$$H(\omega,t-1) = 0, \qquad H(\omega,t) = 1$$

· 명절이 t월, $d_h \leq \omega$이면

$$H(\omega,t-1) = \frac{\omega-k}{\omega}, \qquad H(\omega,t) = \frac{k}{\omega}$$

X-12-ARIMA에서 사용된 RegARIMA 모형의 회귀변수는 다음과 같이 정의한다.

$$\hat{H}(\omega,t) = H(\omega,t) - E[H(\omega,t)]$$

이때 명절요인은 다음과 같이 추정된다.

$$\text{REGARIMA 명절요인} = e^{\hat{\beta}\hat{H}(\omega,t)}$$

명절효과 기간인 ω는 AIC 또는 AICC-통계량이 최소가 되는 값을 선정한다.

(예 3.2) 다음은 Bell과 Hillmer의 식(3.11)에 의해 설과 추석효과에 대한 회귀변수의 설정 예이다. 설이 2월에 있고 설 일자 d_s<ω인 경우(11<LS≤ω+11), 설 효과가 1월과 2월에 영향을 미치게 되므로, 설 명절효과 변수 H_S(ω, t)는 다음과 같이 1월과 2월에 나누어 설정할 수 있다.

$$H_S(\omega,1) = \frac{\omega + 11 - LS}{\omega} \qquad H_S(\omega,2) = \frac{LS - 11}{\omega}$$

추석의 경우, 추석이 9월에 있고 $d_c < \omega$인 경우, 추석 효과는 8월과 9월에 영향을 미친다.

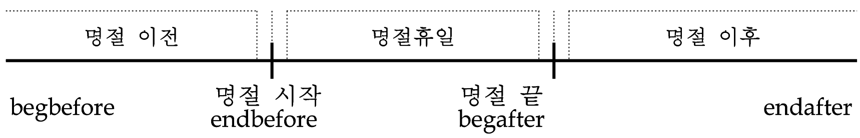
$$H_C(\omega, 9) = \frac{\omega + 24 - LC}{\omega} \qquad H_C(\omega, 10) = \frac{LC - 24}{\omega}$$

이때 설과 추석효과 변수를 갖는 회귀모형과 RegARIMA 모형은 다음과 같다.

$$y_t = \beta_s H_S(\omega, t) + \beta_c H_c(\omega, t) + z_t$$

$$\psi(B)\Phi(B^s)(1-B)^d(1-B^s)^D(y_t - \sum \beta_k H_k(\omega, t)) = \Theta(B)\Theta(B^s)a_t$$

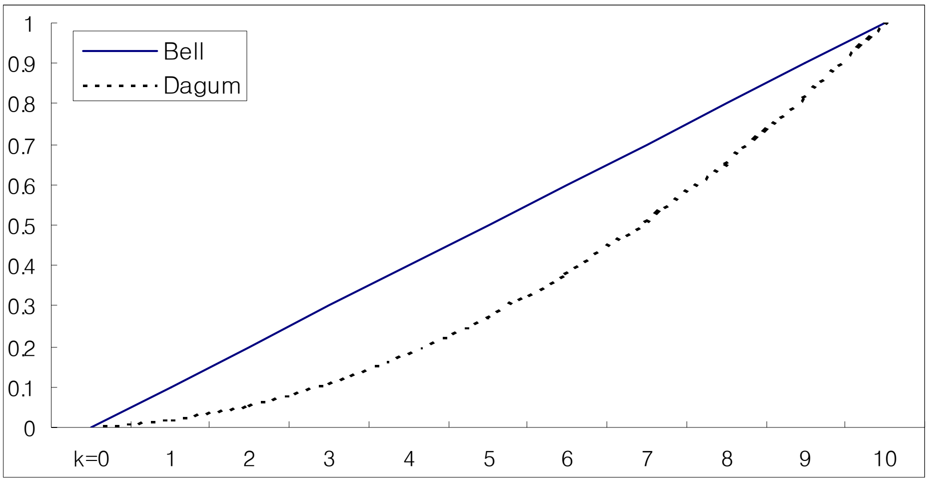
Bell과 Hillmer(1983)는 명절효과가 명절 이전에 영향을 미친다고 가정하였으나, Lin과 Liu(2002)는 명절효과는 명절 전・후 뿐만 아니라 명절휴일 기간에도 영향을 미친다고 가정하고 명절효과 기간을 명절이전 기간, 명절기간, 명절이후 기간 등 3구간으로 구분하여 Bell과 Hillmer의 명절효과 변수를 설정하여 명절효과 요인을 추정하였다. Lin과 Liu의 명절효과 변수 생성 프로그램인 genhol.exe은 "www.census.gov/srd/www/x12a/x12down_pc.html"에서 down load를 할 수 있다.



이긍희(1998), 조신섭・김기웅・남승민(1999)은 명절효과가 설과 추석 전・후에 영향을 미친다고 가정하였으며, 전백근(2002)은 명절길이의 누적길이에 비례($\sum_{j \le i} j / \sum_{i=1}^{\omega} i$)하는 명절효과 변수를 RegARIMA 모형에 설정하여 명절효과 요인을 추정하고 있다(전백근(2006)은 Dagum 방법으로 표현하고 있다).

다음은 명절효과 기간(ω)이 10일이라고 가정하는 경우, Bell과 Hillmer, Dagum 변수를 사용했을 때 명절효과 변수 H(ω,t)의 그림이다. k는 명절효과 기간 내에 t월이 있는 일수이다.

<그림 3.3> ω=10인 경우, Bell과 Hillmer, Dagum의 명절효과 분포



명절효과를 추정하기 위한 X-12-ARIMA에 내장된 회귀변수는 서양에서 적용되고 있는 태양력의 부활절 easter[w]가 있다. 그러나 우리 나라와 같이 태음력의 명절인 설과 추석효과를 추정하기 위하여 내장된 회귀변수는 없다. 따라서 우리 나라의 설과 추석 명절효과를 반영하기 위해서는 X-12-ARIM를 실행하기 전에 명절일자 분포에 따라 명절효과를 반영할 수 있는 회귀변수를 생성해야 한다. Findley와 Monsell 방법에 의한 명절효과변수 값 생성 프로그램은 <참고 3.2>를 참고하기 바란다. 프로그램에 사용된 $E[H(\omega,t)]$는 1900년부터 2050년까지 $H(\omega,t)$의 기대값이다.

(예 3.3) 다음은 제조업생산지수에 대하여 설과 추석 명절요인을 조정하기 위한 계절조정 예이다. 명절효과는 Findley와 Monsell의 방법에 따라 명절일자분포를 사용자 정의의 명절효과 회귀변수로 설정하였다. 명절효과 기간 ω는 4라고 가정하고 (예 3.2)의 설과 추석 명절효과 변수 설정식과 <참고 3.2>의 SAS 프로그램을 이용하여 회귀변수 값을 생성하면 <표 3.6>과 같으며, 이를 "kh4.dat"라고 하자. "kh4.dat" 형태는 다음과 같다.

1980	1	−0.43229	0.00000
1980	2	0.43229	0.00000
1980	3	0.00000	0.00000
…	…	…	…
1980	8	0.00000	0.00000
1980	9	0.00000	0.16042

```
1980  10    0.00000  -0.16042
1980  11    0.00000   0.00000
1980  12    0.00000   0.00000
1981   1   -0.43229   0.00000
1981   2    0.43229   0.00000
1981   3    0.00000   0.00000
   …   …      …    …
```

회귀변수 값 생성 프로그램에 의해서 작성된 kh4.dat 파일의 값을 설과 추석 명절 월인 (1월, 2월) (9월, 10월)로 분리하여 정리하면 <표 3.6>과 같다. 표를 보면, 명절일자 분포에 따라 회귀변수 값도 변동하고 있음을 알 수 있다. 회귀변수 값의 분포는 명절효과 변수의 생성 방법과 명절효과 기간(ω)에 따라 <그림 3.3>과 같이 다른 형태를 띠게 된다.

<표 3.6> 생성된 명절회귀변수 값

	설			추 석		
	명절 분포	회귀변수 값 1월	회귀변수 값 2월	명절 분포	회귀변수 값 9월	회귀변수 값 10월
1985	30 (2.20)	-0.43229	0.43229	22 (9.29)	0.16042	-0.16042
1986	19 (2. 9)	-0.43229	0.43229	11 (9.18)	0.16042	-0.16042
1987	8 (1.29)	0.56771	-0.56771	30 (10. 7)	-0.83958	0.83958
1988	28 (2.18)	-0.43229	0.43229	18 (9.25)	0.16042	-0.16042
1989	16 (2. 6)	-0.43229	0.43229	7 (9.14)	0.16042	-0.16042
1990	6 (1.27)	0.56771	-0.56771	26 (10. 3)	-0.33958	0.33958
1991	25 (2.15)	-0.43229	0.43229	15 (9.22)	0.16042	-0.16042
1992	14 (2. 4)	-0.18229	0.18229	4 (9.11)	0.16042	-0.16042
1993	2 (1.23)	0.56771	-0.56771	23 (9.30)	0.16042	-0.16042
1994	20 (2.10)	-0.43229	0.43229	13 (9.20)	0.16042	-0.16042
1995	10 (1.31)	0.56771	-0.56771	2 (9. 9)	0.16042	-0.16042
1996	29 (2.19)	-0.43229	0.43229	20 (9.27)	0.16042	-0.16042
1997	18 (2. 8)	-0.43229	0.43229	9 (9.16)	0.16042	-0.16042
1998	7 (1.28)	0.56771	-0.56771	28 (10. 5)	-0.83958	0.83958
1999	26 (2.16)	-0.43229	0.43229	17 (9.24)	0.16042	-0.16042
2000	15 (2. 5)	-0.43229	0.43229	5 (9.12)	0.16042	-0.16042
2001	3 (1.24)	0.56771	-0.56771	24 (10. 1)	0.16042	-0.16042
2002	22 (2.12)	-0.43229	0.43229	14 (9.21)	0.16042	-0.16042
2003	11 (2. 1)	0.56771	-0.56771	4 (9.11)	0.16042	-0.16042
2004	1 (1.22)	0.56771	-0.56771	21 (9.28)	0.16042	-0.16042

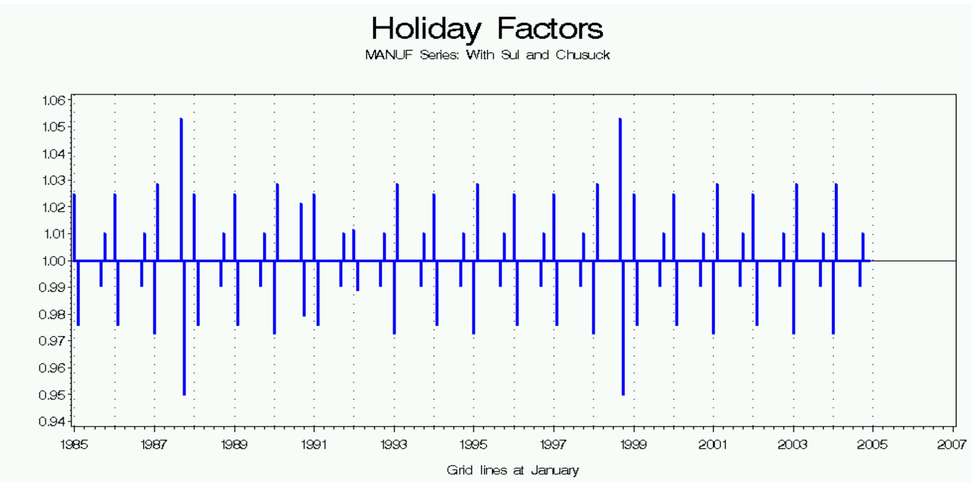
이때 X-12-ARIMA 프로그램은 다음과 같다.

```
series {    ……       }
transform{ function=auto}
regression{ user=(sul chu)
            file="c:\holiday\korh\kh4.dat"
            format="datevalue"
            start=1980.1
            usertype=holiday    }
automdl { }
x11{save=(d10 d11 d12)    }
```

<표 3.6>의 회귀변수 값을 RegARIMA 모형에 적합한 결과 ARIMA 모형은 (0 1 1)(0 1 1)이 선정되었다. 이때 χ^2(p-값)은 104.67(0.00)이며 AICC와 BIC는 975.58과 992.43이다. 명절효과에 대한 추정된 회귀계수와 통계량은 다음과 같다.

변 수	회귀계수	표준편차	t-값
설	-0.0557	0.00677	-8.24
추 석	-0.0618	0.01044	-6.05

<그림 3.4> 설과 추석효과 요인

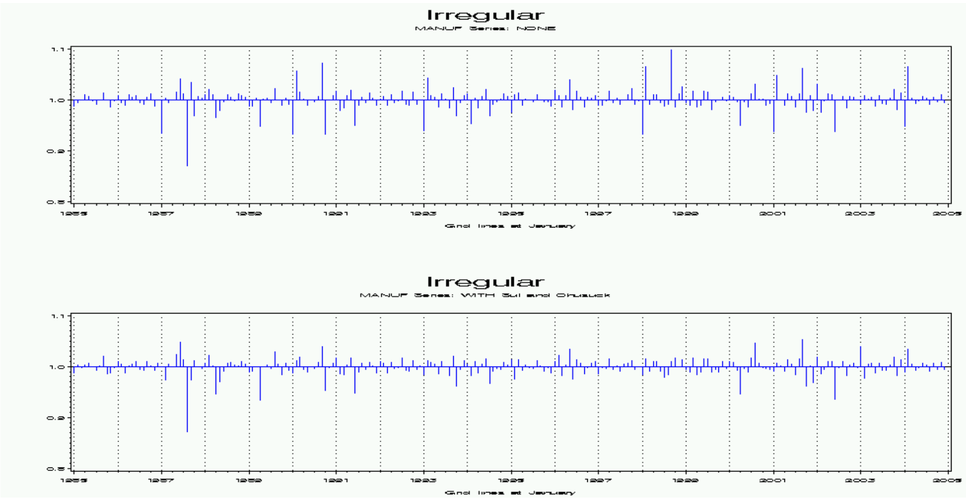


<그림 3.4>는 추정된 명절효과요인이다. 추석의 경우 대부분(약 75%) 9

월에 있으며, 드물게 나타나는 10월 추석은 10월 초에 발생한다. 자료에서 10월 추석은 1987년, 1998년에 각각 10월 7일과 10월 5일에 있어 명절효과요인이 다른 연도의 명절효과요인보다 크게 나타나고 있다.

<그림 3.5>는 명절효과를 반영하지 않은 모형(NONE: 그림 위)과 명절효과를 반영한 모형(WITH: 그림 아래)의 불규칙요인이다. NONE 그림을 보면, 거의 매년 1월과 2월의 불규칙요인이 크거나 작게 나타나고 있어 설 효과가 있음을 나타낸다. 또한 90년, 98년, 20001년의 9월과 10월의 불규칙요인 역시 크게 나타나고 있어 추석효과를 반영하고 있음을 보여주고 있다. 이때 명절효과를 제거한 후의 불규칙요인 크기가 줄어들었음을 보여주고 있다.

<그림 3.5> 명절효과를 반영하지 않은 모형과 반영한 모형의 불규칙요인



다음은 명절효과를 반영하지 않은 모형의 불규칙요인(NONE)과 명절효과를 반영한 모형(WITH)의 불규칙요인 변동에 대한 기술통계이다. 명절효과를 반영하지 않은 모형의 불규칙요인 표준편차는 2.30이나 명절효과를 반영한 모형의 경우에는 1.81로 줄어들어 불규칙요인의 변동성이 감소했음을 알 수 있다.

	평균	표준편차	범위(최대, 최소)
NONE	99.90	2.30	22.79 (109.79, 87.00)
WITH	99.85	1.81	18.12 (105.26, 87.14)

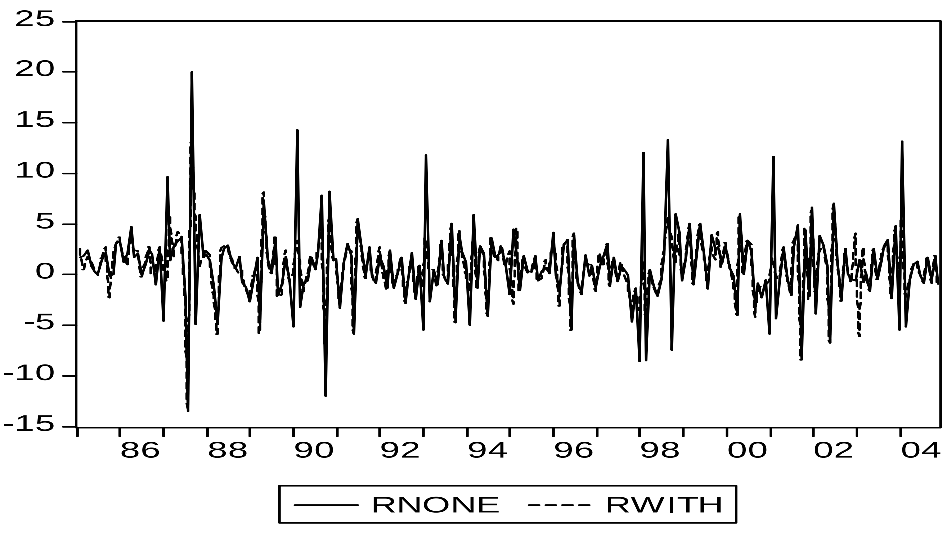
명절 월별로 불규칙요인을 보면(<표 3.7>), 명절효과를 제거한 WITH 모형의 표준편차와 범위가 명절효과를 제거하지 않은 NONE 모형보다 작게 나타나고 있어 명절효과를 제거한 모형이 보다 안정적임을 보여주고 있다.

<표 3.7> 명절효과를 반영하지 않은 모형과 반영한 모형의 불규칙요인

월		평균값	표준편차	최소값	최대값	범 위
1월	NONE	98.29	3.32	93.15	103.11	9.96
	WITH	99.98	1.12	98.47	101.78	3.31
2월	NONE	100.98	3.09	95.30	106.51	11.21
	WITH	100.03	1.72	96.67	103.64	6.97
9월	NONE	101.33	3.04	98.34	109.79	11.45
	WITH	100.53	1.83	97.30	105.26	7.96
10월	NONE	99.60	2.15	93.17	102.12	8.95
	WITH	99.45	1.82	95.30	101.64	6.25

이때 <그림 3.6>은 명절효과를 제거한 계절조정계열과 제거하지 않은 계절조정계열의 전월비 그림이다. RNONE은 명절효과를 제거하지 않은 계절조정계열의 전월비이며, RWITH는 명절효과를 제거한 계절조정계열의 전월비이다.

<그림 3.6> 계절조정계열의 전월비

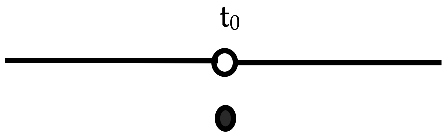


3.2.4. 특이치 효과

특이치(Outlier)란 파업, 이상기후, 정책변화 또는 측정오차 등에 의해서 대부분의 관측치와 다른 측정값을 갖는 소수의 관측치를 말한다. 특이치의 종류는 가법특이치(Additive Outliers: AO), 수준변화(Level Shifts: LS), 임시적 변화 특이치(Temporary Change: TC) 및 램프(Ramp: RP)로 분류할 수 있다.

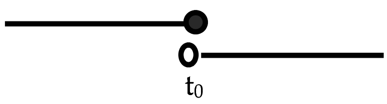
가법특이치(AO)는 임의의 t_0시점에서 전체 계열과는 다른 형태를 나타내는 경우이다. 예를 들어 파업과 같이 어느 한 시점에서 단기적으로 일어나는 현상을 반영하는 경우이다.

$$AO_t^{(t_0)} = \begin{cases} -1 & \text{for} \quad t = t_0 \\ \ \ 0 & \text{for} \quad t \neq t_0 \end{cases}$$



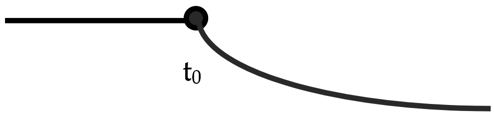
수준변화(LS)는 임의의 t_0시점 이후의 시계열 수준이 이전 시점의 수준과 다른 경우를 말한다. 예를 들어 갑작스런 정책변화나 구조변화로 예전의 수준과는 다른 수준을 갖는 시계열이 계속 관측되는 경우이다. X-12-ARIMA에서는 수준변화가 있는 시계열의 경우에도 계절조정이 가능하다. 즉, 최근의 시계열을 이용하여 ARIMA 모형을 설정한 후, 수준이 다른 과거의 시계열을 예측하여 시계열의 수준을 일정하게 만든 후 계절조정을 한다.

$$LS_t^{(t_0)} = \begin{cases} \ \ 0 & \text{for} \quad t < t_0 \\ -1 & \text{for} \quad t \geq t_0 \end{cases}$$



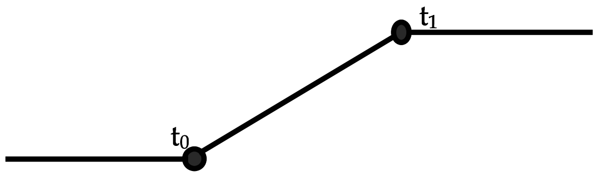
일시적 변화 특이치(TC)는 임의의 어느 시점에 AO 처럼 크게 나타났다가 그 이후에 지수적으로 영향이 감소 또는 증가하는 형태이다.

$$TC_t^{(t_0)} = \begin{cases} 0 & \text{for} \quad t < t_0 \\ \alpha^{t-t_0} & \text{for} \quad t \geq t_0 \end{cases}$$



램프(Ramp: RP)는 시계열의 수준이 점진적으로 변하여 LS 처럼 수준이 수 개월동안 영향을 받는 형태이다.

$$RP_t^{(t_0, t_1)} = \begin{cases} -1 & \text{for } t \leq t_0 \\ \dfrac{(t - t_0)}{(t_1 - t_0)} - 1 & \text{for } t_0 < t < t_1 \\ 0 & \text{for } t \geq t_1 \end{cases}$$



AO는 한 개의 시계열 관측치에 영향을 미치며, LS는 어떠한 시점 이후 일정한 시점만큼 증가시키거나 감소시킨다. TC는 일정한 시점이 후 지수적으로 감소시키거나 증가시키는 형태이며, Ramp는 지정한 기간 동안에 대해서 시계열의 수준을 선형적 증가나 감소를 시킨다.

이들 효과들을 모형화하기 위한 X-12-ARIMA의 회귀변수들은 <표 3.1>에 주어졌다. 그러나 X-12-ARIMA에서는 AO, LS, TC에 대한 특이치가 자동선정 옵션에 의해서 추정되며, 사용자가 RP의 옵션을 사용하드라도 자동 특이치 선정은 AO, LS, TC에 대한 특이치를 식별한다.

시계열 진단의 중요한 관점은 특이치 검색이며, X-12-ARIMA의 outlier 명령은 가법특이치(AO), 수준변화(LS) 및 임시적 변화 특이치(TC)를 자동으로 탐색할 수 있도록 하였다. 이러한 특이치의 탐색은 기본적으로 Chang과 Tiao(1983)의 방법을 이용하였으며, Bell(1983), Otto과 Bell (1990)의 방법에 따라 개선을 하였다. 특이치 검색을 위한 일반적인 검색방법은 GLS의 단계별(stepwise) 회귀방법과 유사하다. 여기서, 특이치 검색을 실시하려는 후보 회귀변수는 모든 시점에 대한 AO와 LS, TC의 변

수이므로, 전체 관측치가 T이라면 AO와 LS, TC를 검색하기 위해서는 3T의 관측치가 필요하다. 그러나 실제로는 3T보다 작은 관측치 수를 이용하여 특이치를 검색한다.

매 시점에 대해서 각 특이치 형태에 대한 t-통계량을 계산하고, 이를 통해 유의적인 특이치를 찾기 위한 t-검정을 실시한다. 이때, 관측치가 특이치로서 유의적인 결과를 갖는 경우 RegARIMA 모형에 AO나 LS, TC 회귀변수를 포함한다. X-12-ARIMA는 2가지의 특이치 검색 방법을 채택하고 있다. addone 방법은 하나의 특이치를 모형에 첨가한 후 전체모형을 재추정하는 방법이다. 한편, addall 방법은 검색된 특이치 모두를 모형에 첨가한 후 모형을 재추정하는 방법이다(PART B의 outlier 명령 참조).

다음은 Findley, Monsell, Bell, Otto와 Chen(1998)의 AO와 LS 특이치 검색을 위한 추가법(Forward Addition)과 제거법(Backward Deletion)이다. X-12-ARIMA에서는 addone은 추가법, addall은 제거법이 적용된다. T_t를 AO, LS 회귀계수에 대한 t-통계량, γ를 임계치, β를 특이치 회귀계수 벡터, (ϕ, θ)를 ARIMA모형의 계수 벡터라고 하자. $a_t(\beta, \phi, \theta)$는 (β, ϕ, θ) 계수벡터를 갖는 RegARIMA(식 3.3)의 잔차 추정치이다. AO 회귀변수는 1≤t≤T의 모든 관측치에 이용할 수 있으나, LS 회귀변수는 시계열의 처음 2개 시점과 마지막 시점을 제외하고 모든 시점에서 이용할 수 있다.

초기화: 사용자에 의해서 설정한 (식 3.3)의 계수 (β, ϕ, θ)를 추정한다. 사전에 정의한 AO와 LS 회귀변수가 모형에 포함된다면, 이들 변수들은 항상 회귀변수 벡터로 유지된다.

추가법(Forward Addition, X-12-ARIMA의 addone 옵션)

1단계: 추정된 $\beta,\ \phi, \theta$에 대해서 다음의 robust 표준오차를 계산한다.

$$\sigma_t^R = \ 1.49 \times median_t |a_t(\beta,\ \phi,\ \theta)|$$

2단계: 추정된 $(\phi,\ \theta)$와 σ_a^R를 이용하여 추가법에 의하여 이미 식별되었거나 미리 지정된 특이치를 제외하고, AO와 LS 회귀변수에 대한 T_t 값을 계산한다. 즉, 일반화 최소자승법에 의해 주어진 회귀변수와 ARIMA 모형의 (ϕ, θ)가 결정되고, 주어진 특이치 회귀변수의 T_t 값을 계산한다.

3단계: $\max|T_t|$ 을 갖는 특이치 회귀변수를 찾는다.

4단계: $|T_t| \geq \gamma$이라면, 이 특이치 회귀변수를 모형에 추가하고 β와 (ϕ, θ)를 재추정한다.

1단계~4단계를 $|T_t| \geq \gamma$를 만족하는 추가의 특이치가 없을 때까지 계속 반복한다.

제거법(Backward Deletion: X-12-ARIMA의 addall 옵션)은 추가법에 의하여 식별된 모든 특이치 변수들을 모형에 포함함으로써 시작한다.

1단계: $(\beta,\ \phi,\ \theta, \sigma_a)$의 최우추정치를 계산한다.

2단계: 추정된 $(\beta,\ \phi,\ \theta, \sigma_a)$을 이용하여, 추가법에 의해 식별되어 모형에 남아 있는 모든 AO와 LS 회귀변수에 대한 T_t를 계산한다.

3단계: $\min |T_t| < \gamma$이라면, 모형에서 이 회귀변수를 제거하고 1단계로 간다.

특이치 검색은 각 시점에 대해서 가장 유의적인 특이치를 찾는 과정이다. AO와 LS, TS 특이치를 찾기 위한 default 과정에서 AO와 LS, TS 회귀변수는 시계열의 전기간에 대해서 적합이 되며, 이때 산출되는 t-통계량을 임계치와 비교한다. AO와 LS, TS 특이치에 대한 default 임계치는 관측치 수에 따라 다음과 같이 임계치가 주어진다.

<특이치 식별을 위한 default 임계치>

관측치 개수	임계치	관측치 개수	임계치
1	1.96	48	3.63
2	2.24	72	3.73
3	2.44	96	3.80
4	2.62	120	3.85
5	2.74	144	3.89
6	2.84	168	3.92
7	2.92	192	3.95
8	2.99	216	3.94
9	3.04	240	3.99
10	3.09	264	4.01
11	3.13	288	4.03
12	3.16	312	4.04
24	3.42	336	4.05
36	3.55	360	4.07

regression 명령문은 AO, LS와 RP(램프 특이치) 옵션을 사용할 수 있으며, RP는 사전에 어떠한 시점(알려진 시점)에서부터 시점까지 영향이 미친다는 것을 아는 경우에 사용한다.

특이치 식별 결과는 다음의 원인에 의해서 특이치 결과가 변할 수 있다. 첫째, 회귀변수나 ARIMA 모형 형태가 변하면 자동으로 선정된 특이치가 바뀔 수 있다. 예를 들어, 명절효과가 큰 시계열의 경우, 명절효과 회귀변수가 RegARIMA 모형에 포함되지 않았거나 사전조정과정에서 충분히 명절효과가 제거되지 않았다면 명절월의 관측치는 특이치로서 식별될 수 있다. 그러나 명절효과 회귀변수를 사용하여 사전조정을 하면 특이치로 식별되지 않을 수 있다. 한편 우리나라 추석의 경우, 분석하는 시계열의 기간이 짧은 경우에는 추석 명절이 특이치로 식별될 수 있으므로 시계열의 기간을 충분히 길게 하여야 한다.

둘째, AO나 LS 특이치로 선정되는 관측치들은 단계별 회귀절차에 의해서 계산된 T_t 값과 지정된(또는 default) 임계치와 비교함으로써 특이치가 결정된다. 새로운 관측치가 시계열에 더해졌을 때, 임계치 보다 근사하게 작은 t-통계량은 임계치보다 큰 t-통계량으로 바뀔 수 있다. 따라서 새로운 관측치가 모형에 첨가됨으로써, 특이치로 식별되었던 관측치가 특이치에서 제외될 수 있다.

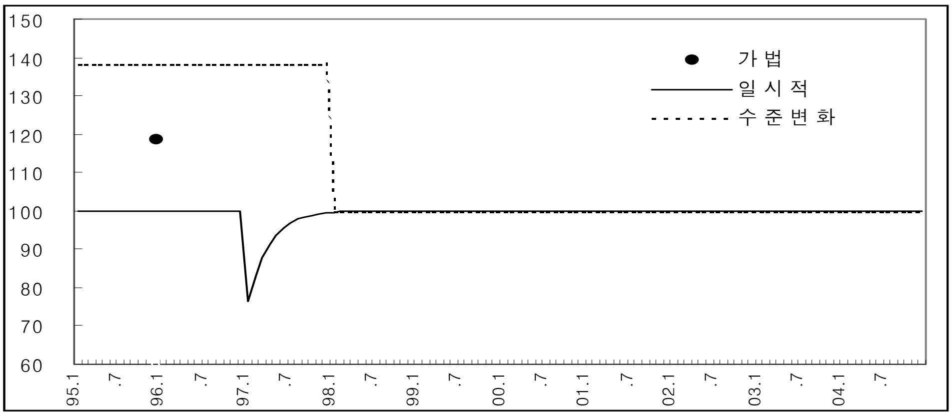
AO, LS와 RP는 실제적으로는 Box와 Tiao(1975)가 설명한 개입모형(intervention)의 단순한 형태이다. X-12-ARIMA는 Box와 Tiao에 의해서 논의된 전 범위에 대한 동적 개입효과(dynamic intervention effect)가 적용되지 않는다. 특이치는 확률적 시계열(사용자 정의의) 회귀변수를 포함하는 RegARIMA와 Box와 Jenkins(1976. 10장, 11장)의 동적 전이함수(dynamic transfer function)에 의해서 검색할 수 있다. 전이함수모형은 일반적으로 설명변수의 미래 값이 미지이기 때문에 예측을 할 때 특별한 처리가 요구되나, RegARIMA모형은 더 일반적인 동적 전이함수모형을 사용할 수 있다. 전이함수모형은 허명회ㆍ박유성(1994)을 참고하면 SAS와 연결하여 쉽게 이해할 수 있다.

(예 3.4) 다음은 1995년 1월부터 2004년 12월까지의 내구소비재에 대한 특이치 검색과 결과에 대한 예이다. 특이치는 가법 특이치, 수준변화, 일시적 변화 특이치 등을 탐색하였다(특이치 탐색 명령은 PART B OUTLIER 참조).

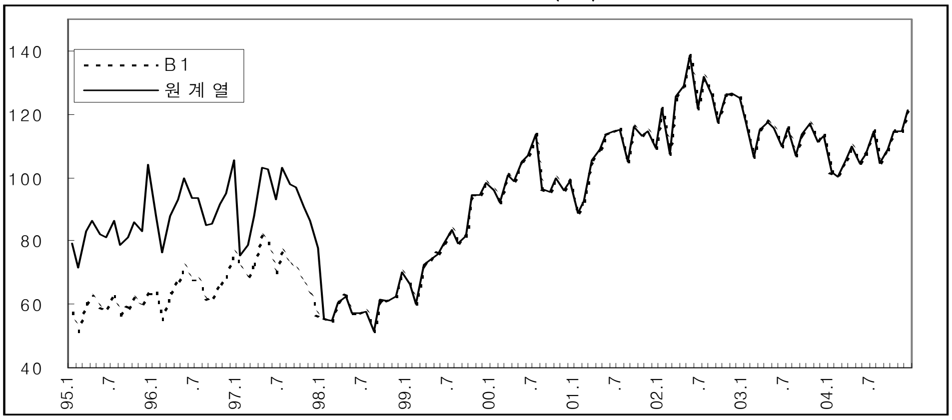
특이치	계수(t-값)	시 점
AO1995.Dec	0.1706(3.99)	1995년 12월
TC1997.Jan	-0.2670(-5.14)	1997년 1월 ~ 1998년 6월
LS1998.Jan	-0.3247(-5.96)	1995년 1월 ~ 1997년 12월

<그림 3.7>은 추정된 특이치 그림이다. 임시적 특이치(TC)는 1997년 1월부터 수준변화가 발생하여 1998년 6월까지 지수적으로 감소하는 형태를 보이고 있으며, 수준변화 특이치(LS)는 1998년 1월 이전과 이후의 수준은 다른 것으로 추정되었다. 이때 1995년 1월부터 1997년 12월까지는 1998년 1월이후 시점보다 높은 수준을 보인다.

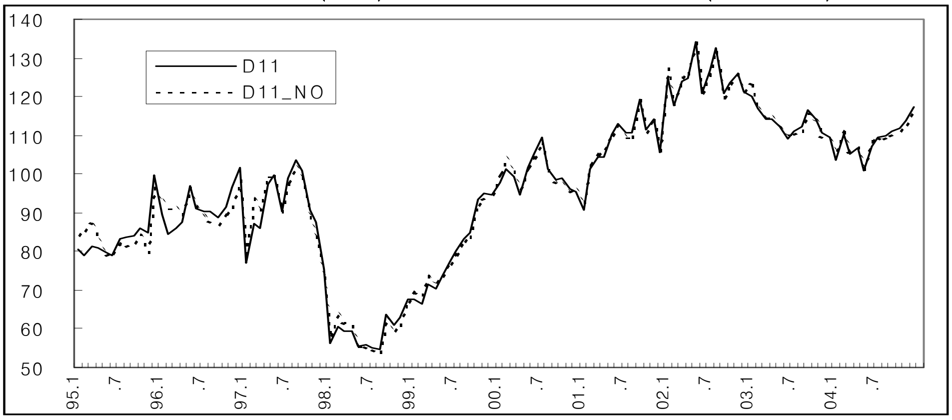
<그림 3.7> 특이치



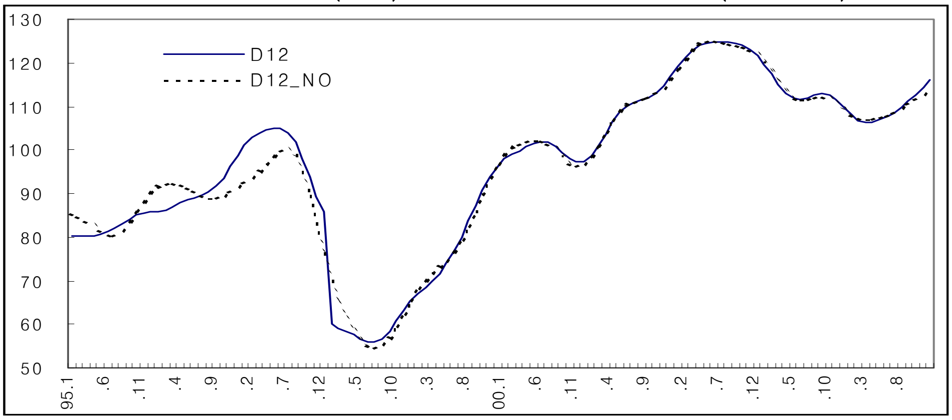
<그림 3.8> 원계열과 사전조정된 원계열(B1)



<그림 3.9> 계절조정계열:
특이치 조정(D11), 특이치 조정하지 않음(D11_NO)



<그림 3.10> 추세·순환 요인:
특이치 조정(D12), 특이치 조정하지 않음(D12_NO)



<그림 3.8>은 추정된 특이치에 의해 사전조정된 원계열(B1)과 특이치를 조정하지 않은 원계열에 대한 그림이다. 이때 계절조정요인은 사전조정된 원계열(B1)로부터 구해지며 <그림 3.9>와 <그림 3.10>의 계절조정계열과 추세·순환요인은 원계열에서 추정된 계절조정요인을 제거한 결과이다. 즉, 사전조정된 원계열은 계절요인을 구하기 위하여 사용되며, 계절조정계열은 사전조정이 되지 않은 원계열로부터 구해진다. 각 그림에서 수준변화로 인해 1998년 이전의 특이치를 조정한 계열과 조정하지 않은 계열의 차이가 크게 남을 볼 수 있다.

3.3 ARIMA모형의 식별

X-12-ARIMA는 요일 및 윤년효과, 명절효과 등의 캘런더효과와 특이치 등을 계절조정하기 전에 사전에 조정해 주기 위하여 Bell과 Hillmer (1983)의 RegARIMA 모형을 채용하고 있다. RegARIMA 모형은 원계열에 사전조정변수(x_{it})를 고려한 회귀모형의 잔차가 Box와 Jenkins(1976)의 ARIMA 모형을 따른다는 가정을 기본으로 하고 있다(자세한 내용은 3.1 RegARIMA 모형 참조). X-12-ARIMA에서 ARIMA 모형을 추정하는 이유는 크게 2가지로 분류할 수 있다.

첫째, X-12-ARIMA의 계절조정방법은 X-11 방법에서 채택하고 있는 선형 filtering 방법인 이동평균법을 채택하고 있어, 중심화 s-개항 이동평균(centered s-term moving average)시 양 끝부분의 s/2개 시점이 결측치로 나타난다. 따라서 이러한 결측치를 없애기 위하여 ARIMA 모형을 이용하여 원계열의 양끝 관측치를 예측하여 연장한다. 둘째, 사전조정을 위한 회귀모형 잔차 z_t는 ARIMA 모형 (p d q)(P D Q)$_s$을 따른다고 가정하고 있다.

$$z_t = y_t - \sum_i \beta_i x_{it}$$

따라서 사전조정을 위한 회귀계수 β_i를 추정하기 위해서는 잔차 z_t의 ARIMA 모형에 대한 (p d q)(P D Q)$_s$ 정보가 필요하다. 즉, 회귀모형과 ARIMA 모형을 혼합한 RegARIMA 모형의 해를 구하기 위하여 필요하다.

$$\phi(B)\Phi(B^s)(1-B)^d(1-B^s)^D(y_t - \sum_i \beta_i x_{it}) \;=\; \theta(B)\Theta(B^s)a_t$$

RegARIMA 모형의 ARIMA 부분은 계절주기 s와 시계열의 구조에 따라 (p d q)(P D Q)에 의해서 결정된다. 이때 ARIMA 모형 결정시, 사전조정을 위한 회귀변수가 RegARIMA 모형 속에 포함되어 있지 않다면, y_t의 표본 자기상관함수(sample autocorrela tion function: ACF)와 표본 편자기상관함수(sample partial autocorrelation function: PACF)를 검토함으로써 할 수 있다. 그러나 사전조정을 위한 회귀변수가 RegARIMA 모형 내에 있는 경우, 회귀효과가 ACF와 PACF을 왜곡시키므로 ARIMA 모형을 수정하는 절차가 필요하다. 일반적으로 차분차수는 시계열 y_t의 ACF

를 검토함으로써 결정할 수 있다. 즉, ACF가 완만하게 감소하는 형태를 보이면 차분이 필요하다고 하겠다. 이때 차분된 회귀변수($\triangle x_{it}$)에 의하여 차분된 시계열($\triangle y_t$)을 회귀함으로써 회귀잔차를 얻을 수 있다. 이들 회귀잔차의 ACF와 PACF의 분석방법은 ARIMA모형의 AR, MA 차수를 결정하기 위한 방법과 동일하다.

X-12-ARIMA에서 ARIMA 모형을 식별하기 위한 명령은 identify이다. 차분차수가 결정되면, X-12-ARIMA의 identify와 regression 명령을 이용하여 차분된 시계열($\triangle y_t$)에 차분된 회귀변수($\triangle x_{it}$)를 회귀한 회귀모형을 설정하고, ARIMA 모형의 AR, MA 차수를 식별하기 위하여 회귀모형의 잔차에 대한 ACF와 PACF를 만들 수 있다. <표 3.12>는 ACF, PACF에 의해서 ARMA모형의 p, d, P, Q를 선정하는 방법이다. ARIMA 모형의 추정은 <참고 3.1 ARIMA 모형의 개요>를 참고하기 바란다.

<표 3.12> ACF와 PACF에 의한 ARIMA 모형 설정

모형	자기상관함수(ACF)	편자기상관함수(PACF)
AR(1)	지수적으로 감소 또는 지수적으로 감소하는 파동곡선	k=1에서 절단
AR(2)	지수적으로 감소 또는 지수적으로 감소하는 파동곡선	k=2에서 절단
AR(p)	지수적으로 감소 또는 지수적으로 감소하는 파동곡선	k=p에서 절단
MA(1)	k=1에서 절단	지수적으로 감소 또는 지수적으로 감소하는 파동곡선
MA(2)	k=2에서 절단	지수적으로 감소 또는 지수적으로 감소하는 파동곡선
MA(q)	k=q에서 절단	지수적으로 감소 또는 지수적으로 감소하는 파동곡선
ARMA(1,1)	지수적으로 감소 (=시차 1부터 AR(1)의 ACF)	지수적으로 감소(=시차 1부터 MA(1)의 PACF)
ARMA(p,q)	k>p부터 지수적 감소	k>q부터 지수적 감소

RegARIMA 모형은 통상최소자승법(Ordinary Least Squares: OLS)에 의해서 (식 3.3)이 적합된다. 예를 들어, 1차 일반 및 계절차분(d=1, D=1)

이 존재하는 시계열이라면, (식 3.11) 회귀모형이 결정되면 회귀잔차 z_t의 ACF와 PACF가 계산된다.

$$(1-B)(1-B^{12})y_t = \sum \beta_i (1-B)(1-B^{12})\ x_{it} + \ z_t \qquad (3.11)$$

Bell과 Hillmer(1983)는 REGARIMA 모형을 식별하기 위해 다음의 절차로 설명하고 있다. 예를 들어 최대 차분차수가 d=1, D=1이라고 가정하면, RegARIMA 모형을 추정하기 위한 절차는 다음과 같다.

(i) OLS 방법에 의하여 회귀잔차 $\tilde{z}_t = y_t - \sum \tilde{\beta}_i x_{it}$를 구한다.

(ii) 회귀잔차 $\tilde{\tilde{z}}_t = y_t - \sum \tilde{\tilde{\beta}}_i x$를 구한다.

(iii) 회귀잔차 $\tilde{z}_t$, $(1-B)\tilde{z}_t$, $(1-B^{12})\tilde{z}_t$와 $(1-B)(1-B^{12})\tilde{z}_t$에 대한 ACF와 PACF를 구한 후, ARMA 모형의 (p q)(P Q)를 결정한다.

ACF와 PACF는 모형의 AR · MA의 차수 뿐만 아니라 차분 차수를 결정하기 위하여 이용한다. 즉, ACF가 완만히 감소한다면 시계열의 차분이 필요하다.

X-12-ARIMA/0.3에서는 ARIMA 모형을 자동 식별하기 위해서 2가지 방법을 제공하고 있다. 첫째는 X-11-ARIMA와 X-12-ARIMA(기존 버전)에서 적용했던 방법과 같이 표준 ARIMA 모형군을 설정하여 ARIMA 모형을 찾는 방법으로 "**PICKMDL**" 명령을 이용한다. 둘째는 TRAMO/SEATS에서 적용한 방법으로 Hannan-Rissanen 반복계산법(Hannan과 Rissanen, 1982)을 기초로 ARIMA(p,d,q)의 차수를 최소 AIC-통계량을 만족하는 (p, d, q)를 찾는 방법이다. 이 방법은 X-12-ARIMA/0.3에서는 “**AUTOMDL**" 명령을 이용한다.

X-12-ARIMA/0.3의 PICKMDL 명령은 X-11-ARIMA와 X-12-ARIMA(기존 버전) 적용하였던 방법으로 X-12-ARIMA(기존 버전)의 AUTOMDL 명령과 동일하다.

3.3.1. PICKMDL 명령에 적용된 방법

X-12-ARIMA에서는 X-11-ARIMA와는 달리 사용자가 ARIMA 표준모형을 임의로 수정하여 시계열에 적합할 수 있다. 기존 X-12-ARIMA의 default ARIMA(p,d,q)(P,D,Q) 모형은 default 파일인 x12a.mdl 파일 속에 다음과 같은 표준 ARIMA 모형이 저장되어 있다.

(0 1 1)(0 1 1)　(0 1 2)(0 1 1)　(2 1 0)(0 1 1)
(0 2 2)(0 1 1)　(2 1 2)(0 1 1)

그러나 X-12-ARIMA/0.3에서는 표준 모형을 저장한 default 파일은 없으며, 사용자가 표준 ARIMA 모형을 저장한 파일을 지정할 수 있다. 그러나 표준 모형을 저장한 파일을 지정하지 않으면 default 모형은 ARIMA(0 1 1)(0 1 1)이 적용된다. 사용자가 ARIMA 표준 모형군을 설정하고자 한다면, ARIMA 모형군을 설정한 저장 파일을 생성하고 확장자를 "*.mdl"로 지정한다. PICKMDL 명령의 자세한 사용방법은 PART B에서 설명한다.

PICKMDL 명령의 자동 모형 선정기준은 다음과 같으며, 자세한 내용은 통계조사국(1995)을 참고할 수 있다.

① 최근 3년 동안의 외삽 값(extrapolated value)에 대한 절대 평균 퍼센트 오차(absolute average percentage error)가 15%보다 작아야 한다.
② 모형의 잔차상관이 있는지 검정하기 위하여 적합된 모형에 대하여 Ljung-Box Q-통계량(Portmanteau-test)을 구하며, 이때 Q-통계량의 p-값이 5% 보다 커야 한다.
③ 과도차분의 징후가 없어야 한다. 즉, 계절 또는 비계절의 MA 모수 값의 합이 0.9보다 작아야 한다. 모형 파일에 있는 모형을 받아들일 수 없으면 모형 선정을 하지 않는다.

이들 기준은 fcslim, qlim 혹은 overdiff 옵션을 이용하여 변경할 수 있다. X-12-ARIMA 계열 프로그램에서 주어지는 모형 선정을 위한 우도함수 통계량은 다음과 같으며, 이들 통계량이 최소가 되는 ARIMA 모형을 최적 ARIMA 모형으로 선정한다.

AIC, AICC(F-adjusted AIC), Hannan Quinn, BIC

위의 통계량들은 우도검정을 이용하여 여러 개의 ARIMA 모형 중에서 최적의 모형을 선정하는 통계이나, t-통계량과 χ^2-통계량은 회귀모형에 의하여 추정된 회귀계수의 적합성을 검정해 준다.

3.3.2. AUTOMDL 명령에 적용된 방법

X-12-ARIMA/0.3에서 AUTOMDL 명령의 수행 과정은 다음과 같이 크게 5단계로 나눌 수 있다;

- default 모형 추정
- 차분 차수 식별
- ARIMA 모형 식별
- default 모형과 식별된 모형의 비교
- 최종모형 검토 단계

각 단계에서의 수행 작업은 다음과 같다.

Default 모형 추정단계에서는 ARIMA(0 1 1)(0 1 1)를 default 모형으로 설정하고, 특이치를 식별한다. 이때 특이치가 식별되면, 요일효과, 명절효과 및 상수항이 유의적인지 검토함으로써 RegARIMA 모형의 초기 모형을 설정한다. 이때 검정을 위해 사용되는 통계량은 Ljung-Box Q-통계량, 잔차 평균에 대한 t-통계량 등이다.

차분 차수 식별 단계에서는 시계열을 차분하고, ARIMA 모형의 적합과 단위근 검정을 실시한다. 차분한 시계열에 대한 ARIMA 모형의 적합은 Gomez와 Maravall(2000)이 TRAMO에 적용한 Hannan-Rissanen의 반복 계산법에 의해서 추정한다. 이때 AR 모형의 모수 값이 1에 근사하다면 단위근을 갖는다고 볼 수 있으며 차분이 필요하다.

ARIMA 모형 식별을 위하여, 먼저 계절 ARIMA 모형에 대한 초기모형을 결정한다. 초기 모형을 결정하기 위하여 (0≤P, Q≤2)에 대한 (3 d 0)(P D Q)의 모든 ARIMA 모형을 설정하고 다음의 BIC2-통계량을 계산한다. 이때 BIC2가 최소가 되는 P와 Q를 선택한다.

$$BIC2_N = (-2\widehat{L_N} + n_p \log N)/N$$

여기서 N은 모형의 일반 및 계절 차분을 실시한 후의 관측치 수이며, $\widehat{L_N}$은 N개의 관측치에 의해 계산된 log 우도함수의 최대값, n_p는 백색잡음

분산을 포함한 모형에서 추정된 모수의 수이다.

계절 ARIMA 모형의 모수에 대한 차수 $\tilde{P}$와 $\tilde{Q}$가 결정되면, 일반 ARIMA 모형에 대한 최적 모형을 식별하기 위하여, (0≤p, q≤3)대해서 $(p\ \tilde{d}\ q)(\tilde{P}\ \tilde{D}\ \tilde{Q})$의 모든 ARIMA 모형에 대한 BIC2를 계산한다. 이때 BIC2가 최소가 되는 p와 q를 선택한다.

추정된 $\tilde{p}$와 $\tilde{q}$를 이용하여, 계절모형의 차수를 다시 결정한다. (0≤P, Q≤2)에 대해서 (p d q)(P D Q)s의 모든 ARIMA 모형에 대한 BIC2를 계산한다. 이때 BIC2가 최소가 되는 P와 Q를 선택한다.

식별된 모형이 default 모형이 아닌 경우, 자동으로 선정된 모형의 잔차와 default 모형의 잔차를 비교한다. 이때 비교하는 통계량은 default 모형과 식별된 모형의 Ljung-Box Q 통계량의 신뢰계수 및 잔차 표준 오차를 이용한다. 모형 선정 기준은 <PART B, AUTOMDL 명령의 참고>를 따른다.

최종적으로 모형을 결정하기 위하여, AR 및 MA 다항식의 단위근을 검토하여 차분의 적정성을 보며 RegARIMA 모형의 회귀계수에 대한 t-검정을 실시한다. 이후 회귀변수들이 적정하다면 ARIMA 모형의 계수에 대한 t-검정을 실시한다.

3.4 모형추정과 추론

X-12-ARIMA에서는 regression과 arima 명령에 의하여 REGARIMA 모형을 설정한다. 이때 estimate 명령은 최대우도(maximum likelihood) 혹은 조건부 최소자승(conditional least square)으로 알려진 조건부 최대우도(conditional maximum likelihood)에 의해서 모형의 모수를 추정한다. 사용자는 AR 모수 추정을 위하여 정확한 최대우도나 조건부 최대우도에 의해서 최대화를 할 수 있으나 MA 모수 추정을 위해서는 정확한 최대우도, AR과 MA 모수 추정을 위해서는 조건부 최대우도를 최대화하도록 설정한다. 일반적으로 정확한 최대우도와 조건부 최대우도의 최대화에 의해 추정된 AR 모수 차이는 작으며, 각 방법은 적절한 것으로 알려져 있다. MA 모수 추정을 위한 정확한 최대우도와 조건부 최대우도사이의 차이는 근본적인 것으로 정확한 최대우도 추정법을 권장한다.

X-12-ARIMA의 MA 모수 추정에 대한 조건부 최대우도 옵션은 다른 software의 결과와 비교하기 위한 것이다. 또한 추정치의 수렴에 문제가 있는 경우, 정확한 최대우도 추정법의 초기치를 생성하기 위한 것이다. AR과 MA 모수추정을 위한 default 방법은 정확한 최대우도 추정법이다.

REGARIMA모형의 우도함수를 최대화하기 위하여 일반화된 반복최소자승기법(iterative generalized least square: IGLS)이 사용된다. 이 기법은 두단계로 이루어져 있다.

1단계: 주어진 AR, MA 모수 값에 대하여, 우도함수를 최대화하는 회귀모수는 일반화된 최소자승(GLS)법에 의해서 구한다. (GLS는 ARIMA 모형에 의해서 결정된 회귀잔차의 공분산구조를 이용한다.)

2단계: 주어진 회귀계수 β_i에 대하여, ARIMA 모형은 다음의 회귀잔차에 대한 최대우도함수에 의해서 적합된다.

$$z_t = y_t - \sum \beta_i x_{it}$$

IGLS는 수렴치에 도달할 때까지 이들 두 단계를 반복한다.

REGARIMA 모형의 모수에 대한 통계적 추론은 ARIMA 모형의 최대우도추정법(maximum likelihood estimation: MLE)에 대한 점근적 결과를 이용한다. 이때 "추정되는 모수는 평균이 참모수 값과 같으며 추정가능한 공분산행렬을 갖는 정규분포이다"라는 기본가정을 갖는다.

X-12-ARIMA에서는 개개의 회귀모수와 특별한 회귀효과를 나타내는 회귀모수의 통계적 유의성을 평가하기 위하여 t-통계량과 χ^2-통계량이 주어진다. 또한 X-12-ARIMA는 최대화된 지수 우도함수값을 인쇄해주기 때문에 다른 모형을 가정하는 경우와의 우도비검정이 가능하다.

X-12-ARIMA는 잔차분산 σ^2의 MLE를 이용한다. 여기서 잔차분산의 추정량은$\hat{\sigma}^2 = SS/(T-d-s \cdot D)$ 이며, SS는 잔차자승합, (T-d-s · D)는 차분후의 유효한 관측치 수이다. 만일 AR모수에 대하여 조건부 우도함수를 이용한다면, (T-d-s · D) 대신 (T-p-d-s · P-s · D)를 사용한다.

3.5 예측

추정된 REGARIMA 모형은 forecast 명령에 의해서 점추정과 예측과 관련된 표준편차, 구간추정 등을 위해서 이용된다. 점추정치는 참모형이 사용된다는 가정하에 현재와 과거시계열 y_t를 바탕으로 평균오차자승(mean squared error: MMSE)이 최소가 되는 선형 예측치의 값이다. 이때 REGARIMA 모형은 적절한 회귀변수가 사용되며 예측기간 내에서는 가법특이치나 수준이동이 발생하지 않는다는 가정을 한다. 그러나 이러한 가정은 실제상황과는 괴리가 있는 비현실적인 가정이다. 현실적인 가정은 REGARIMA 모형의 추정된 모수 값들이 참값에 충분히 가깝게 될 것이라는 것이다.

시계열의 변환(series transformation)은 정상시계열 조건을 요구하는 ARIMA 모형을 만족하기 위한 방법이다. 시계열의 변환은 log, 이승근(square root), 역(inverse), 로지스틱 변환 등을 실시할 수 있다(자세한 내용은 PART B의 TRANSFORM 명령 참조).

시계열이 변환되었거나 또는 사전조정이 되었다면 $(Y_t = F(y_t))$, 먼저 예측치는 변환된 시계열로 예측$(\widehat{Y_t})$되며 이때 예측된 시계열은 원래의 원계열 $(\hat{y}_t = F^{-1}(\widehat{Y_t}))$로 역변환(inverse transform)된다. 다음은 log, 로지스틱, 이승근 변환의 역변환 식이다(SAS 매뉴얼 Chapter30 Sect 5 참조).

	변 환	역 변 환
log	$Y_t = \log(y_t)$	$y_t = \exp(Y_t + \frac{\sigma_t^2}{2})$
이 승 근	$Y_t = \sqrt{y_t}$	$y_t = Y_t^2 + \sigma_t^2$
로지스틱	$Y_t = \log(\frac{c\ y_t}{1 - c\ y_t})$	$y_t = E[\frac{1}{c(1+\exp(-Y_t))}]$

여기서 scaling factor $c = (1 - e^{-6})10^{-ceil(\log_{10}(\max(y_t)))}$, $ceil(x)$는 x와 같거나 큰 가장 작은 정수이다.

예를 들어, $Y_t = \log(y_t)$로 변환되었다면, 먼저 $\widehat{Y_t}$가 예측이 되며, 이후

점추정과 구간추정의 결과는 역변환된 원계열 $\hat{y}_t$의 점추정치와 구간추정치가 계산된다. 여기서 역변환된 시계열 값은 $\hat{y}_t = \exp(\hat{Y}_t + \hat{\sigma}_t^2/2)$으로 계산된다. 따라서 점추정치와 구간추정치의 결과는 y_t의 MMSE 예측결과가 아니며, $Y_t = \log(y_t)$에 대한 MMSE의 결과이다.

3.6 모형 선정 기준

X-12-ARIMA는 모형 선정을 위하여 AIC, AICC, Hannan Quinn, BIC 통계량을 이용한다. n_p를 백색잡음 분산을 포함한 모형에서 추정하는 모수의 수이라고 하자. 만일 시계열을 일반 또는 계절차분하여 모형에 적합하였다면, 관측치 수는 N이며 L_N은 N개의 관측치에 의해 추정된 log 우도함수의 최대값이다. 이들 기준치에 대한 통계량은 다음과 같다.

AIC-통계량(Akaike's Information Criterion) $= -2L_N + 2n_p$

AICC(F-Corrected AIC) $= -2L_N + 2n_p\left(\frac{N}{N - n_p - 1}\right)$

Hannan Quinnen $= -2L_N + 2n_p \log\log N$

BIC-통계량(Baysian's Information Criterion) $= 2L_N + n_p \log N$

주어진 시계열에 대해서 모형들이 적합되고 통계량들이 계산・비교되며, 가장 작은 통계량을 갖는 모형을 최적의 모형으로 선정한다. 다음은 이들 통계량을 사용시 주의해야 할 점이다(U.S. Census Bureau, 2004). 첫째, 이들 통계량의 비교는 같은 차분을 한 모형 내에서 유효하다. 둘째, 모형 추정시, AR과 MA 모수를 정확한 최우추정법에 의해서 추정하여야 한다. 이러한 이유 때문에 조건부 우도함수를 사용시에는 모형 선정 기준 통계가 인쇄되지 않는다.

이들 모형 선정 기준 통계량은 다른 특이치 회귀변수(AO, LS, TC 혹은 Ramp)를 갖는 RegARIMA 모형을 비교하는 데 사용할 수 없다. 특히, 모형에 특이치를 포함할 것인지를 결정하기 위한 기준으로 사용할 수 없다. 그러나 같은 특이치 변수를 갖는 모형의 비교시에는 특이치 회귀 추

정은 같은 방법으로 모형의 ML 값에 영향을 미친다. 즉, 로그 우도(log likehood)의 차분을 하였을 때 이들 효과는 없어진다. 예를 들어, t시점의 가법 특이치 변수 AO_t는 t시점의 결측치와 같은 효과를 갖고 있어 두 모형에서 결측치로서 다루어 진다. 그러나 만일 두 모형에서 다른 AO 변수를 갖는다면, 다른 자료로 모형을 추정하는 것과 같게 되므로 로그 우도나 로그 우도로부터 계산된 통계량을 비교하는 것은 적절하지 않다. 따라서 두 모형을 비교하여 적합한 모형을 선정할 때, 일반 및 계절 차분을 같게 하여야 한다.

다른 특이치 변수를 갖는 모형들을 비교하기 위하여 앞의 모형 선정 기준을 이용한다면, 대개는 보다 큰 기준치를 갖는 모형이 선정되어 예측치는 나쁜 결과를 얻을 수 있다. 또한 선정된 모형에 포함된 특이치의 일부는 1년 또는 2년의 시계열이 추가되면, 특이치로서 식별되지 않으므로서 모형 식별의 불안정을 나타내게 된다.

변환(transformation)과 사전조정(prior adjustments)는 모형 선정 기준치에 영향을 준다. 예를 들어, transformation 명령에 none 이외의 값을 갖는 function 옵션이 있거나 $\lambda \neq 1$인 power 옵션을 사용하였다면, 우도함수에 영향을 미치며, 이에 따라 모형 선정 기준에도 영향을 미친다.

<참고 3.1> ARIMA 모형의 개요

다음과 같이 T개의 관측치를 갖는 시계열이 있다고 가정하자.

$$(z_1, \ldots, z_{t-k}, \ldots, z_{t-1}, z_t, \ldots, z_T)$$

□ 표본자기상관함수(Sample AutoCorrelation Function: SACF)

표본자기상관계수 r_k는 t시점의 관측값 z_t와 시차 k를 갖는 관측값 z_{t-k} 사이의 자기상관을 나타내며, 다음과 같이 정의할 수 있다.

$$r_k = \frac{\gamma(k)}{\gamma(0)} = \frac{\sum_{t=k+1}^{T}(z_t - \bar{z})(z_{t-k} - \bar{z})}{\sum_{t=1}^{T}(z_t - \bar{z})^2}, \quad k = 0,1,2,\ldots,T-1$$

여기서 자기공분산 $\gamma(k)$는 다음과 같이 정의한다.

$$\gamma(k) = Cov(z_t, z_{t-k}) = \frac{1}{T}\sum_{t=1}^{T-k}(z_t - \bar{z})(z_{t-k} - \bar{z})$$

□ 표본편자기상관함수(Sample Partial AutoCorrelation Function: SPACF)

k 시차를 갖는 편자기상관계수 b_{kk}는 다음에 의해서 구해진다.

$$b_{kk} = corr(z_t, z_{t-k} | z_{t-1}, \ldots, z_{t-k+1})$$

즉, z_t를 $z_{t-1}, \ldots, z_{t-k}$에 회귀시켰을 때 z_{t-k}에 걸리는 회귀계수이다.

$$z_t = b_{k0} + b_{k1}z_{t-1} + \ldots + b_{kk}z_{t-k}$$

□ 백색잡음과정(White noise) 및 확률보행과정(random walk)

$(a_1, a_2, \ldots, a_t, \ldots)$를 평균 0, 분산 σ^2인 동일한 분포로부터 독립적으로 얻어지는 확률분포라면, 백색잡음과정과 확률보행과정은 다음과 같다.

- 백색잡음과정: $z_t = \mu + a_t \qquad t = 1,2,\ldots,T$
- 확률보행과정: $z_0 = \mu, \quad z_t = z_{t-1} + a_t, \qquad t = 1,2,\ldots,T,\ldots$

즉, 확률보행과정은 $z_t = \mu + a_1 + a_2 + \ldots + a_{t-1} + a_t, \quad t = 1,2,\ldots,T$ 이므로 백색잡음(random shock)들의 누계이다.

□ 정상시계열(stationary series)

정상시계열 조건을 만족하는 시계열을 정상시계열(stationary series), 만족하지 않는 시계열을 비정상시계열(nonstationary series)이라 한다.

<정상시계열 조건>

① 평균 일정: 모든 t에 대해 $E(z_t) = \mu$

② 분산이 존재하며 상수: 모든 t에 대해 $Var(z_t) = \gamma(0) < \infty$

③ 두 시점사이의 자기 공분산은 시차(time lag)에만 의존:

모든 t와 s에 대해 $Cov(z_t, z_s) = \gamma(|t-s|)$

비정상시계열의 경우, 차분(differencing)을 함으로써 정상시계열로 변환하며, z_t의 차분(∇) 및 후향연산자 B($B^k z_t = z_{t-k}$)를 이용하면 다음과 같이 표현할 수 있다.

- 1차 차분: $\nabla z_t \equiv z_t - z_{t-1} = (1-B)z_t$

  1차 차분은 $(1-B)z_t$이므로 $\phi = 1$인 AR(1) 모형 $(1 - \phi B)z_t$과 같다.

- 2차 차분: $\nabla^2 z_t \equiv \nabla\nabla z_t$

$$
\begin{aligned}
&= (z_t - z_{t-1}) - (z_{t-1} - z_{t-2}) = z_t - 2z_{t-1} + z_{t-2} \\
&= [(1-B)z_t] - [(1-B)z_{t-1}] = (1-B)(z_t - z_{t-1}) \\
&= (1-B)^2 z_t
\end{aligned}
$$

- 비정상시계열이 선형 추세를 갖는 경우: $z_t = \beta_0 + \beta_1 t + \epsilon_t$

  1차 차분하면, $\nabla z_t = (z_t - z_{t-1})$

$$
\begin{aligned}
&= (\beta_0 + \beta_1 t + \epsilon_t) - (\beta_0 + \beta_1 (t-1) + \epsilon_{t-1}) \\
&= \beta_1 + (\epsilon_t - \epsilon_{t-1}) = \beta_1 + \nabla\epsilon_t
\end{aligned}
$$

- 2차 곡선추세를 갖는 비정상시계열 경우: $z_t = \beta_0 + \beta_1 t + \beta_2 t^2 + \epsilon_t$

  1차 차분하면, $\nabla z_t = \beta_1 + 2\beta_2 t + \nabla\epsilon_t$

  2차 차분하면, $\nabla^2 z_t = 2\beta_2 + \nabla^2\epsilon_t$

- 계절주기 s를 갖는 계절차분(seasonal differencing): $\nabla_s z_t = z_t - z_{t-s}$

  D차 계절차분: $\nabla_s^D z_t = (1-B^s)^D z_t$

(예) $\nabla_s^2 z_t = (1-B^s)^2 z_t$
$$= (1-B^s)(1-B^s)z_t = (1-B^s)(z_t - z_{t-s})$$
$$= (z_t - z_{t-s}) - (z_{t-s} - z_{t-2s}) = z_t - 2z_{t-s} + z_{t-2s}$$

□ ARMA(p, q): 자기회귀이동평균과정(AutoRegressive Moving Average process):

○ 자기회귀과정(AutoRegressive process): AR(p)

- AR(p) 모형은 가장 최근 값인 z_{t-1}이 가장 많은 정보를 갖고 있으며, 시간이 갈수록 정보의 양은 감소하는 모형이다.
- 모든 k>0에 대해 $Cov(z_t, z_{t-k}) = 0$이라면, AR(p) 모형은 다음과 같이 표현할 수 있다.

$$z_t = \phi_1 z_{t-1} + \phi_2 z_{t-2} + \ldots + \phi_p z_{t-p} + a_t \quad (1)$$
$$= \phi_1 B z_t + \phi_2 B^2 z_t + \ldots + \phi_p B^p z_t + a_t$$

(식 1)로부터, $z_t - \phi_1 B z_t - \phi_2 B^2 z_t - \ldots - \phi_p B^p z_t = a_t$
$$(1 - \phi_1 B - \phi_2 B^2 - \ldots \phi_p B^p) z_t = a_t$$
$$\phi(B) z_t = a_t$$

여기서 $\phi(B) = 1 - \phi_1 B - \phi_2 B^2 - \ldots - \phi_p B^p$이다.

○ 이동평균과정(Moving Average process): MA(q)

- MA(q) 모형은 백색잡음과정을 이용하여 다음과 같이 표현할 수 있다.

$$z_t = a_t - \theta_1 a_{t-1} - \ldots - \theta_q a_{t-q}$$
$$= (1 - \theta_1 B - \theta_2 B^2 - \ldots - \theta_q B^q) a_t$$
$$= \theta(B) a_t$$

여기서 $\theta(B) = 1 - \theta_1 B - \theta_2 B^2 - \ldots - \theta_q B^q$이다.

- MA(q) 모형은 $(a_t, a_{t-1}, \ldots, a_{t-q})$에 가중치 $(1, -\theta_1, -\theta_2, \ldots, -\theta_q)$를 부여하여 시계열 z_t를 계산, 동일한 가중치로 $(a_{t+1}, a_t, \ldots, a_{t-q+1})$와 결합하여 z_t의 한 시점 이후인 z_{t+1}를 계산하므로 시계열은 이동평균에 의해서 생성된다.

○ARMA 모형 식별 방법

- SACF, SPACF에 의해 AR, MA 모형의 모수(ϕ, θ) 차수 결정
- 추정된 AR, MA 모수에 대한 t-검정(p-값)에 의해 t-값이 너무 작으면(p-값 너무 크면) 추정된 AR, MA 모수는 부적절하고 할 수 있다.
- 적합된 모형의 잔차에 대한 Ljung-Box Q-통계량을 이용한 포트만토 검정(Portmanteau test)을 실시한다. 귀무가설은 “ARIMA 모형이 타당하다”이다. SAS 결과는 “Autocorrelation check for Residuals"를 분석한다. 이때 Q-통계량의 p-값(Pr>Chisq)이 0.05(5%)보다 크면 식별된 모형은 타당하다고 볼 수 있다.

$$Q = n(n+2)\sum_{k=1}^{K} r_k/(n-k) \sim \chi^2(K-p-q)\text{분포}$$

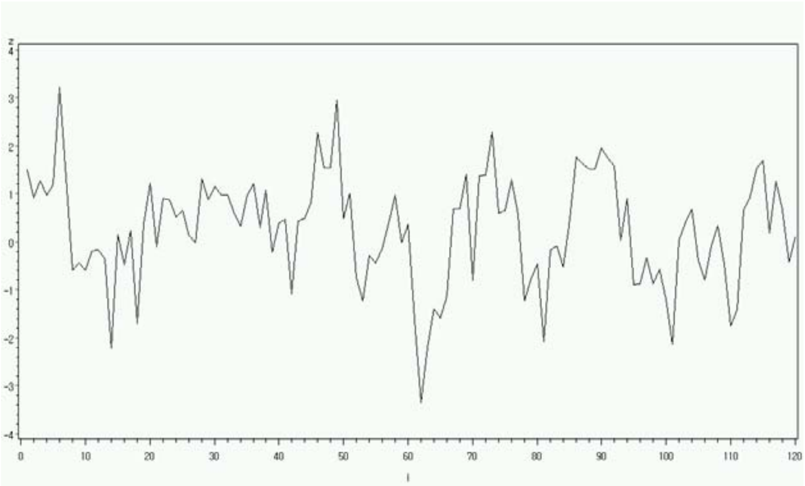
여기서 n은 차분후의 자료수

ACF와 PACF에 의한 ARIMA 모형 설정

모형	자기상관함수(ACF)	편자기상관함수(PACF)
AR(1)	지수적으로 감소 또는 지수적으로 감소하는 파동곡선	k=1에서 절단
AR(2)	지수적으로 감소 또는 지수적으로 감소하는 파동곡선	k=2에서 절단
AR(p)	지수적으로 감소 또는 지수적으로 감소하는 파동곡선	k=p에서 절단
MA(1)	k=1에서 절단	지수적으로 감소 또는 지수적으로 감소하는 파동곡선
MA(2)	k=2에서 절단	지수적으로 감소 또는 지수적으로 감소하는 파동곡선
MA(q)	k=q에서 절단	지수적으로 감소 또는 지수적으로 감소하는 파동곡선
ARMA(1,1)	지수적으로 감소 (=시차 1부터 AR(1)의 ACF)	지수적으로 감소(=시차 1부터 MA(1)의 PACF)
ARMA(1,1)	k>p부터 지수적 감소	k>q부터 지수적 감소

(예 1) $\phi_1 = 0.6$을 갖는 AR(1) 모형: $z_t = \phi_1 z_{t-1} + a_t$

$(1 - 0.6B) z_t = a_t$



Autocorrelations

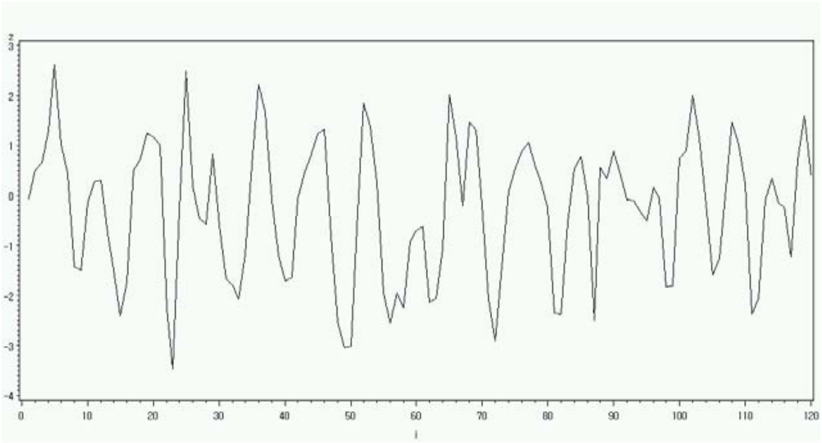
```
Lag  Covariance   Correlation  -1 9 8 7 6 5 4 3 2 1 0 1 2 3 4 5 6 7 8 9 1   Std Error
  0   1.267391     1.00000     |                    |********************|          0
  1   0.718459     0.56688     |                .   |***********         |   0.091287
  2   0.491370     0.38770     |               .    |********            |   0.117001
  3   0.329740     0.26017     |               .    |*****               |   0.127257
  4   0.119799     0.09452     |               .    |**  .               |   0.131615
  5  -0.021404     -.01689     |               .    |    .               |   0.132180
  6  -0.211536     -.16691     |               . ***|    .               |   0.132198
  7  -0.361080     -.28490     |             ******|     .               |   0.133942
  8  -0.259012     -.20437     |              . ****|     .              |   0.138900
  9  -0.186343     -.14703     |              .  ***|     .              |   0.141384
 10  -0.189864     -.14981     |              .  ***|     .              |   0.142652
 11  -0.182229     -.14378     |              .  ***|     .              |   0.143957
 12  -0.245000     -.19331     |              . ****|     .              |   0.145149
 13  -0.138792     -.10951     |              .   **|     .              |   0.147279
 14  -0.125879     -.09932     |              .   **|     .              |   0.147956
 15  -0.065246     -.05148     |              .    *|     .              |   0.148511
```

Partial Autocorrelations

```
Lag  Correlation  -1 9 8 7 6 5 4 3 2 1 0 1 2 3 4 5 6 7 8 9 1
  1    0.56688     |                 .  |***********         |
  2    0.09777     |                 .  |** .                |
  3    0.00961     |                 .  |   .                |
  4   -0.11865     |                 . **|  .                |
  5   -0.07881     |                 . **|  .                |
  6   -0.17328     |                 .***|  .                |
  7   -0.15644     |                 .***|  .                |
  8    0.11101     |                .   |** .                |
  9    0.06260     |                .   |*  .                |
 10   -0.06197     |                 .  *|  .                |
 11   -0.09065     |                 . **|  .                |
 12   -0.17329     |                 .***|  .                |
 13    0.02751     |                 .  |*  .                |
 14   -0.05429     |                 .  *|  .                |
 15    0.08768     |                 .  |** .                |
```

(예 2) $\phi_1 = 0.87$, $\phi_2 = -0.52$를 갖는 AR(2) 모형: $z_t = \phi_1 z_{t-1} + \phi_2 z_{t-2} + a_t$

$$(1 - 0.87B + 0.52B^2) z_t = a_t$$



Autocorrelations

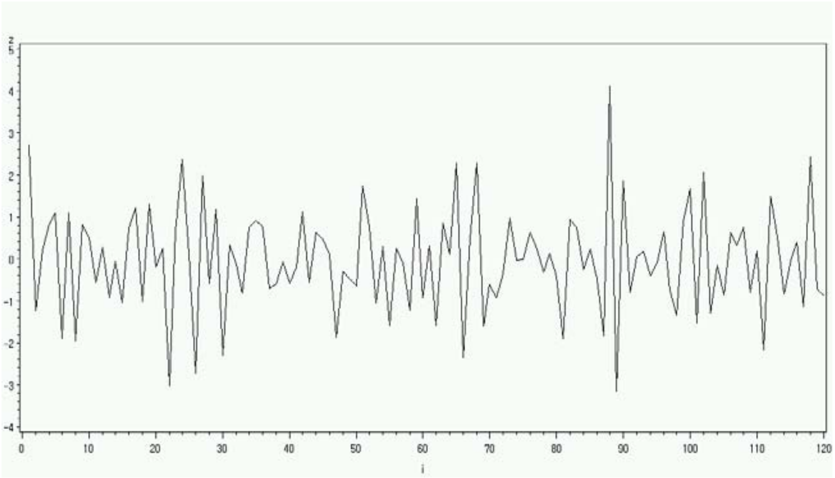
Lag	Covariance	Correlation	-1 9 8 7 6 5 4 3 2 1 0 1 2 3 4 5 6 7 8 9 1	Std Error
0	1.855803	1.00000	|********************|	0
1	1.029192	0.55458	| . |*********** |	0.091287
2	-0.117298	-.06321	| . *| . |	0.116014
3	-0.748800	-.40349	| ********| . |	0.116301
4	-0.721490	-.38878	| ********| . |	0.127434
5	-0.360968	-.19451	| .****| . |	0.136961
6	0.024010	0.01294	| . | . |	0.139244
7	0.285325	0.15375	| . |*** . |	0.139254
8	0.389252	0.20975	| . |**** . |	0.140662
9	0.222899	0.12011	| . |** . |	0.143245
10	0.020028	0.01079	| . | . |	0.144081
11	-0.115954	-.06248	| . *| . |	0.144088
12	-0.215204	-.11596	| . **| . |	0.144314
13	-0.280411	-.15110	| . ***| . |	0.145088
14	-0.095183	-.05129	| . *| . |	0.146394
15	0.137624	0.07416	| . |* . |	0.146543

Partial Autocorrelations

Lag	Correlation	-1 9 8 7 6 5 4 3 2 1 0 1 2 3 4 5 6 7 8 9 1
1	0.55458	| . |*********** |
2	-0.53545	| ***********| . |
3	-0.10674	| . **| . |
4	-0.07609	| . **| . |
5	-0.08771	| . **| . |
6	-0.02160	| . | . |
7	0.03153	| . |* . |
8	0.06296	| . |* . |
9	-0.06839	| . *| . |
10	0.07908	| . |** . |
11	-0.00567	| . | . |
12	-0.07349	| . *| . |
13	-0.07453	| . *| . |
14	0.12174	| . |** . |
15	-0.03951	| . *| . |

(예 3) $\theta = 0.8$를 갖는 MA(1) 모형: $z_t = a_t - \theta_1 a_{t-1}$인 경우

$$z_t = (1 - 0.8B)a_t$$



Autocorrelations

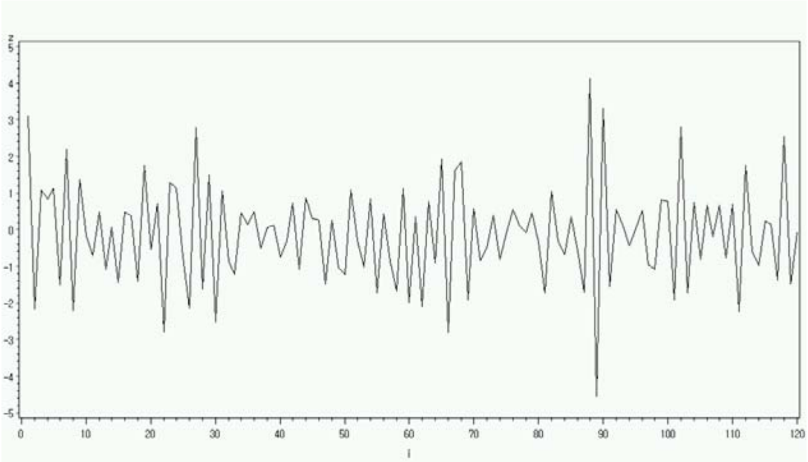
```
Lag  Covariance  Correlation  -1 9 8 7 6 5 4 3 2 1 0 1 2 3 4 5 6 7 8 9 1   Std Error
  0   1.514588    1.00000     |                    |********************|         0
  1  -0.683244    -.45111     |           *********|    .                |  0.091287
  2   0.043751    0.02889     |                .   |*   .                |  0.108282
  3  -0.105236    -.06948     |                .  *|    .                |  0.108346
  4   0.034505    0.02278     |                .   |    .                |  0.108717
  5  -0.092250    -.06091     |                .  *|    .                |  0.108757
  6   0.085524    0.05647     |                .   |*   .                |  0.109041
  7  -0.162361    -.10720     |                . **|    .                |  0.109284
  8   0.246027    0.16244     |                .   |***.                |  0.110157
  9  -0.070266    -.04639     |                .  *|    .                |  0.112135
 10  -0.015886    -.01049     |                .   |    .                |  0.112295
 11  -0.020404    -.01347     |                .   |    .                |  0.112303
 12   0.139649    0.09220     |                .   |**  .                |  0.112317
```

Partial Autocorrelations

```
Lag  Correlation  -1 9 8 7 6 5 4 3 2 1 0 1 2 3 4 5 6 7 8 9 1
  1   -0.45111    |           *********|    .                |
  2   -0.21922    |                ****|    .                |
  3   -0.20111    |                ****|    .                |
  4   -0.13702    |                .***|    .                |
  5   -0.17714    |                ****|    .                |
  6   -0.10153    |                . **|    .                |
  7   -0.21528    |                ****|    .                |
  8   -0.02310    |                .   |    .                |
  9    0.00033    |                .   |    .                |
 10   -0.02748    |                .  *|    .                |
 11   -0.02437    |                .   |    .                |
 12    0.10797    |                .   |**  .                |
```

(예 4) $\theta_1 = 0.8,\ \theta_2 = -0.5$를 갖는 MA(2) 모형: $z_t = a_t - \theta_1 a_{t-1} - \theta_2 a_{t-2}$

$$z_t = (1 - 0.8B + 0.5B^2)a_t$$



Autocorrelations

Lag	Covariance	Correlation	-1 9 8 7 6 5 4 3 2 1 0 1 2 3 4 5 6 7 8 9 1	Std Error
0	1.940288	1.00000	|********************|	0
1	-1.285670	-.66262	| *************| . |	0.091287
2	0.567961	0.29272	| . |****** |	0.125104
3	-0.140206	-.07226	| . *| . |	0.130687
4	0.129589	0.06679	| . |* . |	0.131020
5	-0.168168	-.08667	| . **| . |	0.131303
6	0.214717	0.11066	| . |** . |	0.131779
7	-0.305054	-.15722	| . ***| . |	0.132551
8	0.322068	0.16599	| . |*** . |	0.134096
9	-0.172170	-.08873	| . **| . |	0.135798
10	0.063393	0.03267	| . |* . |	0.136280
11	-0.095081	-.04900	| . *| . |	0.136345
12	0.245931	0.12675	| . |*** . |	0.136492
13	-0.348114	-.17941	| .****| . |	0.137469

Partial Autocorrelations

Lag	Correlation	-1 9 8 7 6 5 4 3 2 1 0 1 2 3 4 5 6 7 8 9 1
1	-0.66262	| *************| . |
2	-0.26089	| *****| . |
3	-0.00108	| . | . |
4	0.15745	| . |***. |
5	0.03791	| . |* . |
6	0.06399	| . |* . |
7	-0.12397	| . **| . |
8	-0.01479	| . | . |
9	0.07882	| . |** . |
10	0.06409	| . |* . |
11	-0.05643	| . *| . |
12	0.07594	| . |** . |
13	-0.05566	| . *| . |

○ARMA(p, q) 과정은 AR(p) 모형과 MA(q) 모형이 혼합된 모형으로 다음과 같이 표현 가능하다.

$$z_t = \phi_1 z_{t-1} + \ldots + \phi_p z_{t-p} + a_t - \theta_1 a_{t-1} - \ldots - \theta_q a_{t-q}$$
$$z_t - \phi_1 z_{t-1} - \ldots - \phi_p z_{t-p} = a_t - \theta_1 a_{t-1} - \ldots - \theta_q a_{t-q}$$
$$(1 - \phi_1 B - \ldots - \phi_p B^p) z_t = (1 - \theta_1 B - \ldots - \theta_q B^q) a_t$$
$$\phi(B) z_t = \theta(B) a_t$$

○계절성이 없는 ARIMA(p,d,q) 모형은 다음과 같이 표현할 수 있다.

$$(1 - \phi_1 B - \phi_2 B^2 - \ldots - \phi_p B^p)(1-B)^d z_t = (1 - \theta_1 B - \ldots - \theta_q B^q) a_t$$
$$\phi(B) \nabla^d z_t = \theta(B) a_t$$
$$\phi(B)(1-B)^d z_t = \theta(B) a_t$$

(예) ARIMA(1,0,1): $(1 - \phi_1 B) z_t = (1 - \theta_1 B) a_t$

$z_t - \phi_1 z_{t-1} = a_t - \theta_1 a_{t-1}$

ARIMA(1,1,2): $(1 - \phi_1 B)(1-B) z_t = (1 - \theta_1 B - \theta_2 B^2) a_t$

$(1 - \phi_1 B)(z_t - z_{t-1}) = (1 - \theta_1 B - \theta_2 B^2) a_t$

$z_t - z_{t-1} - \phi_1 z_{t-1} + \phi_1 z_{t-2} = a_t - \theta_1 a_{t-1} - \theta_2 a_{t-2}$

○계절성이 있는 ARIMA(P,D,Q)s은 다음과 같이 표현할 수 있다.

$$\Phi(B^s) \nabla_s^D z_t = \Theta(B^s) a_t$$
$$\Phi(B^s)(1-B^s)^D z_t = \Theta(B^s) a_t$$

여기서 $\Phi(B^s) = 1 - \Phi_1 B^s - \Phi_2 B^{2s} - \ldots - \Phi_P B^{Ps}$

$\Theta(B^s) = 1 - \theta_1 B^s - \Theta_2 B^{2s} - \ldots - \Theta_Q B^{Qs}$

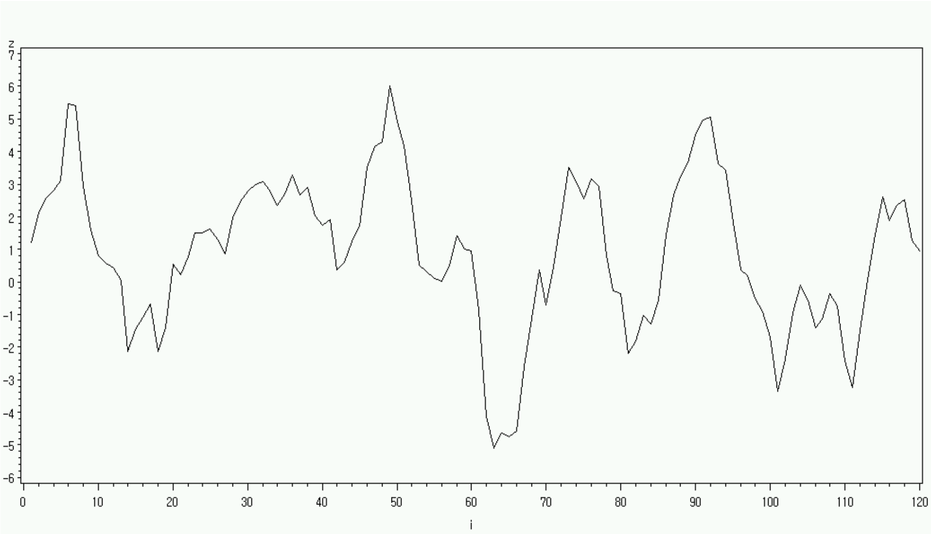
(예 5) 다음은 $\phi = 0.8,\ \theta = -0.5$를 갖는 ARMA(1,0,1)의 예이다.
즉, $z_t = 0.8 z_{t-1} + 0.5 a_{t-1} + a_t$이며, 후향연산자로 표현하면
$(1 - 0.8B) z_t = (1 + 0.5B) a_t$이다.

< SAS PROGRAM >

PROC ARIMA DATA=dataset name;
    IDENTIFY VAR=var.name(d1,d2) NLAG=number;
    ESTIMATE P=lag(s.l) Q=lag(s.l) METHOD=ULS(또는 ML);

```
proc arima;
  identify var=z nlag=15;
  estimate q=1 p=1 noconstant method=uls; run;
```

<SAS 결과>



※ identify var=z; estimate p=1 q=1;

Autocorrelations

Lag	Covariance	Correlation	-1 9 8 7 6 5 4 3 2 1 0 1 2 3 4 5 6 7 8 9 1	Std Error
0	5.345196	1.00000	|********************|	0
1	4.796145	0.89728	| . |****************** |	0.091287
2	3.863797	0.72285	| . |************** |	0.147485
3	2.909238	0.54427	| . |*********** |	0.174529
4	1.884179	0.35250	| . |*******. |	0.188143
5	0.881841	0.16498	| . |*** . |	0.193568
6	-0.032309	-.00604	| . | . |	0.194736
7	-0.694492	-.12993	| . ***| . |	0.194738
8	-0.986453	-.18455	| . ****| . |	0.195459
9	-1.113747	-.20836	| . ****| . |	0.196906
10	-1.221912	-.22860	| . *****| . |	0.198735
11	-1.300554	-.24331	| . *****| . |	0.200914
12	-1.308522	-.24480	| . *****| . |	0.203355
13	-1.154949	-.21607	| . ****| . |	0.205796
14	-0.975712	-.18254	| . ****| . |	0.207678
15	-0.761223	-.14241	| . ***| . |	0.209011

Partial Autocorrelations

```
Lag   Correlation   -1 9 8 7 6 5 4 3 2 1 0 1 2 3 4 5 6 7 8 9 1
  1     0.89728      |                 .  |*****************  |
  2    -0.42209      |          ********|  .                  |
  3     0.00395      |                 .  |  .                |
  4    -0.23537      |             *****|  .                  |
  5    -0.05012      |                 . *|  .                |
  6    -0.11310      |                 .**|  .                |
  7     0.10182      |                 .  |** .               |
  8     0.11908      |                 .  |** .               |
  9    -0.10314      |                 .**|  .                |
 10    -0.11267      |                 .**|  .                |
 11    -0.09736      |                 .**|  .                |
 12    -0.01139      |                 .  |  .                |
 13     0.12299      |                 .  |** .               |
 14    -0.04239      |                 . *|  .                |
 15     0.10920      |                 .  |** .               |
```

Autocorrelation Check for White Noise

To Lag	Chi-Square	DF	Pr > ChiSq	--------Autocorrelations--------					
6	220.10	6	<.0001	0.897	0.723	0.544	0.352	0.165	−0.006
12	255.49	12	<.0001	−0.130	−0.185	−0.208	−0.229	−0.243	−0.245

Parameter	Estimate	Standard Error	t Value	Approx Pr > |t|	Lag
MA1,1	−0.48195	0.08570	−5.62	<.0001	1
AR1,1	0.84593	0.05211	16.23	<.0001	1

Variance Estimate	0.88361
Std Error Estimate	0.940005
AIC	329.8852
SBC	335.4602
Number of Residuals	120

Correlations of Parameter Estimates

Parameter	MA1,1	AR1,1
MA1,1	1.000	0.332
AR1,1	0.332	1.000

Autocorrelation Check of Residuals

To Lag	Chi-Square	DF	Pr > ChiSq	--------Autocorrelations--------					
6	3.42	4	0.4903	0.009	0.075	0.111	−0.003	0.004	−0.095
12	15.70	10	0.1084	−0.241	−0.051	0.010	−0.050	0.014	−0.173
18	18.74	16	0.2825	0.062	−0.065	0.023	−0.010	0.020	0.111
24	20.52	22	0.5508	0.024	0.011	0.001	−0.001	0.074	−0.075

Autoregressive Factors
Factor 1: 1 - 0.84593 B**(1)

Moving Average Factors
Factor 1: 1 + 0.48195 B**(1)

① ACF, PACF에 의해 ARIMA(2 0 0) 가정

Parameter	Estimate	Standard Error	t Value	Approx Pr > |t|	Lag
AR1,1	1.29011	0.08377	15.40	<.0001	1
AR1,2	-0.41533	0.08387	-4.95	<.0001	2

Variance Estimate	0.886436
Std Error Estimate	0.941507
AIC	330.2186
SBC	335.7936
Number of Residuals	120

② ACF가 점진적으로 감소하므로 차분, ARIMA(1 1 1) 가정

Parameter	Estimate	Standard Error	t Value	Approx Pr > |t|	Lag
MA1,1	-0.54069	0.17595	-3.07	0.0026	1
AR1,1	-0.12729	0.20724	-0.61	0.5403	1

Variance Estimate	0.959957
Std Error Estimate	0.979774
AIC	335.0467
SBC	340.6049
Number of Residuals	119

③ 추정된 AR1 계수의 t-값이 작으므로 ARIMA(0 1 1) 가정

Parameter	Estimate	Standard Error	t Value	Approx Pr > |t|	Lag
MA1,1	-0.43263	0.08319	-5.20	<.0001	1

Variance Estimate	0.954264
Std Error Estimate	0.976864
AIC	333.3394
SBC	336.1185
Number of Residuals	119

④ 차분은 AR 모형으로 표현 가능 ARIMA(1 0 1) 가정

Parameter	Standard Estimate	Error	Approx t Value	Pr > |t|	Lag
MA1,1	-0.48195	0.08570	-5.62	<.0001	1
AR1,1	0.84593	0.05211	16.23	<.0001	1

Variance Estimate	0.88361
Std Error Estimate	0.940005
AIC	329.8852
SBC	335.4602
Number of Residuals	120

요 약 표

모형	AIC	추정된 모수(t-값)
ARIMA(2 0 0)	330.22	AR1: 1.29 (15.40) AR2: -0.42 (-4.95)
ARIMA(1 1 1)	335.05	MA1: -0.54 (-3.07) AR1: -0.13 (-0.61)
ARIMA(0 1 1)	333.34	MA : -0.43 (-5.20)
ARIMA(1 0 1)	**329.89**	**MA1: -0.48 (-5.62)** **AR1: 0.85 (16.23)**

ARIMA 모형 식별을 위하여 먼저 SACF, SPACF를 검토한다. SACF인 "Autocorrelations"를 보면 상관관계가 완만하게 감소하고 있어 시계열의 차분이 필요함을 보여주고 있다. SPACF인 “Partial Autocorrelations"은 AR 차수가 2인 AR(2) 모형을 제시하고 있다.

이때 차분된 시계열의 경우, ARIMA(1 1 1)과 ARIMA(0 1 1), 차분하지 않은 시계열의 경우 ARIMA(2 0 0)과 ARIMA(1 0 1) 모형을 비교한다. 물론 다른 형태의 ARIMA 모형도 고려할 수 있으나, 고려한 모형에 대한 추정된 모수의 t-검정시 t-값이 작으므로(p-값이 큼) 비교대상에서 제외하였다. 비교하는 4개의 모형들의 t-값은 ARIMA(1 1 1) 모형의 AR 모수를 제외하고 유의적임을 보여주고 있다. 따라서 이들 모형 중에서 최소 AIC 값을 갖는 ARIMA(1 0 1)를 최적의 모형으로 선정한다.

그러나 ARIMA(1 0 1)의 AR(1) 계수가 0.85로 매우 큼으로 “□ 정상시계열”에서 설명한 바와 같이 AR(1) 형태는 차분 형태로 식별될 수 있어, ARIMA(0 1 1) 모형으로 식별될 수도 있다.

<참고 3.2> Bell-Hillmer 명절효과 변수에 대한 SAS 프로그램

1980년부터 2010년까지의 명절효과 변수 값을 생성하기 위한 프로그램

```
/* *********************************************************************
   PRIOR ADJUST VARIABLE OF KOREA HOLIDAY(SUL, CHUSUK)
  ******************************************************************* */

data kholiday;
  infile 'c:\holiday\korea\kholiday.txt';              우리나라 음력 명절 자료
  input year sm sd cm cd;

  data holiday; set kholiday;
   if year<1980 or year>=2011 then delete;
     ls=sd+31-21; if sm=1 then ls=sd-21;
     lc=cd+30-7;  if cm=9 then lc=cd-7;
  file 'c:\temp\test.dat';
   put year 6.0 ls 4.0  lc 4.0;
 %let sul=SUL4.dat;                                  명절효과 기간(ω)는 4일로 가정
 %let chu=CHUSUK4.dat;
 %let dsn=kh4.dat;
  /*                                          명절효과 변수의 기댓값
 W=1; m1=0.38333; m2=0.61667; m8=       0; m9=0.79167; m10=0.20833;
 W=2; m1=0.40208; m2=0.59792; m8=       0; m9=0.80625; m10=0.19375;
 W=3; m1=0.41667; m2=0.58333; m8=       0; m9=0.82361; m10=0.17639; */

 W=4; m1=0.43229; m2=0.56771; m8=       0; m9=0.83958; m10=0.16042;

/* W=5; m1=0.45000; m2=0.55000; m8=       0; m9=0.85750; m10=0.14250;
  W=6; m1=0.46806; m2=0.53194; m8=       0; m9=0.87569; m10=0.12431;
  W=7; m1=0.48393; m2=0.51607; m8=       0; m9=0.89048; m10=0.10952;
  W=8; m1=0.50052; m2=0.49948; m8=0.00208; m9=0.90208; m10=0.09583;
  W=9; m1=0.51713; m2=0.48287; m8=0.00556; m9=0.90926; m10=0.08519;
 W=10; m1=0.53500; m2=0.46500; m8=0.01333; m9=0.91000; m10=0.07667;
 W=11; m1=0.55189; m2=0.44811; m8=0.02273; m9=0.90758; m10=0.06970;
 W=12; m1=0.56806; m2=0.43194; m8=0.03403; m9=0.90208; m10=0.06389;
 W=13; m1=0.58494; m2=0.41506; m8=0.04551; m9=0.89551; m10=0.05897;
 W=14; m1=0.60179; m2=0.39821; m8=0.05714; m9=0.88810; m10=0.05476;
 W=15; m1=0.61889; m2=0.38111; m8=0.07000; m9=0.87889; m10=0.05111;
 W=16; m1=0.63542; m2=0.36458; m8=0.08307; m9=0.86901; m10=0.04792;
 W=17; m1=0.65221; m2=0.34779; m8=0.09730; m9=0.85760; m10=0.04510;
 W=18; m1=0.66898; m2=0.33102; m8=0.11088; m9=0.84653; m10=0.04259;
```

```
 W=19; m1=0.68553; m2=0.31447; m8=0.12522; m9=0.83443; m10=0.04035;
W=20; m1=0.70125; m2=0.29875; m8=0.14000; m9=0.82167; m10=0.03833;*/
     m2=1-m1; m10=1-m8-m9;    ks=w+11; kc=w+24;
     w1=0.0; w2=0.0; w3=0.0; w4=0.0; w5=0.0; w6=0.0;
     w7=0.0; w8=0.0; w9=0.0; w10=0.0; w11=0.0; w12=0.0;

/********* Lunar New Year **********/                설 변수의 설정
   if ls=11 then do w1=1.0; w2=0.0; end;
     if ls<11 and w<sd then w1=1.0; else if ls<11 and w>=sd then do
                w12=(w-sd+1)/w; w1=(sd-1)/w;    end;
         else if ls>=12 and ls<=ks then do
              w1=(ks-ls)/w; w2=1-w1; end;
           else if ls>=12 and ls>ks then do w1=0; w2=1.0; end;

/******* Mid-August Autumn *********/                추석 변수의 설정
   if lc=24 then do w9=1.0; w10=0.0;  end;
   if lc<24 and w<cd then w9=1.0; else if lc<24 and w>=cd then do
                w8=(w-cd+1)/w; w9=(cd-1)/w;     end;
        else if lc>=25 and lc<=kc then do w9=(kc-lc)/w; w10=1-w9;
                                        end;
    else if lc>=25 and lc>kc then do w9=0.0; w10=1.0; end;

data sul;   set holiday;                        명절 및 이외 월자료의 생성
 sw1=0.0; sw2=0.0; sw3=0.0; sw4=0.0; sw5=0.0; sw6=0.0;
 sw7=0.0;  sw8=0.0;  sw9=0.0;  sw10=0.0;  sw11=0.0;  sw12=0.0;
sw1=w1-m1; sw2=w2-m2;

   file 'c:\holiday\korh1\ts';
   put (sw1 sw2 sw3 sw4 sw5 sw6 sw7 sw8 sw9 sw10 sw11
sw12)(f12.7);   run;
data chusuk;  set holiday;
 cw1=0.0; cw2=0.0; cw3=0.0; cw4=0.0; cw5=0.0; cw6=0.0;
 cw7=0.0; cw8=0.0; cw9=0.0; cw10=0.0; cw11=0.0; cw12=0.0;
   cw8=w8-m8; cw9=w9-m9;cw10=w10-m10;

  file 'c:\holiday\korh1\tc';
   put (cw1 cw2 cw3 cw4 cw5 cw6 cw7 cw8 cw9 cw10 cw11 cw12)
```

```
      (f12.7);    run;
 data t1;
  infile 'c:\holiday\korh1\ts';     input sx@@;
   n=_n_;      year=int((n-1)/12)+1980;
            month=mod(n,12); if month=0 then month=12;
   file "c:\holiday\korh1\&sul";
       put year 5.0 month 4.0 sx 12.7; run;

  data t2;
  infile 'c:\holiday\korh1\tc';    input cx@@;
   n=_n_;       year=int((n-1)/12)+1980;
            month=mod(n,12); if month=0 then month=12;
  file "c:\holiday\korh1\&chu";
        put year 5.0 month 4.0 cx 12.7; run;

data data;
 merge t1 t2;
 file "c:\holiday\korh1\&dsn";
        put year 5.0 month 4.0 sx 12.7 cx 12.7; run;
```

<참고 3.3> RegARIMA 잔차에 대한 t-검정 및 Box-and-Whisker plot SAS 프로그램

```
/*****************************************************************
 *****  USING RegARIMA Rsidual : RSD                        *****
 *****  PAIRED T-TEST                                      *****
*****************************************************************/

/*  수정해야 할 부분:  infile VARI, START    */
                %let VARI=manuf;
                %let START=1985;
data rsd;
 infile "C:\X12A03\SA05\&vari..rsd" firstobs=3;
 input year 4.0 month 2.0 rsd;
data irr;
 infile "c:\x12a03\sa05\&vari..rsd" firstobs=3;
 input year 4.0 MONTH 2.0 irr;
 TITLE h=2.0 "******** &vari  *************";
data test;   merge rsd irr;
data mon1; set test; if month ne 1 then delete; rsd1=rsd;
data mon2; set test; if month ne 2 then delete; rsd2=rsd;
data mon3; set test; if month ne 3 then delete; rsd3=rsd;
data mon4; set test; if month ne 4 then delete; rsd4=rsd;
data mon5; set test; if month ne 5 then delete; rsd5=rsd;
data mon6; set test; if month ne 6 then delete; rsd6=rsd;
data mon7; set test; if month ne 7 then delete; rsd7=rsd;
data mon8; set test; if month ne 8 then delete; rsd8=rsd;
data mon9; set test; if month ne 9 then delete; rsd9=rsd;
data mon10; set test; if month ne 10 then delete; rsd10=rsd;
data mon11; set test; if month ne 11 then delete; rsd11=rsd;
data mon12; set test; if month ne 12 then delete; rsd12=rsd;
 data lslc;
  infile 'c:\holiday\korea\lslc.txt';  /* 명절 일자 분포임 */
  input year ls lc;
  if year lt &start then delete;

data holiday;
```

```
 merge mon1 mon2 mon3 mon4 mon5 mon6 mon7 mon8 mon9 mon10
mon11 mon12 lslc;by year;
 /* ****************** LAST YEAR? *************************** */
if year gt 2004 then delete;
           dhrs=rsd2-rsd1; dhrc=rsd10-rsd9;
           hrs=rsd1; nhrs=rsd2;  hrc=rsd9; nhrc=rsd10;
   if ls lt 11 then dhrs=rsd1-rsd2;
       if ls lt 11 then do hrs=rsd1; nhrs=rsd2;  end;
       if ls ge 11 then do hrs=rsd2; nhrs=rsd1;  end;
   if lc lt 24 then dhrc=rsd9-rsd10;
       if lc lt 24 then do hrc=rsd9; nhrc=rsd10;  end;
       if lc ge 24 then do hrc=rsd10; nhrc=rsd9;  end;

   if ls lt 11 then dhis=irr1-irr2; if ls ge 11 then dhis=irr2-irr1;
   if lc lt 24 then dhic=irr9-irr10; if lc ge 24 then dhic=irr10-irr9;
  file "c:\mks\sim\&vari._rsd.dat";
    put year 4.0 (rsd1-rsd12)(12.7);
file "c:\mks\sim\&vari._HRS.dat";
 put year 4.0 (hrs nhrs rsd3-rsd8 hrc nhrc rsd11 rsd12)(12.7);
file "c:\mks\sim\test.dat";
  put '  ' "HS"  hrs 20.8 '   ' "NS" nhrs 20.8 '   ' "HC" hrc 20.8 '   '
"NC" nhrc 20.8; run;
/*  *********** T-test *********      */
 data tests;
 infile 'c:\mks\sim\test.dat';     input class$ value @@;
    if class="HC" or class="NC" then delete;
     TITLE " ********* TEST SUL EFFECT for &vari ********";
      proc ttest; class class; var value;
data testc;
  infile 'c:\mks\sim\test.dat';  input class$ value @@;
    if class="HS" or class="NS" then delete;
    TITLE " ********* TEST CHUSUCK EFFECT for &vari ********";
    proc ttest; class class; var value; run;
/*  ************   BOX PLOT   *************  */
data BOX;
 infile "c:\mks\sim\&vari._HRS.dat";
   input year @;
```

```
  do month=1 to 12;
     input value @;
     output;
     end;
proc sort out=sort; by month;
  symbol color = salmon h = .8;
  goptions ftext=swiss;
  axis1 minor=none color=black label=(angle=90 rotate=0);
    title "Series Name(&vari): Holiday(1, 9) and Non-Holiday(2, 10)";
     proc boxplot data=sort;
     plot value*month/boxstyle=skhletal
                      cframe   = vligb
                      cboxes   = dagr
                      cboxfill = ywh
                      vaxis    = axis1;
  run;
```

Ⅳ. 계절조정

4.1 X-11의 계절조정 절차

X-12-ARIMA는 A, B, C, D 반복 절차에 의해서 사전조정(prior adjustment) 및 계절조정(seasonal adjustment)이 실시된다. 여기서 A 반복 절차는 RegARIMA 모형에 의해 사전조정이 되는 단계이며, B, C, D는 X-11 방법에 의해 계절조정이 되는 단계이다. 상세한 단계별 과정은 "<참고 4.2> X-12-ARIMA의 산출과정"를 참조하기 바란다.

표 B1은 특이치, 요일효과, 명절효과 등의 사전조정 요인을 원계열에서 계절조정하기 전에 조정한 사전조정계열이다. X-11 방법에 의한 계절조정은 표 B1 계열부터 시작한다.

B 반복절차가 끝나면, 표 C1를 작성하기 위하여 사전조정된 원계열(표 B1)을 표 B20으로 나눈다. C 반복절차가 끝나면, 표 D1를 작성하기 위하여 원계열(표 C1)을 표 C20으로 나눈다. 각 절차의 표는 X11 명령의 print 옵션에 의해서 Output Screen에서 볼 수 있으며, save 옵션에 의해서 파일로 저장할 수 있다. 여기서 추세치(요인)은 추세·순환치(요인)를 지칭한다. 통계조사국 통계분석과(1995)은 실제 자료에 의해 각 반복 절차를 자세히 설명하고 있다.

Statge 1 초기 추정치
1.1 초기 추세치
1.2 초기 추세치가 제거된 계열
1.3 극단값(extreme value)의 대체
1.4a 추세가 제거된 시계열의 월이 평활화된 계열에 대한 초기 "biased" 계절요인
1.4b 초기 "unbiased" 계절요인
1.5 Stage 1의 계절조정된 계열(계절조정계열)

Step 1.1 초기 추세치(표 B2, C2, D2)
사전조정된 원계열에서 계절성을 제거하여야 하기 때문에 다음과 같이 중

심화 12-개항 이동평균을 하여 초기 추세치를 추정한다.

$$T_t^{(1)} = T_t^{2\times 12} = \frac{1}{24}Y_{t-6} + \frac{1}{12}Y_{t-5} + \ \dots \ + \frac{1}{12}Y_t + \ \dots \ + \frac{1}{12}Y_{t-5} + \frac{1}{24}Y_{t+6}$$

(예) 2×12 추세 필터: 98년 1월부터 2000.12까지 중심화 12개월 이동평균의 예

$$\frac{1}{24}\begin{pmatrix} 98.1 + & 98.2 + & \dots & + 98.7 + & \dots & + 98.12 & \\ & 98.2 + & \dots & + 98.7 + & \dots & + 98.12 + & 99.1 \end{pmatrix} = 98.7$$

Step 1.2 초기 추세치가 제거된 계열(표 B3, C4, D4)

원계열에 대한 초기 추세치의 비율을 계산한다. 이 비율을 계절-불규칙요인 “SI-비율”이라 한다.

$$SI_t^{(1)} = \frac{Y_t}{T_t^{(1)}}$$

Step 1.3 SI 비율의 극단값(extreme value)의 대체(표 B4)

SI 비율의 극단값은 다음과 같은 절차에 의해 대체값이 계산된다:

① 계절요인을 추정한다.
② SI로부터 계절요인을 제거함으로써 불규칙요인을 추정한다.
③ 이동 표준편차를 계산한다.
④ 표준편차의 함수를 기초로 가중치를 정한다.
⑤ 같은 월에 대해서 4개의 가장 가까운 값과 해당 월의 값을 기초로 극단값을 조정한다.

극단값에 대한 대체값 추정은 “<참고 4.1> 극단값 조정절차”에서 자세히 다루도록 한다.

Step 1.4a 추세가 제거된 초기 “biased" 계절요인

초기 계절요인을 구한다. 필요하다면, Step 1.3의 대체된 SI 비율에 대해서 5-개항 가중이동평균(3 × 3)을 적용한다.

$$S_t^{'(1)} = S_t^{3\times 3} = \frac{1}{9}SI_{t-24} + \frac{2}{9}SI_{t-12} + \frac{3}{9}SI_t + \frac{2}{9}SI_{t+12} + \frac{1}{9}SI_{t+24}$$

(예) 1990년 1월에 대한 3×3 filter의 예

$$\frac{1}{9}\left(\begin{array}{lllll} 1998.1 + & 1999.1 + & 2000.1 + & & \\ & 1999.1 + & 2000.1 + & 2001.1 + & \\ & & 2000.1 + & 2001.1 + & 2002.1 \end{array}\right) = 2000.1$$

Step 1.4b 초기 “불편위” 계절요인(표 B5, C5, D5)
Step 1.4a 결과를 중심화 12-개항 이동평균법에 의해 초기 “불편위 (unbiased)" 계절요인을 계산한다.

$$S_t^{(1)} = \frac{S_t^{'(1)}}{\frac{1}{24}S_{t-6}^{'(1)} + \frac{1}{12}S_{t-5}^{'(1)} + \ldots + \frac{1}{12}S_t^{'(1)} + \ldots + \frac{1}{12}S_{t+5}^{'(1)} + \frac{1}{24}S_{t+6}^{'(1)}}$$

※ 계절요인을 중심화하는 이유는 가능하면 계절조정계열의 연간합과 원계열의 연간합이 일치하도록 만들어 주기 위함이다.

Step 1.5 Stage 1의 계절조정계열(표 B6, C6, D6)
초기 계절조정계열을 계산한다.

$$A_t^{(1)} = \frac{Y_t}{S_t^{(1)}}$$

Stage 2
2.1 Henderson 추세 필터에 의해 초기 추세치를 구한다.
2.2 Henderson 추세를 제거하여 추세가 제거된 계열을 구한다.
2.3 극단값을 대체한다.
2.4a 추세가 제거된 계열의 월간 평활로부터 편위 계절요인을 구한다.
2.4b 불편위 계절요인을 구한다.
2.5 Stage 2의 계절조정계열을 구한다.
B와 C 반복절차에서만 다음 단계를 수행한다.
2.6 불규칙요인을 추정한다.
2.7 불규칙요인에 대한 가중치를 구한다.
2.8 극단값 조정요인을 구한다.

Step 2.1 초기 추세요인(표 B7, C7, D7)
길이 2H+1의 핸더슨 추세필터에 의해 중간단계의 추세치를 구한다.

$$T_t^{(2)} = \sum_j h_j^{(2H+1)} A_{t+j}^{(1)}$$

핸더슨 추세필터는 noise-signal 비율인 $\bar{I}/\bar{C}$(여기서 C는 추세순환요인)에 의해서 결정한다.

- $\bar{I}/\bar{C}$ <1.0 이면, 9개항 필터(H=4)
- 1.0 $\leq \bar{I}/\bar{C}$ <3.5 이면, 13개항 필터(H=6)
- 3.5 $\leq \bar{I}/\bar{C}$ 이면, 23개항 필터(H=11)

Step 2.2 추세요인이 제거된 계절·불규칙요인(표 B8, C9, D8)
원계열에서 추세요인을 제거함으로써 SI 비율을 계산한다.

$$SI_t^{(2)} = \frac{Y_t}{T_t^{(2)}}$$

Step 2.3 SI 비율의 극단값 대체(B 반복에서만 적용, 표 B9)
극단값에 대한 대체값을 다음 절차에 의해 구한다:

① 계절요인을 추정한다.
② SI로부터 계절요인을 제거함으로써 불규칙요인을 추정한다.
③ 이동 표준편차를 계산한다.
④ 표준편차의 함수를 기초로 가중치를 정한다.
⑤ t월과 인근의 4개년도 t월 값을 기초로 극단값 조정

극단값과 대체값의 추정은 “<참고 4.1> 극단값 조정절차”에서 자세히 다루도록 한다.

Step 2.4 극단값 대체(D 반복에서만 적용, 표 D9)
SI 비율의 극단값에 대한 최종 대체값을 구하기 위하여, 표 D8의 SI 비율(Step 2.2: 원계열에서 초기 추세요인이 제거된 SI 비율임)과 극단값이 조정된 원계열인 표 D1계열을 추세로 나누어 구한 SI 비율과 비교를 한다. 두 값이 차이가 난다면, 차이가 있는 값은 대체가 되며 표 D9과 같이 인쇄된다.

Step 2.4a　편위 계절요인(B와 C 반복에서만 적용)
Stage 2의 계절요인을 3×5 계절이동평균법을 이용하여 구한다.

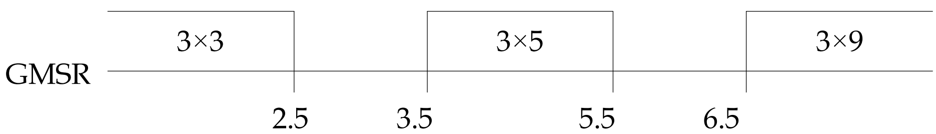
$$
\begin{aligned}
S_t^{'(2)} &= \hat{S}_t^{3\times 5} \\
&= \frac{1}{15}SI_{t-36}+\frac{2}{15}SI_{t-24}+\frac{3}{15}SI_{t-12}+\frac{3}{15}SI_t+\frac{3}{15}SI_{t+12}+ \\
&\quad \frac{2}{15}SI_{t+24}+\frac{1}{15}SI_{t+36}
\end{aligned}
$$

(예) 1990년 1월에 대한 3×5 filter의 예

$$
\frac{1}{15}\left(\begin{array}{l}
1998.1+1999.1+2000.1+2001.1+2002.1+ \\
\qquad 1999.1+2000.1+2001.1+2002.1+2003.1+ \\
\qquad\qquad 2000.1+2001.1+2002.1+2003.1+2004.1
\end{array}\right) = 2001.1
$$

Step 2.4b　불편위 계절요인(D 반복에서만 적용)
다음과 같은 전체 이동계절성비율(Global Moving Seasonality Ratio: GMSR, <참고 4.2> 참조)에 의해 결정되는 계절이동평균으로 최종 계절요인을 구한다.

· $GMSR \le 2.5$ 이면 3×3 (5-개항 계절이동평균)
· $3.5 < GMSR \le 5.5$ 이면 3×5 (7-개항 계절이동평균)
· $GMSR > 6.5$ 이면 3×9 (11-개항 계절이동평균)



GMSR가 gray($2.5<GMSR\le3.5$ or $5.5<GMSR\le6.5$)범위이면, X-12-ARIMA는 시계열 끝 시점에서부터 1년간의 관측치를 제외하고 GMSR를 다시 계산한다. GMSR이 여전히 gray 범위에 있으면, 1년씩 최대 5년까지 제외하여 gray 범위에서 벗어날 때까지 재반복 계산한다. 최대 5년까지 제외하였으나 GMSR이 gray 범위에 있으면 3×5 계절이동평균법을 사용한다.

Step 2.4c　불편위 계절요인(표 B10, C10, D10)

중심화 12-개항 이동평균법에 의해 불편위 계절요인을 계산한다(Step 1.4b와 유사). 최종 불편위 계절요인은 GMSR에 따라 계절이동평균법을 적용하여 구한다.

Step 2.5 계절조정계열(표 B11, C11, D11)
Stage 2단계의 계절조정계열을 계산한다.

$$A_t^{(2)} = \frac{Y_t}{S_t^{(2)}}$$

Step 2.6 불규칙요인 추정(표 B13, C13)
Step 2.5의 계절조정계열을 헨더슨 추세필터에 의해서 구한 Step 2.1의 초기 추세요인으로 나누어 불규칙요인을 구한다.

$$I_t^{(2)} = \frac{A_t^{(2)}}{T_t^{(2)}}$$

Step 2.7 불규칙요인에 대한 가중치(표 B17, C17)
시계열의 변동에 따른 극단값의 가중치를 계산한다(극단값의 추정은 <참고 4.1> 참조).

Step 2.8 극단값 조정요인(표 B20, C20)
Step 2.7의 가중치를 이용하여 극단값 조정요인을 다음과 같이 계산한다. 임의의 t에 대해서, X_t를 다음과 같이 정의하자.

$$X_t = \frac{I_t^{(2)}}{1 + W_t(I_t^{(2)} - 1)}$$

이때, 극단값 조정계열(Y_t^{eva})은 극단값의 가중치에 따라 다음과 같이 구한다.

$$W_t = 1 \text{이면, } Y_t^{eva} = Y_t$$
$$W_t < 1 \text{ 이면, } Y_t^{eva} = \frac{Y_t}{X_t}$$

극단값이 조정된 계열은 다음 반복절차의 시작점이 된다. 즉, 극단값 조정

계열인 표 B20 이후의 반복에서는 표 C1이 되며, 표 C20 이후 절차에서는 표 D1이 된다.

Stage 3(D 반복에서만 적용)
3.1 핸더슨 추세필터로 추세 추정
3.2 최종 불규칙요인 추정

Step 3.1 최종 추세요인(표 D12)
최종 추세요인을 다음과 같이 추정한다.

$$T_t^{(3)} = \sum_j h_j^{(2H+1)} A_{t+j}^{(2)}$$

Step 3.2 최종 불규칙요인(표 D13)
다음과 같이 Stage 2의 계절조정계열을 Step 2의 최종 추세요인으로 나눔으로써 최종 불규칙요인을 계산한다.

$$I_t^{(3)} = \frac{A_t^{(2)}}{T_t^{(3)}}$$

다음은 시계열의 최종 요인 분해 형태이다.

$$Y_t = T_t^{(3)} S_t^{(2)} I_t^{(3)}$$

4.2. 이동평균방법

X-12-ARIMA에 적용된 X-11의 이동평균법은 원계열(A1 또는 B1)을 추세 · 순환, 계절, 불규칙요인으로 각각 분해하는 계산과정으로 시계열의 특성을 고려하여 이용자가 여러 가지 이동평균기법을 선택적으로 적용할 수 있도록 되어 있다.

계절변동 요인들을 산출하기 위한 분해과정에서 사용되는 이동평균법은 대칭가중이동평균과 비대칭가중이동평균으로 구분할 수 있다. 대칭이동평균은 (2n+1)개월 이동평균을 할 때 중앙항인 n항에 위치하며, 비대칭이동평균은 처음과 끝의 n개 관측치를 추정하는데 사용된다. 이 두 가지 이동평균의 가중치의 합은 1이므로, 이동평균을 해도 원계열의 평균은 변하지 않는다.

이동평균을 할 때 원계열의 구성요인과 비교하여 계절조정계열의 수축국면과 확장국면이 시간적으로 바뀌지 않도록 하는 것이 중요하다. 즉, 원계열의 정점이 계절조정계열에서는 저점으로 바뀌는 등의 국면변화가 없어야 한다.

4.2.1 중심화 12개월 이동평균(Centered 12-term moving average)

중심화 12개월 이동평균은 추세 · 순환요인을 추정하기 위해 사용한다. 이 이동평균은 선형추세의 중심값을 정확히 재산출하고 가법모형에서 12개월 주기로 나타나는 안정계절성을 제거할 수 있으며, 승법모형에서는 안정계절성이 곱해진 상수추세를 정확히 재산출할 수 있다.

12개월 이동평균의 큰 단점은 2 · 3년의 단기주기의 순환요인의 경기순환점을 놓칠 우려가 있고, 불규칙 요인이 상대적으로 크면 12개월 이동평균을 해도 충분히 평활화되지 않는다는 점이다. 주기가 3년이고 진폭이 100인 사인(Sine)곡선의 순환요인을 갖는 계열을 12개월 이동평균을 한다면 주기는 변함이 없으나 진폭이 82.5로 감소한다. 2년 주기의 곡선에 대해서는 진폭이 75로 감소하고, 5년 이상 주기 곡선의 진폭은 매우 작게 감소한다. 그러나 대부분의 경제시계열은 40개월 이상의 장기 순환변동으로 이루어져 있으므로 12개월 이동평균은 일반적으로 추세 · 순환요인을 잠정 추정하는데 유용하다.

$$\overline{X}_7 = \frac{1}{2}\left(\frac{X_{i1} + \ ... \ + X_{i12}}{12} + \frac{X_{i2} + \ ... \ + X_{(i+1)1}}{12}\right)$$

4.2.2 중심화 24개월 이동평균(Centered 24-term moving average)

중심화 24개월 이동평균은 주로 2·3년 주기의 짧은 순환변동에 의해 좌우되거나 추세요인이 급격하게 변동하는 계열에서 사용된다. 중심화 24개월 이동평균에 의해 2~3년 주기의 순환변동을 갖는 계열의 진폭을 각각 18%, 5%를 감소시키며, 중심화 12개월 이동평균을 사용했을 때보다 불규칙변동을 더 제거할 수 있다. 24개월 이동평균을 사용하면 양단의 12개항은 결항이 된다. 결항항의 보정방법은 이동평균항에 가까운 6개월은 비대칭형의 다음과 같은 가중치가 사용되고, 양단의 6개항은 결항이 된다. 가중치는 다음의 표와 같다.

월	t-24	t-23	t-22	t-21	t-20	t-19	t-18	t-17	t-16	t-15	t-14	t-13
t-6	0	.019	.011	-.002	-.016	-.026	-.031	-.027	-.017	-.001	.019	.044
t-7	0	.018	.005	-.011	-.024	-.029	-.027	-.018	-.002	.017	.041	.065
t-8	0	.012	-.004	-.018	-.026	-.025	-.018	-.004	.014	.034	.057	.081
t-9	0	.002	-.012	-.021	-.023	-.016	-.004	.012	.030	.047	.069	.093
t-10	0	-.011	-.020	-.021	-.014	-.003	.012	.027	.043	.059	.081	.102
t-11	0	-.022	-.023	-.016	-.003	.012	.028	.041	.056	.073	.091	.105
t-12	-.011	-.027	-.019	-.005	.011	.027	.042	.056	.072	.089	.103	.111

월	t-12	t-11	t-10	t-9	t-8	t-7	t-6	t-5	t-4	t-3	t-2	t-1	t
t-6	.070	.064	.072	.085	.100	.110	.114	.110	.100	.084	.064	.039	.014
t-7	.090	.065	.079	.094	.107	.113	.111	.101	.085	.056	.042	.018	-.006
t-8	.106	.071	.087	.101	.110	.109	.101	.087	.069	.049	.027	.003	-.022
t-9	.115	.081	.096	.105	.106	.099	.087	.072	.054	.036	.014	-.010	-.032
t-10	.118	.094	.103	.104	.098	.086	.071	.056	.040	.024	.002	-.019	-.034
t-11	.115	.106	.107	.099	.086	.071	.055	.042	.027	.010	-.008	-.023	-.031
t-12	.106	.111	.103	.089	.072	.056	.042	.027	.011	-.005	-.019	-.027	-.011

4.2.3 헨더슨 이동평균(Henderson moving average)

헨더슨 이동평균은 보험통계에서 사용된 적분공식에 의해 개발된 것으로, 헨더슨 이동평균에 의하여 얻어진 평활된 값들은 불규칙요인이 제거된 값들이다. 2차·3차 포물선에 헨더슨 이동평균을 적용하면 평활된 값들은 포물선 상에 있으며, 확률적인 자료에 핸더슨 이동평균을 적용할 때에도 최소자승법에 의해 추정된 2차 포물선의 계수보다 더 평활화된 결과를 산출한다. 헨더슨 이동평균의 가중치는 1916년 헨더슨(Robert Henderson)이 한 계열에서 3개월의 시차를 갖는 계열들의 자승합이 최소가 되는 값을 산출하였다.

헨더슨 이동평균은 잠정계절조정계열에서 추세·순환요인의 보다 개선된 추정치를 얻기 위해 적용된다. 헨더슨 이동평균하여 얻은 값은 가중최소자승법에 의해 추정한 3차 다항분포의 중앙값을 평활하여 얻어지는 것과 같은 결과이다. 추세·순환이 1～2년의 단기간의 포물선을 따른다는 가정은 이러한 이동평균이 경제시계열에 적합하도록 한다.

X11ARIMA에 사용된 헨더슨 이동평균은 계절조정계열에 적용하는 것이며, 1년 이상의 진폭을 가지는 곡선에서 유용하다. 가장 많이 사용되는 헨더슨 13개월 이동평균은 20개월 이상의 주기를 나타내는 추세·순환요인의 진폭을 축소시키지는 않는다. 그러나 6개월 이하의 단기주기의 불규칙요인을 제거하기 위한 좋은 방법이다.

헨더슨 이동평균은 잠정 계절조정계열에서 추세·순환요인과 불규칙요인을 분리하기 위하여 사용된다. 이동평균 항수를 길게 할수록 평활화되어 불규칙요인은 명확히 제거되나, 반대로 추세·순환요인이 명확하게 나타나지 않게 된다. 따라서 추세·순환요인과 불규칙요인의 상대적 크기($\bar{I}/\bar{C}$)를 계산하여 그 결과에 따라 이동평균 항수를 정한다. 다음 표는 추세·순환요인($\bar{C}$)과 불규칙요인($\bar{I}$)의 상대적 크기($\bar{I}/\bar{C}$)에 따른 헨더슨 이동평균 항수와 가중치이다. 분기 계열인 경우에는 추세·순환요인과 불규칙요인의 상대적 크기에 관계없이 헨더슨 5개항 이동평균을 사용한다.

$\bar{I}/\bar{C}$	헨더슨 이동평균개월수
0.00 – 0.99	9개월 이동평균
1.00 – 3.49	13개월 이동평균
3.50 이상	23개월 이동평균

헨더슨 9개월 이동평균 가중치

월	t-8	t-7	t-6	t-5	t-4	t-3	t-2	t-1	t
t	0	0	0	0	-.156	-.034	.185	.424	.581
t-1	0	0	0	-.049	-.011	.126	.282	.354	.298
t-2	0	0	-.022	.000	.120	.259	.315	.242	.086
t-3	0	-.031	-.004	.120	.263	.324	.255	.102	-.029
t-4	-.041	-.010	.119	.267	.330	.267	.119	-.010	-.041

헨더슨 13개월 이동평균 가중치

월	t-12	t-11	t-10	t-9	t-8	t-7	t-6	t-5	t-4	t-3	t-2	t-1	t
t	0	0	0	0	0	0	-.092	-.058	.012	.120	.244	.353	.421
t-1	0	0	0	0	0	-.043	-.038	.002	.080	.174	.254	.292	.279
t-2	0	0	0	0	-.016	-.025	.003	.068	.149	.216	.241	.216	.148
t-3	0	0	0	-.009	-.022	.004	.066	.145	.208	.230	.201	.131	.046
t-4	0	0	-.011	-.022	.003	.067	.145	.210	.235	.205	.136	.050	-.018
t-5	0	-.017	-.025	.001	.066	.147	.213	.238	.212	.144	.061	-.006	-.034
t-6	-.019	-.028	.000	.066	.147	.214	.240	.214	.147	.066	.000	-.028	-.019

헨더슨 23개월 이동평균 가중치

월	t-10	t-9	t-8	t-7	t-6	t-5	t-4	t-3	t-2	t-1	t
t	-.064	-.049	-.028	.002	.039	.084	.133	.182	.227	.263	.288
t-1	-.035	-.024	-.004	.025	.061	.101	.141	.176	.203	.219	.224
t-2	-.019	-.005	.018	.049	.082	.116	.146	.166	.177	.176	.166
t-3	-.004	.015	.042	.073	.103	.129	.147	.154	.150	.134	.112
t-4	.015	.040	.068	.098	.121	.137	.142	.136	.119	.095	.006
t-5	.039	.067	.095	.119	.134	.139	.131	.114	.088	.059	.027
t-6	.068	.096	.118	.134	.138	.132	.114	.089	.059	.027	.001
t-7	.096	.120	.135	.140	.133	.116	.090	.060	.031	.005	.015
t-8	.120	.137	.140	.136	.118	.094	.064	.034	.008	-.010	-.021
t-9	.138	.143	.137	.120	.095	.067	.037	.011	-.007	-.017	-.019
t-10	.144	.138	.122	.097	.068	.039	.013	-.005	-.015	-.016	-.011
t-11	.138	.122	.097	.068	.039	.013	-.005	-.015	-.016	-.111	-.004

월	t-22	t-21	t-20	t-19	t-18	t-17	t-16	t-15	t-14	t-13	t-12	t-11
t	0	0	0	0	0	0	0	0	0	0	0	-.077
t-1	0	0	0	0	0	0	0	0	0	0	-.046	-.041
t-2	0	0	0	0	0	0	0	0	0	-.022	-.025	-.025
t-3	0	0	0	0	0	0	0	0	-.008	-.014	-.018	-.015
t-4	0	0	0	0	0	0	0	-.001	-.008	-.013	-.012	-.003
t-5	0	0	0	0	0	0	.003	-.006	-.011	-.011	-.002	.015
t-6	0	0	0	0	0	.002	-.006	-.012	-.011	-.003	.015	.039
t-7	0	0	0	0	.001	-.007	-.013	-.011	-.003	.015	.039	.068
t-8	0	0	0	-.002	-.007	-.013	-.013	-.003	.014	.039	.068	.097
t-9	0	0	-.003	-.010	-.015	-.014	-.005	.014	.040	.069	.097	.122
t-10	0	-.004	-.011	-.016	-.015	-.005	.013	.039	.068	.097	.122	.138
t-11	-.004	-.011	-.016	-.015	-.005	.013	.039	.068	.097	.122	.138	.148

4.2.4 가중 5개항(3×3) 이동평균과 가중 7개항(3×5) 이동평균 (Weighted 5-term & 7-term moving average)

가중 5개항 이동평균은 각 월을 연도별로 3개항 이동평균한 계열을 다시 3개항 이동평균(3×3 이동평균)하는 방법이다. 마찬가지로 가중 7개항 이동평균은 각 월을 연도별로 5개항 이동평균한 계열을 다시 3개항 이동평균(3×5 이동평균)하는 방법이다. 이 두 이동평균은 각 월의 계절·불규칙요인에 각각 적용되고, 이때 사용한 가중치는 모두 양수이며 이동평균 기간 내의 일직선상에 있다. 이로써 5년과 7년의 연간 계절성의 변화를 선형화할 수 있다. 따라서 이러한 이동평균은 7년 이상의 계열에 대해 비선형 추세를 따르는 점진적인 계절성의 변화율을 대략적으로 계산할 수 있다.

가중 5개항 이동평균은 일정한 방향으로 급격한 변화를 하는 시계열에 대해 매우 신축성있게 적용되지만, 이동평균 기간이 짧아 계절·불규칙요인이 잘 평활화 되려면 불규칙요인이 작아야 한다. 반면, 가중 7개항 이동평균은 보다 더 신축적인 최종계절요인을 추정하는데 사용된다. 불규칙요

인이 큰 계열에 대해서 더 긴 이동평균 기간에 적용할 가중치를 제공하며, 이에 따라 더욱 평활된 계절요인을 산출한다.

이동평균법의 항수선택은 Lothian의 방법(Friendley와 Monsell, 1990)에 따라 불규칙요인과 계절요인의 상대적 크기인 이동계절성의 변화율(Moving Seasonality Ratio: MSR)인 I/S에 따른다. 5년 이상의 전체 계열에 대해 N년(12월까지 수치가 있는 가장 마지막 해)을 포함하여 계산된 I/S_N을 이용하여 이동평균항수를 선택한다.

I/S_N비	이동평균항수
2.5 이하	5개항(3×3)이동평균
3.5～5.5	7개항(3×5)이동평균
6.5 이상	11개항(3×9)이동평균

I/S_N비율이 2.5～3.5 또는 5.5～6.5이면, 계열의 마지막 년도는 제외하고 I/S_{N-1}비를 계산하여 위의 과정을 재실행한다. 그러나 만약 이 비율이 여전히 2.5～3.5 또는 5.5～6.5이면, 최근 5년간(backcasting된 수치는 제외)의 자료로 I/S_{N-2}비를 계산한다. 이 I/S_{N-2}비가 다시 위의 조건에 부적합하다면, 3×3이동평균을 사용한다. 이러한 I/S_N비율을 이용한 이동평균항수의 선택옵션은 5년 미만의 계열에서는 적용되지 않는다.

다음 표들은 MSR에 따라 결정되는 이동평균의 가중치들이다. 표에서 N은 시계열의 가장 마지막 년을 나타내고, N+1의 가중치는 N항에서 1년을 예측한 계절요인의 가중치를 나타낸다.

11개항(3×9) 이동평균 가중치

년도	N-10	N-9	N-8	N-7	N-6	N-5	N-4	N-3	N-2	N-1	N
N+1	0	0	0	0	-.014	.031	.096	.180	.208	.236	.265
N	0	0	0	0	0	.051	.112	.173	.197	.221	.246
N-1	0	0	0	0	.028	.092	.144	.160	.176	.192	.208
N-2	0	0	0	.032	.079	.123	.133	.143	.154	.163	.173
N-3	0	0	.034	.075	.113	.117	.123	.128	.132	.137	.141
N-4	0	.034	.073	.111	.113	.114	.116	.117	.118	.120	.084
N-5	.037	.074	.111	.111	.111	.111	.111	.111	.111	.074	.037

7개항(3×5) 이동평균 가중치

년 도	N-6	N-5	N-4	N-3	N-2	N-1	N
N+1	0	0	-.034	.134	.300	.300	.300
N	0	0	0	.150	.283	.283	.283
N-1	0	0	.067	.183	.250	.250	.250
N-2	0	.067	.133	.217	.217	.217	.150
N-3	.067	.133	.200	.200	.200	.133	.067

5개항(3×3) 이동평균 가중치

년 도	N-4	N-3	N-2	N-1	N
N+1	0	-.056	.148	.426	.481
N	0	0	.185	.407	.407
N-1	0	.111	.259	.370	.259
N-2	.111	.222	.333	.222	.111

4.3 계절성 검정(Seasonality test)

계절조정 시, 원계열이 계절성을 갖고 있느냐하는 문제는 중요한 의미를 갖는다. 계절성은 매년 주기적으로 반복되어 나타나는 형태일 수도 있고, 산업구조나 생활 양식의 변화로 매년마다 계절성이 조금씩 변화될 수도 있다. 주기적으로 반복되는 계절성을 안정계절성(stable seasonality)이라 하며, 변화하는 계절성을 이동계절성(moving seasonality)이라 한다.

계절성 존재여부를 파악하기 위한 방법은 그래프적인 방법과 통계량에 의한 F-검정과 Kruskal-Wallis 검정이 있다. 그래프적인 방법은 원계열이나 스펙트럼을 그려봄으로서 계절성 존재 여부를 쉽게 파악할 수 있다. Cleveland와 Devlin(1980)의 스펙트럼분석에 의해 원계열이 갖고 있는 계절성 및 요일효과를 파악할 수 있다. 이 방법은 X-12-ARIMA 뿐만 아니라 TRAMO/SEATS 프로그램에도 적용하고 있다. 원계열에 계절성이 있는 경우에는 스펙트럼 그림의 계절주기에서 정점을 보이게 된다. 한편 요일효과가 있는 경우에는 요일주기에서 정점을 보이게 된다(자세한 내용은

"4.4.2 스펙트럼분석"에서 다룬다).

통계량에 의한 검정은 원계열에 계절성이 존재한다면, 추정된 계절·불규칙요인(SI)의 월(분기)별 평균 간에 차이가 있게 된다는 점을 이용한다. 이때 안정계절성 검정에는 일원분류 분산분석(one-way ANOVA)의 F-검정과 Kruskal-Wallis의 비모수적 검정방법이 이용된다. 한편 이동계절성은 월간 뿐만 아니라 연간 평균 간의 차이를 나타내게 되며 이원분류 분산분석(two-way ANOVA)의 F-검정이 이용된다(Dagum(1988), Higginson (1975)).

4.3.1 F-검정

X-12-ARIMA 결과 중 B1표(사전조정된 원계열)에서는 사전조정된 원계열에 대한 안정계절성 존재에 대한 검정을 하고 있으며, D8에서는 최종 계절·불규칙요인(SI)의 비율(차이)에 의한 안정계절성과 이동계절성의 존재여부를 검정한다. D11에서는 계절조정계열에 대한 전체 시계열 뿐만 아니라 최근 3년의 잔차 계절성 검정을 해 준다. 이때 검정을 위한 귀무가설은 다음과 같다. B1에 대한 계절성 검정은 F2.I 표에 주어진다.

B1 : 계절성이 존재하지 않는다.
D8 : 식별 가능한 계절성이 존재하지 않는다.
D11: 잔차 계절성이 존재하지 않는다.

계절·불규칙요인 SI_{ij}(i=1,2,⋯,N년, j=1,2,⋯,12(4))에서 j월의 연평균은 $\overline{SI}_{.j} = \frac{1}{N}\sum_{i=1}^{N} SI_{ij}$이며, i년의 월평균은 $\overline{SI}_{i.} = \frac{1}{12}\sum_{j=1}^{12} SI_{ij}$이다. 또한 전체 평균 $\overline{SI} = \frac{1}{n}\sum_{i=1}^{N}\sum_{j=1}^{12} SI_{ij}$, n=12N이다. 이때 전체 평방합(SST) $\sum_{i=1}^{N}\sum_{j=1}^{12}(SI_{ij} - \overline{SI})^2$을 일원분류 분산분석의 경우, 다음과 같이 월간 안정계절성을 나타내는 월(분기)간 평방합(SSS)과 불규칙변동에 의한 오차 평방합(SSR)으로 분리를 할 수 있다.

$$\sum_{i=1}^{N}\sum_{j=1}^{12}(S_{ij}-\overline{SI})^2=\sum_{i=1}^{N}\sum_{j=1}^{12}(SI_{ij}-\overline{SI}_{.\,j})^2+N\sum_{j=1}^{12}(\overline{SI}_{.\,j}-\overline{SI})^2$$

전체 평방합 = 오차 평방합 + 월간 평방합
(SST) (SSR) (SSS)

이때 안정계절성 존재에 대한 F-검정은 오차분산과 월간분산의 비이다. 이에 대한 분산분석표를 만들면 다음과 같다.

<일원분류 분산분석표>

변동요인	평 방 합	자유도	평 균 평 방 합	F - 값
월 간	SSS	11	MSS = SSM/11	F_S= MSS/MSR
오 차	SSR	n-12	MSR= SSR/(N-12)	
전 체	SST	n-1	MST = SST/(n-1)	

여기서 MST = σ_T^2, MSS = σ_S^2, MSR = σ_R^2이다.

F-값인 F_S와 F-분포의 F(11,n-12,α)값과 비교하여 F_S > F(11,n-12,α)이면 “안정계절성이 없다”라는 귀무가설을 기각하게 된다. X-12-ARIMA에서는 귀무가설의 기각여부에 대해서 인쇄를 해주고 있다.

이원분류 분산분석의 경우, 전체 평방합(SST) $\sum_{i=1}^{N}\sum_{j=1}^{12}(SI_{ij}-\overline{SI})^2$은 월간 안정계절성을 나타내는 월(분기)간 평방합(SSS)과 연간 계절변동의 이동을 나타내는 연간 평방합(SSM), 불규칙변동에 의한 오차 평방합(SSR)으로 분리를 할 수 있다.

$$\sum_{i=1}^{N}\sum_{j=1}^{12}(SI_{ij}-\overline{SI})^2=12\sum_{i=1}^{N}(\overline{SI}_{i\,.}-\overline{SI})^2+N\sum_{j=1}^{12}(\overline{SI}_{.\,j}-\overline{SI})^2$$
$$+\sum_{i=1}^{n}\sum_{j=1}^{12}(SI_{ij}-\overline{SI}_{.\,j}-\overline{SI}_{i\,.}+\overline{SI})^2$$

전체 평방합 = 연간 평방합 + 월간 평방합 + 오차 평방합
(SST) (SSM) (SSS) (SSR)

이때 안정계절성 존재에 대한 F-검정은 오차분산과 월간분산의 비이

며, 이동계절성 존재에 대한 F-검정은 오차분산과 연간분산의 비이다. 이에 대한 분산분석표는 다음과 같다.

<이원분류 분산분석표>

변동요인	평방합	자유도	평 균 평 방 합	F - 값
월 간 연 간 오 차	SSM SSM SSR	11 N-1 11(N-1)	MSS= SSS/11 MSM= SSM/(N-1) MSR= SSR/11(N-1)	F_S= MSS/MSR F_M=MSM/MSR
전 체	SST	n-1		

여기서 MST = σ_T^2, MSM = σ_M^2, MSS = σ_S^2, MSR = σ_R^2이다.

이때, F_S>F(11,11(N-1),α)이면 "안정계절성이 없다"라는 귀무가설을 기각하게 된다. 또한 F_M>F(N-1,11(N-1),α)이면 "이동계절성이 없다"는 귀무가설을 기각하게 된다. X-12-ARIMA에서는 귀무가설의 기각여부에 대해서 인쇄를 해주고 있다.

X-12-ARIMA 및 X-11-ARIMA에서는 일반적인 1% 또는 5% 유의수준의 검정 결과가 아닌, 안정계절성 검정은 0.1%, 비모수적 안정계절성 검정은 1%, 이동계절성 검정은 5% 유의수준에서 검정해 주며, 그 결과를 인쇄해 준다.

4.3.2 Kruskall-Wallis 검정

계절·불규칙요인을 월별로 크기 순서로 작은 것부터 번호를 부여하여 안정계절성을 검정하는 방법이 크루스칼-왈리스(Kruskall-Wallis)의 비모수 검정이다. 12개월의 모집단이 동일하다(즉, 안정계절성이 존재하지 않는다)는 귀무가설을 검정하는 이 비모수적방법은 다음과 같이 계산한다.

먼저 계절·불규칙요인 전체를 월별로 작은 것부터 크기순으로 번호를 부여한다. 계절·불규칙요인 SI_{ij}의 순위를 R_{ij}이라 할 때, SI_{ij} 대신에 R_{ij}를 대체하여 월별 순위합을 구한다.

이때 검정 통계량 $H = \frac{12}{n(n+1)}[\sum_{i=1}^{12}\frac{W_i^2}{n_i}]-3(n+1)$로, 귀무가설 하에

서 근사적으로 자유도 11(=12-1)인 χ^2분포를 따른다. 즉, 검정 통계량 H가 유의수준 α에서 $\chi^2(11)$ 값보다 클 때, 귀무가설은 기각되며 유의수준 α에서 안정계절성이 존재하게 된다.

	월			
	1	2	……	12
	R_{11}	R_{21}	……	$R_{12 \cdot 1}$
	R_{12}	R_{22}	……	$R_{12 \cdot 2}$
	:	:		:
	R_{1N}	R_{2N}	……	$R_{12 \cdot N}$
순위합	W_1	W_2		W_{12}

4.3.3 식별 가능한 계절성 존재에 대한 결합 검정

이 검정은 앞에서 서술한 안정계절성 및 이동계절성 검정을 위한 F-검정 및 Kruskall-Wallis 검정을 동시에 수행하며, 식별 가능한 계절성을 검정해 준다. 만일 어떠한 시계열이 안정계절성은 거의 없고 급속히 변하는 이동계절성만이 존재한다면, 계절성은 X-12-ARIMA에 의해서 정확히 식별할 수 없기 때문에 계절요인이 정확히 추정되지 못하게 된다. 따라서 이때 식별 가능한 계절성 존재에 대한 결합 검정을 해 준다.

식별 가능한 계절성 대한 결합검정은 앞에서 서술된 3개 검정 통계량을 이용하여 다음과 같이 4가지 Case에 대해서 검정을 한다.

Case 1. 유의수준 0.1%에서 안정계절성의 존재에 대한 F_S-검정이 기각되지 않으면, "식별 가능한 계절성이 없다"라는 귀무가설이 채택된다.

Case 2. F_S-검정이 기각되어 안정계절성은 있다고 판단이 되나, 5% 유의수준의 이동계절성 F_M-검정이 기각되지 않는다면, 다음의 T_1과 T_2 통계량을 이용한다. 이때 이들의 평균이 1이상이면, "식별 가능한 계절성이 없다"라는 귀무가설을 채택한다.

$$T_1 = \frac{7}{F_M - F_S} \qquad T_2 = \frac{3F_M}{F_S}$$

Case 3. F_S-검정 및 F_M-검정이 기각되고, 2개의 T 통계량 중 하나가 기각되거나 Kruskall-Wallis 검정이 1% 유의수준에서 기각되지 않는다면, "식별할 수 있는 계절성이 있음(Identifiable Seasonality present)"을 인쇄한다.

Case 4. F_S, F_M, 그리고 Kruskal-Wallis 검정이 기각되면, “식별 가능한 계절성이 없다”라는 귀무가설이 기각되며 "식별할 수 있는 계절성이 있음(Identifiable Seasonality present)"을 인쇄한다.

4.3.4 계절성 검정 예

다음은 산업생산지수(1985.1월 ~ 2004.12월)를 이용한 예이다. 계절조정 시, 명절효과, 요일효과 등의 사전조정은 실시하지 않았다. 계절조정을 하기 전에 사전조정된 원계열에 계절성이 존재하는지(F2.I) 먼저 확인하여야 하나 여기에서는 편의상 최종 계절・불규칙요인인 D8표부터 설명한다.

최종 계절・불규칙요인(D8)의 계절성 검정결과는 X-12-ARIMA의 결과표 D8.A를 보면 다음과 같은 결과표를 인쇄해 준다.

D 8.A F-tests for seasonality

① Test for the presence of seasonality assuming stability.

	Sum of Squares	Dgrs.of Freedom	Mean Square	F-Value
Between months	2935.6685	11	266.87896	41.289**
Residual	1473.7305	228	6.46373	
Total	4409.3991	239		

**Seasonality present at the 0.1 per cent level.

② Nonparametric Test for the Presence of Seasonality Assuming Stability

Kruskal-Wallis Statistic	Degrees of Freedom	Probability Level
152.3201	11	0.000%

Seasonality present at the one percent level.

③ Moving Seasonality Test

	Sum of Squares	Dgrs.of Freedom	Mean Square	F-value
Between Years	105.9492	19	5.576272	1.451
Error	803.3075	209	3.843576	

No evidence of moving seasonality at the five percent level.

④ COMBINED TEST FOR THE PRESENCE OF IDENTIFIABLE SEASONALITY
IDENTIFIABLE SEASONALITY PRESENT

결과 ①의 안정계절성 검정을 위한 F-검정(H_0: “안정계절성이 존재하지 않음”)의 검정 결과는 다음과 같다. F_S-value는 41.289로 0.1% 유의수준에서 안정계절성 존재한다. ②의 안정계절성을 검정하기 위한 Kruskal-Wallis의 비모수적 검정(H_0: “안정계절성이 존재하지 않음”)은 Kruskal-Wallis 통계치는 152.3201로 p-값은 0.000이므로 귀무가설이 기각되며 1% 수준에서 안정계절성이 존재한다고 결론을 내린다. ③의 이동계절성 검정을 위한 F-검정(H_0: “이동계절성이 존재하지 않음”)은 F_M-value는 1.451로 5%에서 이동계절성이 존재한다는 증거는 없다고 결론을 내린다.

④의 식별 가능한 계절성 존재에 대한 결합 검정은 안정계절성 검정통계량 F_S가 유의수준 0.1%에서 귀무가설 기각되고 Kruskal- Wallis 검정결과 1%유의수준에서 기각되므로, 식별할 수 있는 안정계절성이 존재한다고 결론을 내린다. 그러나 이동계절성 검정 통계량 F_M은 유의수준 5%에서 귀무가설이 채택되므로 식별할 수 있는 이동계절성이 존재하지 않는다. 따라서 Case 2에 해당되며, 이때 T_1=-0.17, T_2=0.11이므로, 귀무가설을 기각하므로 “식별 가능한 계절성이 있다”라는 결과를 얻는다.

한편, 계절조정계열(D11)에 대한 잔차계절성 검정 결과는 다음과 같이 D11.A에 인쇄된다.

```
Test for the presence of residual seasonality.
     No evidence of residual seasonality in the entire series at the
         1 per cent level.  F =      1.01
      No evidence of residual seasonality in the last 3 years at the
         1 per cent level.  F =      0.39
      No evidence of residual seasonality in the last 3 years at the
         5 per cent level.
```

위의 최종계절조정계열(D11)의 잔차계절성에 대한 F-검정 결과는 전체계열을 이용하는 경우와 최근 3년 시계열을 이용하는 경우로 나누어 다음과 같이 검정해 준다.

• 전체 계열: F=1.01로 1% 유의수준에서 H_0(“잔차계절성이 없음”) 채택
• 최근 3년: F=0.39로 1% 유의수준에서 H_0(“잔차계절성이 없음”) 채택
5% 유의수준에서 H_0(“잔차계절성이 없음”) 채택

마지막으로 이들 검정통계량을 종합한 결과는 다음과 같이 F2.I에서 인쇄해 준다.

```
 F 2.I:                                                    Statistic  Prob.level
   F-test for stable seasonality from Table B 1.      :  28.245     0.00%
   F-test for stable seasonality from Table D 8.      :  41.289     0.00%
   Kruskal-Wallis Chi Squared test
             for stable seasonality from Table D 8. : 152.320     0.00%
   F-test for moving seasonality from Table D 8.      :   1.451    10.64%
```

이때 사전조정한 원계열에 대한 계절성 존재에 대한 검정은 B.1표의 F-검정에 의해 실시되며, F-값은 28.245(p-값: 0.00)이므로 H_0("계절성이 존재하지 않음")가 기각된다. 따라서 사전조정한 원계열에 계절성이 존재한다고 할 수 있다.

4.4 계절조정결과의 진단

계절조정방법은 SAVL, BV4, X11, X-11-ARIMA, X-12-ARIMA, TRAMO/SEATS, SEASABS, DEMETRA[1] 등이 있다. 그러나 계절성과 계절조정 결과를 평가하기 위한 객관적인 통계의 정의가 불가능하므로 이들 계절조정 방법 중에서 어떠한 방법이 더 우월하다고 확정적으로 이야기할 수는 없다(Findley and Monsell, 1990). X-12-ARIMA에서는 계절조정 결과에 대한 자가 진단방법으로 X-11-ARIMA에서 적용된 M-통계량과 Q-통계량(Lothian and Morry, 1978) 뿐만 아니라 이동기간분석, 수정률, 스펙트럼분석 등을 제시하고 있다.

계절조정의 목적은 계절조정계열을 구하기 위하여 원계열에서 계절요인을 추출하는데 있다. X-11 방법의 반복적인 계산 특성상 최종 계절조정계열(final seasonally adjusted series)을 산출하기 위해서는 계절요인 뿐만 아니라 추세·순환요인과 불규칙요인을 정확히 추정하는 것이 중요하다. 계절조정계열이나 불규칙요인에 잔차 계절성이나 캘런더효과(명절 또는

---

1) DEMETRA는 Eurostat에서 TRAMO/SEATS와 X-12-ARIMA를 결합한 사용자 친화적인 계절조정프로그램이다.

요일효과)등을 찾을 수 있다면, 계절조정이나 사전조정이 만족스럽게 이루어졌다고 할 수 없다. 계절조정결과에서 잔차 효과를 찾을 수 없다고 하더라도, 추가된 시계열에 의하여 다시 계절조정을 하였을 때 월간 변화율 등과 같은 계절조정 결과가 크게 변한다면 계절조정이 적절하다고 보기 어렵다. 이러한 불안정성은 계절조정하고자 하는 시계열에 내재하는 계절성이나 추세요인이 민감하게 변동하거나, 사용하는 소프트웨어의 옵션을 부정확하게 사용하기 때문에 생길 수 있다.

신뢰할 수 있는 계절조정 결과는 잔차계절성(residual seasonality)과 안정성(stability) 등이 확보되어야 한다. 잔차계절성이란 계정조정계열이나 불규칙요인에 계절성이 남아 있지 않아야 한다는 것이다. 안정성이란 계절조정계열에 새로 시계열을 추가 또는 제외하여 계절조정을 하였을 때, 새로 추정된 계절조정계열이나 추세·순환요인의 결과에는 큰 변화가 없어야 한다. Hood와 Findley(2001)는 장기 시계열의 경우, 계절조정계열이나 최종 불규칙요인의 스펙트럼 분석 시 계절 또는 요일주기에서 정점을 보인다면, 잔차 계절성 또는 요일효과가 남아 있음을 보여주고 있다.

Cleveland와 Terpenning(1982)은 계절조정 결과의 안정성 등을 보기 위하여 월별 계절요인 및 계절요인-불규칙요인의 그림과 계절조정계열의 수정률 분석하였다. Findley와 Hood (1999)는 X-12-ARIMA/GRAPH에 Cleveland와 Terpenning의 그래프적 방법을 이용하여 최종 불규칙요인의 등분산성(heteroskedastically)과 추세를 제거한 계열(SI 비율과 SI 비율의 극단값을 대체한 SI 비율)을 분석함으로써 계절조정계열의 안정성을 분석하고자 하였다.

Dosse와 Planas(1996)은 X-12-ARIMA와 TRAMO/SEATS의 계절조정 결과에 대한 안정성을 비교하기 위하여 계절조정계열과 추세요인의 초기 추정치와 최종 추정치간의 수정률(r_k)을 다음과 같이 정의하였다.

$$r_k = A_{t|t+k+1} - A_{t|t+k}$$

여기서, $A_{t_0|t}(t_0 \le t \le T)$는 시계열$(Y_1, Y_2, ..., Y_t)$에 의해 추정된 t_0시점 관측치Y_{t_0}에 대한 계절조정계열(또는 추세요인)이다.

Absolute Revision Variation: $ARV = \sum | r_{k+1} - r_k | \ / \ A_{t|t+T_0}$

Smoothness of Revision: $SMR = \sum (r_{k+1} - r_k)^2 / A_{t,|T}^2$

Sum of Squared Revision: $SQR = \sum r_k^2 / A_{t|T}^2$

Mean Convergence $MC = \dfrac{1}{T_0} \sum_{k=0}^{T_0 - 1} \left(\sum_{i=0}^{k} \dfrac{r_i^2}{SQR} \right)$

Smoothness of Convergence $SC = \sum_{k=0}^{T_0-2} \left(\sum_{i=0}^{k+1} r_i^2 - \sum_{i=0}^{k} r_i^2 \right) / SQR$

X-12-ARIMA에서 계절조정 결과의 적정성은 11개의 M-통계량과 이들을 가중평균한 Q-통계량에 의해서 분석할 수 있다. 또한 잔차 계절성을 분석하기 위하여 F-검정 통계량, 계절조정계열과 최종 불규칙요인의 스펙트럼 등이 주어진다. 시계열 $(Y_1, Y_2, \ldots, Y_t)$에 의한 계절조정 결과에 잔차 계절성이나 요일주기가 나타나지 않더라도, 추가(또는 제외)된 시계열 자료$(Y_1, Y_2, \ldots Y_t, \ldots, Y_{t+n})$에 의해 다시 계절조정을 했을 때, 계절조정계열이나 추세요인의 이전 추정치와 차이인 수정률(Revision)이 높거나, 이동기간분석(Sliding span analysis)시 신뢰할 수 없는 비율이 높다면 계절조정 결과의 안정성은 만족스럽지 않다고 할 수 있다.

4.4.1 M-통계량 및 Q-통계량

X-12-ARIMA 결과의 “F2 표”는 추정된 추세·순환, 계절인자 및 불규칙요인과 관련한 각종 통계량이 주어진다. 이들 통계량은 계절조정의 결과를 평가하기 위한 월 간격길이(different span), 원계열의 변화에 대한 상대적 기여도, 연(run)의 평균 연속기간 등에 관한 정보들을 제공해 준다. 다음은 “F2 표”를 이용한 11개의 검정통계량과 Q통계량을 소개한다. Lothian과 Morry(1978)가 421개의 캐나다 시계열을 분석하여 구한 가중치로 11개 통계량를 조합한 Q통계량을 이용하여 계절조정한 결과에 대한 최종 분석과 평가를 “F3 표”의 마지막에 주어지고 있다.

(1) M_1 : 3개월 간격의 차분에 대한 불규칙요인의 상대적 기여도

계절조정 시 계절요인과 불규칙요인을 정확히 분리하는 것은 매우 중요하다. 계절요인 변동에 비해 불규칙요인 변동이 너무 크다면 두 요인은 성공적으로 분리될 수 없다. 차분(differencing)은 시계열 속에 내재되어 있는 선형적 추세를 제거하여 정상계열로 만드는 장점은 있으나, 차분을 하면 다른 시계열 구성요인의 분산에 영향을 미친다. 따라서 3개월 간격의 차분을 하는 경우, 원계열의 분산에 대한 불규칙요인 분산의 상대적 최대 허용 기여율은 10%이다. 만일 3개월 간격의 차분에 대한 불규칙요인의 기여율인 $R_{I(3)}$가 10%보다 크다면 아래의 M_1검정통계량은 사용할 수 없다.

$$M_1 = \frac{R_{I(3)}}{10} \tag{4.1}$$

여기서, $R_{I(3)} = \bar{I}^2_{(3)} / \bar{Y}^2_{(3)} \times 100\%$ 이며

$$\bar{I}_{(3)} = \{\frac{\sum_{t=4}^{n} | I_t - I_{t-3} |}{(n-3)}\}^2 \tag{4.1.1}$$

$$\bar{Y}^2_{(3)} = \bar{I}^2_{(3)} + \overline{TC}^2_{(3)} + \bar{S}^2_{(3)} \text{ 이다.}$$

여기서 $\overline{TC}^2_{(3)}$과 $\bar{S}^2_{(3)}$은 식(4.1.1)에 따른다. 이때 만일 M_1>1.0 이면, 시계열의 전체분산 중에서 불규칙요인의 분산이 크게 기여하고 있음을 나타낸다. 이는 불규칙요인과 계절요인의 분리가 제대로 되지 않았음을 의미한다.

(2) M_2: 정상 시계열의 분산에 대한 불규칙요인 분산의 상대적 기여도

정상 시계열로 만들기 위해 사용되는 추세 제거법은 차분법이지만, 여기서는 3차 차분 대신 D12의 추세·순환요인에 선형회귀를 적합시켜 추세요인을 추정한다. 이때 시계열이 승법모형이라면, 지수적 증가 형태로 가정하며 구성 요인들은 대수(log)변환을 한다. 정상계열의 원계열 B1′를 만들기 위하여 원계열에서 추세치(D12)를 제거한다. D12′은 D12에서 추세치를 제거한다. F2.F에 나타나는 요인의 상대적 기여도는 다음과 같이 계산된다.

$$\text{I의 기여도} = \frac{\text{불규칙 요인}(D13)\text{의 분산}}{\text{추세가 제거된 원계열}(B1')\text{의분산}}$$

$$\text{C의 기여도} = \frac{\text{추세가 제거된 순환요인}(D12')\text{의 분산}}{\text{추세가 제거된 원계열}(B1')\text{의 분산}}$$

$$\text{S의 기여도} = \frac{\text{최종계절요인}(D10)\text{의 분산}}{\text{추세가 제거된 원계열}(B1')\text{의 분산}}$$

불규칙요인 I의 기여도가 10%이상이면 다음과 같은 M_2통계량을 사용할 수 없다.

$$M_2 = 100 \times \frac{I\text{의 기여도}}{10}$$

이때 $M_2 > 1.0$이면, 불규칙요인의 분산이 추세가 제거된 정상시계열의 분산에 너무 많이 영향을 미치고 있음을 나타낸다.

(3) M_3 : 불규칙요인에 대한 추세·순환요인의 전월비 크기 비교

불규칙요인의 전월비가 최종 계절조정계열(CI)의 움직임을 주도하고 있다면, 불규칙요인과 추세·순환요인을 분리하는 것은 어려우며 계절조정 결과의 질을 보장할 수 없게 된다. 따라서 불규칙요인과 추세·순환요인의 전월비 관계를 $\bar{I}/\bar{C}$ 비율에 의해서 측정한다. 여기서 $\bar{I}$, $\bar{C}$는 D13과 D12의 절대변화의 평균이다. 만일 $\bar{I}/\bar{C} > 3$이면 불규칙변동의 크기가 너무 크다는 것을 나타낸다.

$$M_3 = \begin{cases} (\bar{I}/\bar{C} - 1) \ / \ 2 \ , & \text{월별자료의 경우} \\ (\bar{I}/\bar{C} - .33) / \ .67 \ , & \text{분기별자료의 경우} \end{cases}$$

만일 $M_3 > 1.0$이면 불규칙변동이 추세·순환변동에 비해 크므로 계절조정이 잘 되었다는 보장을 할 수 없다.

(4) M_4 : 연(run)의 평균 연속기간에 의한 불규칙요인의 자기상관

D13의 최종 불규칙요인(잔차)에 대한 임의성(randomness) 검정은 연의 평균연속기간(Average Duration Run: ADR)을 이용하며, 그 결과는 F2.D에 있다. 전월비의 부호가 바뀌는 횟수와 관련된 이 비모수적방법은 잔차 I_t가 다음과 같은 1차 자기회귀모형에 따른다는 대립가설에 대한 잔차의 임의성 검정이다.

$$I_t = \rho I_{t-1} + \epsilon_t$$

여기서 ρ는 자기회귀 계수이며, ϵ_t는 순수한 임의과정이다.

$$ADR \begin{cases} > 1.75 & \text{양의 자기상관} \\ = 1.50 & \text{잔차는 임의성} \\ < 1.30 & \text{음의 자기상관} \end{cases}$$

을 나타낸다. M_4 통계량은 다음과 같은 정규분포의 점근식을 이용한다.

$$M_4 = \frac{\left| \frac{n-1}{ADR} - \frac{2(n-1)}{3} \right|}{\sqrt{(16n - 29)/90}} \times \frac{1}{2.58}$$

여기서 2.58은 양측검정에서 표준정규분포의 1% 유의수준에 대한 임계치이다. 이때 만일 $M_4 > 1.0$이면, 잔차에는 유의적인 자기상관이 존재하며, 계절조정이 잘 되었다고 할 수 없다.

(5) M_5: 불규칙요인 변화율보다 큰 추세·순환요인 변화율의 월 간격

M_5 통계량은 계절조정계열(CI)에서 분리되어 나온 불규칙요인과 추세·순환요인에 대한 변동의 상대적 크기를 측정한다. 두 요인의 변동비율을 계산할 때, 월 간격을 길게 할수록 순환성은 커지나 불규칙요인은 제거되지 않게 된다. 따라서, 월 간격이 길어질수록 순환성의 크기가 불규칙성의 크기를 능가하게 된다. M_5는 추세·순환변동이 불규칙변동보다 크게 만드는 이동평균 기간의 길이를 검정해 준다. k=1,2,⋯,12(분기의 경우 1,2,⋯,4)에 대한 $\bar{I}_{(k)}/\bar{C}_{(k)}$ 비율에 의하여 MCD(Months for Cyclical

Dominance)는 다음과 같이 계산한다. MCD 계산 결과는 "F2.E 표“에서 주어진다.

$$MCD = k, \qquad \overline{I_{(k)}}/\overline{C_{(k)}} \leq 1 \text{이고} \quad \overline{I_{(k-1)}}/\overline{C_{(k-1)}} > 1 \text{ 일때}$$

MCD 통계량은 정수형이다. 그러나 $\bar{I}_{(k)}/\overline{C}_{(k)}$비율에 의해서 이동평균기간인 MCD를 결정할 때, 월 간격 개월 수(k)가 정수 사이에 있는 경우는 구별할 수 없으므로 이를 보정해주기 위해 다음과 같은 보간법을 이용하여 MCD′를 구한다.

$$MCD' = K + \frac{R_{(K)} - 1.0}{R_{(K)} - R_{(K+1)}}$$

여기서 $K = MCD - 1$ 이며 $R_{(K)} = \bar{I}_{(K)}/\overline{C}_{(K)}$이다. 이때,

$$M_5 = \begin{cases} \dfrac{MCD - 0.5}{5.0} & \text{월별자료} \\[2ex] \dfrac{QCD - 0.17}{1.67} & \text{분기자료} \end{cases}$$

이다. $M_5 > 1.0$이면 MCD에서 제공되는 이동평균기간은 기각되며, 또한 MCD 통계량이 6이상이면 MCD를 이용한 방법은 옳지 않다.

(6) M_6 : 계절요인과 불규칙요인의 연간 변화율의 비교

X-11 방법에 의한 계절조정 시, 계절요인에서 불규칙요인을 분리하기 위하여 SI 비율에 따라 가중 7개항 이동평균(3 ×5)을 사용한다. 이때 불규칙요인의 연간변동이 작다면, 즉 $\bar{I}/\overline{S}$의 비율이 낮다면, (3 ×5) 이동평균은 계절변동에 신축적이지 못하다. 반면, $\bar{I}/\overline{S}$의 비율이 너무 높다면 (3 ×5) 이동평균은 과도하게 신축적인 결과를 보인다. 이는 계절인자의 결과 속에 불규칙 변동의 일부가 포함되기 때문이다.

캐나다의 경우 1.5≤ $\bar{I}/\bar{S}$ ≤6.5이면 (3×5) 이동평균을 사용한다. 그러나 $\bar{I}/\bar{S}$<1.5이면 보다 짧은 이동평균 기간, $\bar{I}/\bar{S}$>6.5이면 보다 긴 이동평균 기간이 필요하다(4.2.4절 참조). 이동평균 기간을 검정하기 위하여 다음의 M_6 통계량을 사용한다.

$$M_6 = \left| \frac{\bar{I}/\bar{S} - 4.0}{2.5} \right|$$

이때 $M_6 > 1.0$이면, 적절한 이동평균 기간이 사용되지 않은 것이다. 따라서 $\bar{I}/\bar{S} < 2.5$ 이면 (3 ×3) 이동평균을, $\bar{I}/\bar{S}$>6.5이면 보다 긴 이동평균 기간인 (3×9) 이동평균으로 계절조정을 한다.

(7) M_7 : 안정계절성에 대한 이동계절성의 상대적 크기

최종 계절인자 추정치에 대한 절대오차가 크지 않다면, 시계열에 내재된 계절성을 식별할 수 있다. 이러한 오차는 안정계절성 검정 통계량인 F_S와 이동계절성 검정통계량인 F_M의 값과 관계가 있다. 낮은 F_S은 큰 오차를 나타내는 반면, 높은 F_M은 연간변동으로 인해 생기는 큰 오차를 나타낸다. 시계열에 내재된 계절성을 X11방법에 의해서 인식할 수 있는지는 위의 2개의 F-통계량에 의해서 구해지는 M_7에 의해서 검정할 수 있다.

$$M_7 = \sqrt{\frac{1}{2}\left(\frac{7}{F_S} + \frac{3F_M}{F_S}\right)}$$

$M_7 > 1.0$이면 식별할 수 있는 계절성이 없다.

다음 M_8~M_{11}은 연간 변동과 관련된 내용이다. 변동은 다음과 같이 2가지 형태로 분류할 수 있다;

- 유사임의변동(quasi random fluctuation)
- 전계열이 같은 방향으로 변하는 변동(선형변동)

유사임의변동은 계절인자의 연간 변화를 절대평균으로 측정할 수 있으며,

전계열이 같은 방향으로 변하는 변동은 변화에 대한 단순 산술평균이 체계적인(또는 선형) 변동의 크기로 나타난다. 임의변동은 M_8과 M_{10}으로 측정된다. M_9와 M_{11}은 선형변동의 크기를 측정하며, M_8과 M_9는 계절요인(D10)으로부터 계산된다. 대부분의 사용자가 관심있는 최근의 계절조정결과에 대한 평가는 M_{10}과 M_{11}에 의해 알 수 있다. 최근의 계절요인이 유의적인 선형변동을 보인다면 계절조정 시, 사용된 최근의 가중치에 의해서 추정된 계절요인이 영향을 받았기 때문이다.

$M_8 \sim M_{11}$을 계산하기 위한 표준화 계절요인 S_t'를 다음같이 정의하자.

$$S_t' = \frac{S_t - \overline{S}}{S_t\text{의 표준편차}}$$

여기서 $\overline{S}$는 평균 계절인자이다.

(8) M_8 : 전 계열에 걸친 계절요인의 변화 크기

최종 계절요인(D10)의 전계열에 대한 전년동기에 대한 연간 평균절대변화율은 표준화 계절요인을 사용하여 다음과 같이 계산한다.

$$| \Delta\overline{S} | = \frac{1}{J(N-1)} \sum_{j=1}^{J} \sum_{i=2}^{N} | S_{ij}' - S_{(i-1)j}' |$$

여기서 N은 년수, J는 계절주기로 월별자료이면 12, 분기별 자료이면 4이다. 이때 채택 가능한 최대변화는 10%이며, M_8 통계량은

$$M_8 = 100 \times | \Delta\overline{S} | \times \frac{1}{10}$$

이다. 이때, $M_8 > 1.0$이면, 전 계열에 걸쳐 계절요인의 변화가 크므로 계절요인 추정에 오류가 있다는 것을 의미한다.

(9) M_9 : 전 계열에 대한 계절요인의 평균 선형변동의 크기

각 월에 대한 연간변동의 평균은 절대변화의 평균으로 구한다. 이때 연간변동의 평균이 0이라면, 변동은 연간 임의변동에 가깝다. 대부분의 변화가 각 월에 대해서 같은 방향이라면, 절대변화의 평균은 다음의 연간평균변화 $\Delta\overline{S'}$와 비슷하다.

$$\Delta\overline{S'} = \frac{1}{J(N-1)}\sum_{j=1}^{J}\sum_{i=2}^{N}(S_{ij}' - S_{(i-1)j}')$$

이때 채택 가능한 최대변화는 10%이며, M_9 통계량은

$$M_9 = 100 \times \frac{\sum_{j=1}^{J}|S'_{J(N-1)+j} - S'_j|}{J(N-1)} \times \frac{1}{10}$$

이다. 이때 $M_9 > 1.0$이면, 전 계열에 걸쳐 같은 방향으로 변동이 발생하는 체계적인 변화가 있다는 것을 의미한다.

(10) M_{10} : 최근 3년 계절요인의 변화 크기

이 통계량은 최근 3년(N-2, N-3, N-4)의 표준화 계절요인을 사용하며, M_8 통계량과 같은 방법을 사용한다. 연간 절대변화율은

$$|\Delta\overline{S'}|_R = \frac{1}{3J}\sum_{j=1}^{J}\sum_{i=N-4}^{N-2}|S_{ij}' - S_{(i-1)j}'|$$

로 계산된다. 이때 채택 가능한 최대변화는 10%이며, M_{10} 통계량은

$$M_{10} = 100 \times |\Delta\overline{S'}|_R \times \frac{1}{10}$$

이다. 이때, $M_{10} > 1.0$이면, 최근 계열에 있어서 계절요인의 변화가 크므로 계절요인 추정에 오류가 있다는 것을 의미한다.

(11) M_{11} : 최근 3년 계절요인의 평균선형변동의 크기

최근 3년(N-2, N-3, N-4)간의 표준화 계절요인을 사용하여 M_9 통계량과 같은 방법을 사용한다. 연간 변화율은 다음과 같이 계산한다.

$$\Delta \overline{S}'_R = \frac{1}{3J}\sum_{j=1}^{J}\sum_{i=N-4}^{N-2}(S_{ij}' - S_{(i-1)j}')$$

이때 채택 가능한 최대변화는 10%이며, M_{11} 통계량은

$$M_{11} = 100 \times \Delta \overline{S}'_R \times \frac{1}{10}$$

이다. 이때 $M_{11} > 1.0$일 때, 체계적인(선형) 변동에 대한 가중치의 효과가 고르게 분포되어 최근의 계절요인이 심하게 왜곡되어 있음을 나타낸다.

11개의 통계량은 계절조정의 각기 다른 요인들에 대해서 검정을 해주나, 계절조정에 대한 전반적인 검정을 해주는 통계량은 아니다. 각 통계량은 일반적인 시계열에 대해서 적용되나, 일반적이지 않은 시계열에서는 잘못된 결론을 유도할 수 있으므로 이용 시 유의하여야 한다.

M_1과 M_2는 계절변동에 대한 불규칙변동의 비율을 나타낸다. 대부분의 불규칙요인은 계절조정계열의 분산에 대해서 약 5～10% 기여하는 순환주기를 가지므로 한계수준을 10%로 한다. 순환요인이 없는 계열이라면, 불규칙변동이 전체분산에 대해 13～14%로 M_1과 M_2가 1보다 크지만 계절조정 결과는 채택 가능하다고 할 수 있다. $M_1 > 1$이면 불규칙변동이 시계열의 총분산에 크게 작용하고 있음을 나타내며, $M_2 > 1$이면 불규칙요인의 분산이 추세가 제거된 정상계열의 총분산에 너무 많이 영향을 미치고 있음을 나타낸다.

만일 시계열이 평평한(거의 상수)순환주기를 갖는다면, $\overline{I}/\overline{C} > 3$으로 M_3는 기각되나 계절조정이 잘못되었다고는 할 수 없다. 실제로 X-11 프로그램은 추세·순환요인을 추정하기 위하여 23개월 헨더슨 이동평균을 사용함으로써 추세요인 추정의 약점을 보완하고 있다. 그러나 사용자의

목적이 경기순환분석에 있다면, 높은 M_3는 심각한 문제를 야기시킬 수 있다. 이것은 최종 계절조정계열 속에 추세·순환요인의 분해를 방해하는 불규칙변동이 존재함을 의미하기 때문이다.

M_4가 1보다 크다는 것은 최종 불규칙변동 속에 유의적인 자기상관이 존재함을 의미한다. 이것은 표본조사에 의해 발생된 것일 수도 있으므로, 이 통계량은 계절조정의 질을 판단하는 것과는 큰 관련이 없다.

M_5에서 불규칙변동이 너무 크면, 정확한 계절요인을 분리시키는데 장애가 되는 거의 일정한 추세·순환요인을 포함하고 있음을 나타낸다.

M_6은 잘못된 계절조정방법을 고칠 수 있는 유일한 통계량이다. M_6는 계절조정의 정도를 높이기 위하여 계절·불규칙요인에 대한 적절한 이동평균법을 적용하도록 해 준다.

M_7은 계절성에 대한 검정을 해주며, 계절조정에서 가장 중요한 통계량이라고 할 수 있다. 만일 M_7의 결과, 계절성을 인식할 수 없다면 계절조정을 하지 않는 것이 좋다. 급격한 추세성장과 요인들이 승법적으로 구성되어 있으나, 계절조정 시 가법옵션을 사용하여 F_M값이 매우 큰 경우에는 계절조정을 할 수 있다.

M_8과 M_{10}이 기각되더라도 M_9와 M_{11}이 1보다 작다면 심각한 문제가 아닐 수 있으며, 사용자는 계절요인의 편위(bias)만 조심하면 된다. M_9와 M_{11}이 1보다 크지만 사용자가 역사적 계절인자(historical seasonal factor)에만 관심이 있다면, 중심화 계절이동평균의 가중치가 어떠한 선형 변동을 따르기 때문에 이들 요인에 대한 추정치는 정확할 수 있다. 그러나 현재의 계절요인에 관심이 있다면, 높은 M_9와 M_{11}은 추정에 유의적인 오차가 존재함을 의미한다.

앞에서 서술한 11개의 통계량이 모두 기각되면 계절조정계열은 채택될 수 없다. 그러나 일부만 채택되고 나머지는 기각되면, 11개의 통계량에 가중치를 부여한 Q-통계량을 이용하여 계절조정 결과를 평가할 수 있다. 이때, 가중치는 캐나다 시계열을 분석하여 어느 한 통계량이 계절조정계열을 기각하는 원인이 될 수 없도록 계절조정의 전체적인 평가에 대한 상대적 중요성에 따라 부여한 것이다. 11개 통계량에 대한 가중치는 <표 4.1>

과 같다. Q-통계량은 다음과 같다.

$$Q = \frac{\sum_{i=1}^{11} w_i \, M_i}{\sum_{i=1}^{11} w_i}, \qquad 0 < Q < 3$$

만일 계절요인을 추정하기 위해 (3×5) 이외의 이동평균을 선택한다면 M_6 통계량은 필요가 없으며 따라서 $w_6 = 0$이다. 분석하는 시계열이 6년 이하의 시계열이나 불안정계열이라면 M_8, M_9, M_{10}, M_{11}은 계산되지 않으며 가중치 $w_8 = w_9 = w_{10} = w_{11} = 0$이다.

<표 4.1> M-통계량 가중치

통 계 량(M_i)		w_i
M_1	3개월 간격의 차분 에 대한 불규칙요인의 상대적 기여도	0.13
M_2	정상시계열 분산에 대한 불규칙요인 분산의 상대적 기여도	0.13
M_3	불규칙요인에 대한 추세·순환요인의 전월비 크기 비교	0.10
M_4	연(run)의 평균 연속기간에 의한 불규칙요인의 자기상관	0.05
M_5	불규칙요인 변화율보다 큰 추세·순환요인 변화율의 월간격	0.11
M_6	계절성요인과 불규칙요인의 연간 변화율의 비교	0.10
M_7	안정계절성에 대한 이동계절성의 상대적 크기	0.16
M_8	전 계열에 걸친 계절요인의 변화 크기	0.07
M_9	전 계열에 대한 계절요인의 평균 선형변동의 크기	0.07
M_{10}	최근 3년 계절요인의 변화 크기	0.04
M_{11}	최근 3년 계절요인의 평균선형변동의 크기	0.04

이때 $Q > 1.0$ 이면, 계절조정 결과는 좋다고 할 수 없다. 또한 식별할 수 있는 계절성에 대한 검정이 기각된다면, 계절조정요인은 기각된다. Q-통계량이 계절조정요인에 대해서 일반적인 평가를 해주지만 작은 변화에도 유의적인 결과를 줄 수 있으므로 주의해야 한다. 특히, 미국의 Census Bureau에서는 좋은 계절조정 결과의 조건으로 작은 수정률(revision)과 함께 안정 계절성에 대한 이동 계절성의 상대적 크기인 M_7과 M-통계량의 가중 합인 Q-통계량이 1보다 작을 것을 강조하고 있다.

4.4.2 스펙트럼분석

스펙트럼 분석은 시계열의 기간이 충분히 긴 시계열에서 주기성이 있는 계절요인이나 요일요인이 시계열에 존재하는지 찾기 위한 그래프적 방법이다. 계절요인은 매년 동일한 월에 반복적으로 나타나는 요인으로 일반적으로 월자료는 12개월, 분기자료인 경우 4분기의 주기를 갖는다. 요일주기는 매월 동일한 요일에 나타나는 요인으로 7일마다 반복해서 나타나게 된다. 일반적으로 산업생산 등은 토・일요일보다는 주중에 생산활동이 활발하게 이루어지므로 토・일요일이 적은 월에는 많은 월보다 생산이 많이 된다. 이처럼 주기성을 갖는 요인들은 다음의 Fourier transformation을 기초로 하는 스펙트럼 분석에 의해 탐색할 수 있다. t=1,2,⋯,T에 대해서 원계열 y_t의 Fourier 함수는 다음과 같이 표현할 수 있다.

$$\begin{aligned} y_t &= \frac{a_0}{2} + \sum_{i=1}^{m}[a_i\cos(2\pi\frac{i}{T}t) + b_i\sin(2\pi\frac{i}{T}t)] \\ &= \frac{a_0}{2} + \sum_{i=1}^{m} r_i\cos(2\pi\frac{i}{T}t) \end{aligned}$$

여기서 m은 T가 짝수이면 T/2, 홀수이면 (T-1)/2이다. 한편, i/T는 Fourier 빈도(frequency) 또는 주기를 나타내며, 월간자료의 경우, 1/12이다. Periodogram 또는 진폭(amplitude) r_i는 다음과 같이 추정할 수 있다 (Box and Jenkins, 1976).

$$r_i = \frac{T}{2}(a_i^2 + b_i^2), \quad i = 1,2,...,m$$

X-12-ARIMA에서는 계절성과 요일효과를 탐색하기 위하여 다음과 같은 자기회귀 스펙트럼 추정치(autoregressive spectral estimator; $\hat{s}(\lambda)$)를 이용하고 있다(U.S. Census Bureau, 2004).

$$\hat{s}(\lambda) = 10 \bullet \log_{10}\left\{\frac{\widehat{\sigma_m^2}}{2\pi\left|1 - \sum_{j=1}^{m}\widehat{\psi}_j e^{i2\pi j\lambda}\right|^2}\right\}, \quad 0 \le \lambda \le 0.5$$

여기서 $\hat{\psi}_j$는 $y_{t-j}-\bar{y}$ $(1 \le j \le m)$에 대한 $y_t-\bar{y}$의 선형회귀계수이며, $\bar{y}=\sum_{t=1}^{T} y_t/T$, $\hat{\sigma}_m^2$은 회귀잔차의 표본분산이다.

X-12-ARIMA는 m=30에서 스펙트럼 분석을 위한 $\hat{s}(\lambda)$ 값을 구하며, X-12-ARIMA에서 제공되는 스펙트럼 그림은 다음 2가지 경우를 제외하고 $\lambda_k=k/120,\ 0 \le k \le 60$에 대하여 61번째 빈도의 $\hat{s}(\lambda)$ 값을 나타낸다. k/120가 요일주기(월자료인 경우 0.348과 0.432)에 근사한 경우, $\hat{s}(\lambda)$ 값의 열에 "T"를 인쇄하고, 계절주기에서는 $\hat{s}(\lambda)$ 값의 열에 "S"를 인쇄한다. 그리고 다른 주기에서의 값에는 "*"를 인쇄한다.

X-12-ARIMA는 Cleveland와 Devlin(1980)가 적용하였던 계절주기와 요일주기를 이용하고 있다. 즉, 계절성과 요일효과를 포함하고 있는 월자료의 시계열은 계절주기 k/12, k=1,2,⋯,6(S_1, S_{12} ,⋯, S_6)와 요일주기 0.348(T1), 0.432(T2)에서 정점을 보이며, 이를 계절정점(seasonal peak)과 요일정점(trading peak)이라 한다.

X-12-ARIMA 스펙트럼은 경험적 결과에 의하여 $\max_k \hat{s}(\lambda_k)-\min_k \hat{s}(\lambda_k)$를 계산하며, 계절주기나 요일주기 주변의 "*"보다 6"*"가 높은 경우 계절정점이나 요일정점이 있다고 한다. 이때 "시각적으로 유의함(visual significance)" 이라는 경고문을 준다(Soukup과 Findley, 1999).

X-12-ARIMA에서는 RegARIMA 잔차, 원계열, 계절조정계열과 불규칙요인에 대한 8년 간의 자료에 의해 스펙트럼이 default로 추정된다. 추정된 스펙트럼에서 계절주기나 요일주기에서 정점이 나타난다면, 시계열에 계절성이나 요일효과가 존재한다고 할 수 있다. 특히, 계절조정계열 또는 불규칙요인에 대한 스펙트럼이 계절주기에서 정점을 보인다면, 계절 filter의 부적절성으로 잔차 계절성이 존재함을 나타내므로 다른 계절 filter나 분석기간을 짧게 하는 것을 고려해 볼 수 있다.

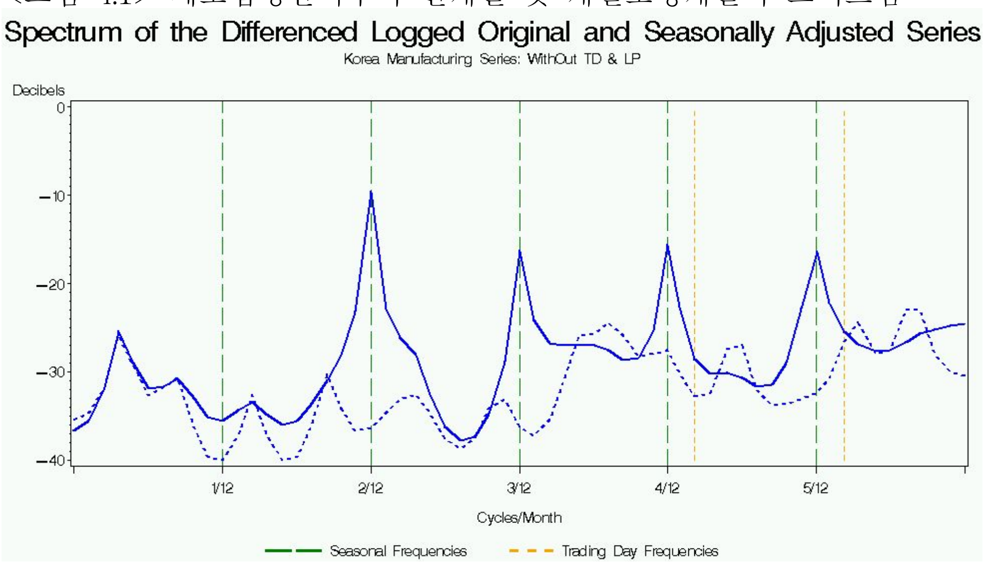
한편, 요일주기에서 정점을 보인다면, 원계열에 요일효과가 존재한다는 것을 의미하므로 요일효과변수를 RegARIMA 모형에 포함하여 요일효과를 조정할 필요가 있다.

원계열을 이용하여 스펙트럼 분석을 하는 경우, 계절성이 요일효과보다 크게 작용하므로 요일정점이 나타나지 않을 수 있다. 따라서 요일효과를 검증하기 위한 스펙트럼 분석은 계절조정계열이나 RegARIMA 모형의 잔차를 이용하여야 한다.

(예 4.1) 다음은 1985년 1월부터 2003년 12월까지의 제조업생산지수를 이용하여 계절성과 요일효과를 찾기 위한 스펙트럼 분석결과이다. <그림 4.1>부터 <그림 4.3>은 SAS를 이용한 X-12-ARIMA/GRAPH(사용법은 <참고 4.3> X-12-ARIMA/GRAPH 참조)에 의해 제공된 결과이며 <그림 4.4>와 <그림 4.5>는 X-12-ARIMA에서 제공된 결과이다.

<그림 4.1>은 요일효과 및 윤년효과 변수를 RegARIMA 모형에 반영하지 않은 경우의 제조업생산지수 원계열 및 계절조정계열의 스펙트럼 그래프이다. 원계열 스펙트럼(실선)의 경우, 계절정점이 2/12(S2)와 3/12(S3), 4/12(S4), 5/12(S5) 등에서 나타나고 있어 계절성이 있는 것으로 나타났다. 그러나 요일효과는 요일주기(T1, T2)에서 정점이 나타나고 있지 않아 요일효과는 없는 것처럼 보이나, 이는 원계열을 이용 시 계절성 때문에 요일효과가 제대로 나타나지 않기 때문이다. <그림 4.2>의 RegARIMA 잔차 스펙트럼은 요일주기 T1에서 정점을 보이고 있어 요일효과가 있음을 나타내고 있다. 또한 <그림 4.5>의 계절조정계열의 스펙트럼에서도 요일주기 T1에서 정점을 보이고 있다.

<그림 4.1> 제조업생산지수의 원계열 및 계절조정계열의 스펙트럼



X-12-ARIMA/GRAPH의 metafile인 x12g.gls의 스펙트럼 옵션은 다음과 같다.

spectrum: spcosa spcirr spcrsd

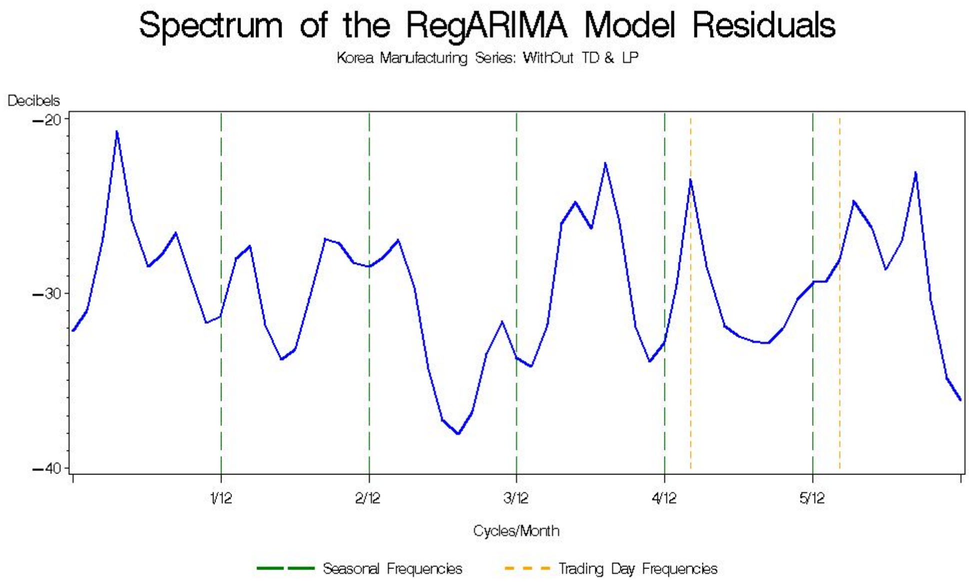
X-12-ARIMA/GRAPH를 이용하는 경우는 다음 경로를 이용한다.

Spectrum Graphs/Original and Seasonally Adjusted Series

<그림 4.2>는 요일효과 및 윤년효과를 반영하지 않은 RegARIMA 모형에 대한 제조업생산지수의 RegARIMA 잔차 스펙트럼이다. 그림을 보면 계절주기에서 정점을 보이고 있지 않으므로 ARIMA (311)(011) 모형은 적절한 것으로 보인다. 그러나 요일주기 T1에서 정점을 보이고 있어 요일효과가 있는 것으로 보여진다.

X-12-ARIMA/GRAPH => RegARIMA Model Residuals

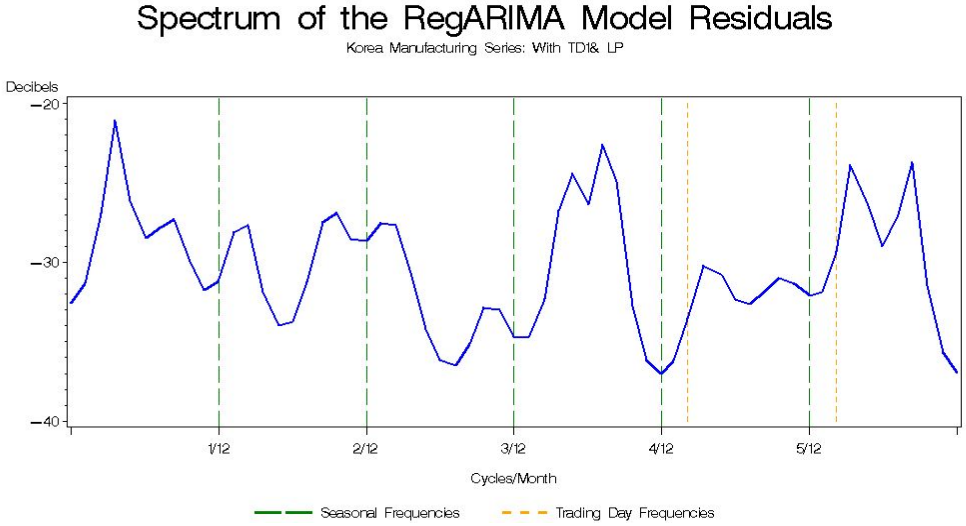
<그림 4.2> 제조업생산지수의 RegARIMA 잔차 스펙트럼



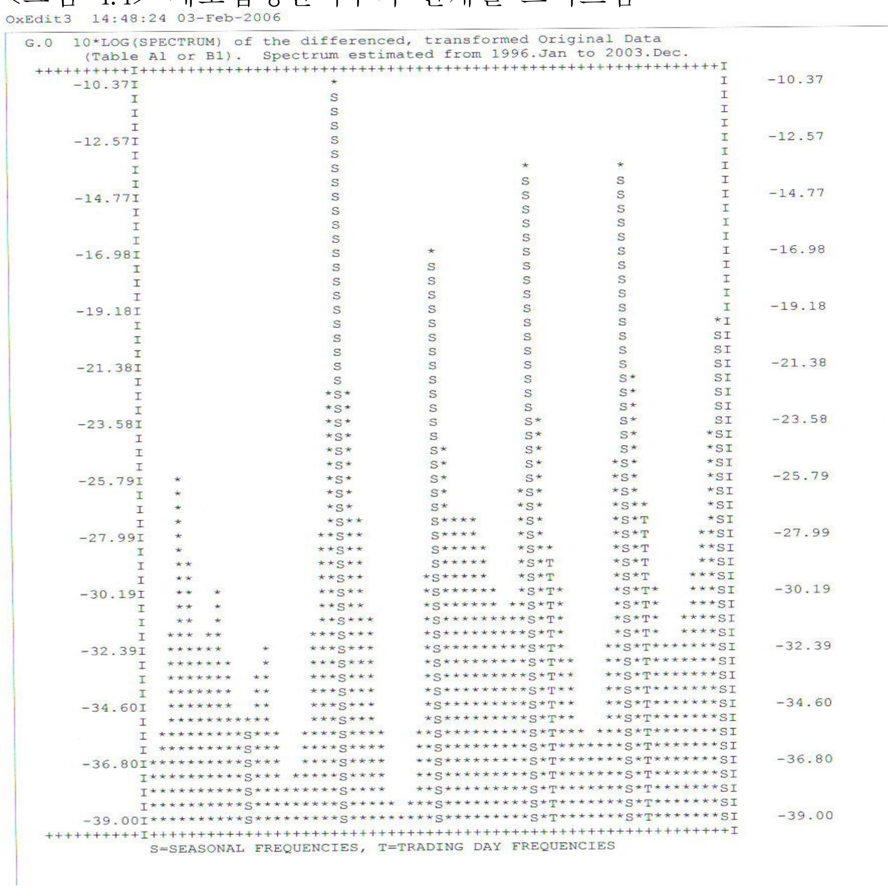
<그림 4.3>은 1개 요일효과와 윤년효과 변수를 반영한 RegARIMA 모형의 잔차 스펙트럼이다. 그림을 보면 계절주기와 요일주기에서 정점이 보이지 않고 있어 요일효과와 계절성은 제거되었음을 보여주고 있다.

윤년효과변수를 반영한 RegARIMA 모형에 대한 모형 추정 결과를 보면, 1개 요일효과 변수의 회귀계수는 0.0022(t값: 4.13), 윤년 회귀계수는 0.0041(t값: 0.28), 이들 두 변수에 대한 모형의 적합성 검정 통계량인 X^2 값은 17.12(p값: 0.00), 특이치는 AO87.AUG(t값: -4.87)로 추정되었다. 이때 ARIMA 모형은 (311)(011)으로 AICC-통계량은 952.6이다.

<그림 4.3> 제조업생산지수의 RegARIMA 잔차 스펙트럼

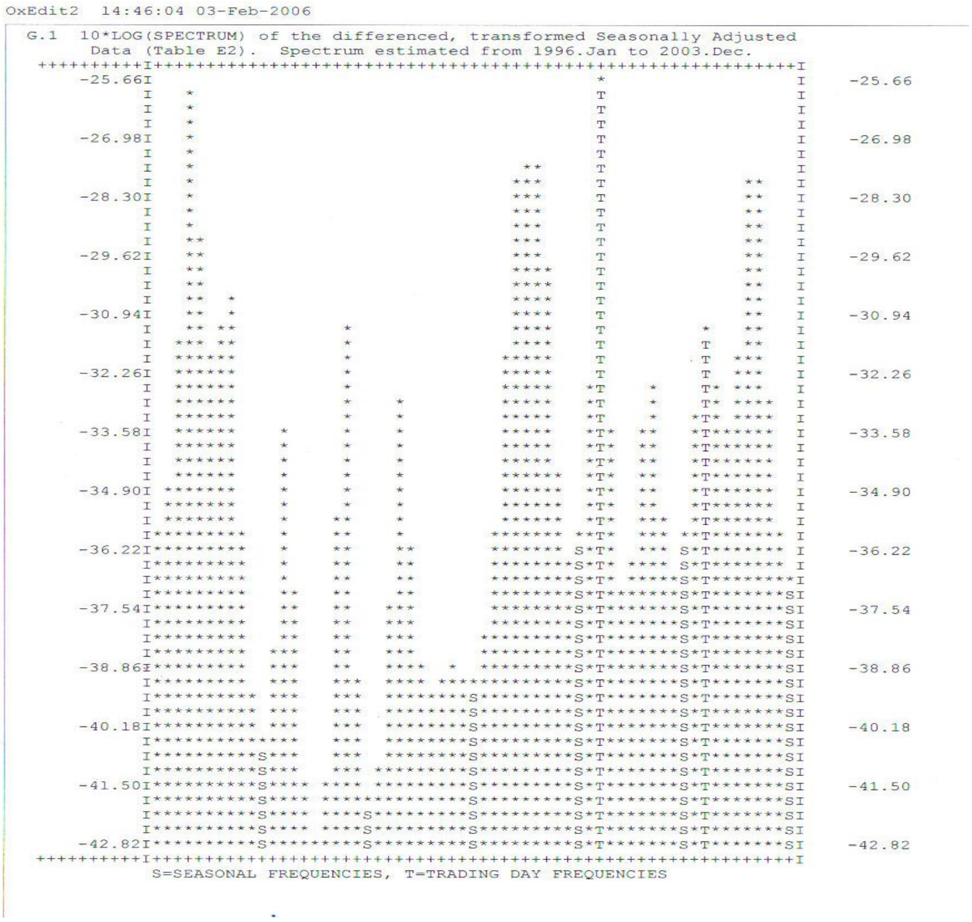


<그림 4.4> 제조업생산지수의 원계열 스펙트럼



<그림 4.4>는 X-12-ARIMA 프로그램에서 제공되는 차분 및 log 변환된 원계열의 스펙트럼이다. 이때 스펙트럼 분석을 위한 기간은 최근 8년(1996년 1월 부터 2003년 12월)으로 Default 기간이다. 그림을 보면 계절주기(S)의 경우, 2번째 주기와 3번째 주기 등에서 주변의 다른 '6*'보다 높게 나타나고 있어 계절성이 있음을 나타내고 있다. 한편, 요일주기(T)에서는 주변의 '*'보다 높지 않아 요일효과는 없는 것처럼 보여진다. 그러나 계절조정계열의 스펙트럼을 보면(<그림 4.5>), T1에서 요일주기가 있음을 알 수 있으며 원계열과 계절조정계열 스펙트럼의 경우 강한 계절성이 요일효과를 잠식한 것으로 보여진다.

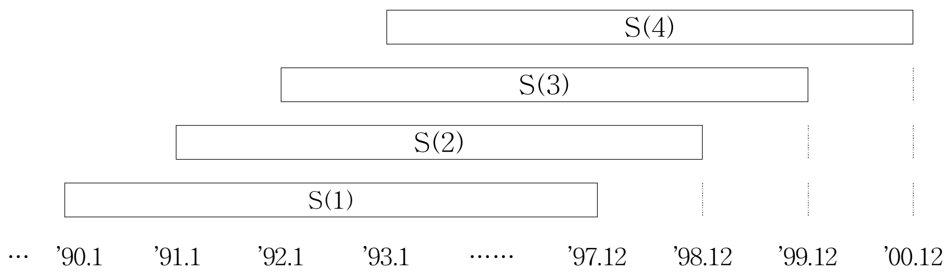
<그림 4.5> 제조업생산지수의 계절조정계열 스펙트럼



주) S: 계절주기, T: 요일주기

4.4.3 이동기간분석(Sliding Span Analysis)

이동기간분석은 계절조정을 하기 위한 시계열의 기간을 중첩이 되도록 4개의 부분 시계열로 구성한 후(부분시계열을 중첩기간(overlapping span) 시계열이라고도 한다.), 각 부분 시계열에 대하여 계절조정을 한다. 이때 4개 부분 시계열에 대한 계절조정 결과의 최대치와 최소치를 비교하여 결과가 크게 차이가 난다면 계절조정계열은 신뢰할 수 없음을 나타내며 안정성이 결여되었다고 할 수 있다. 즉, S(1)은 [1990.1월~1997.12], S(2)는 [1991.1월~1998.12월]이며, 이때 [1992. 1월~1997.12월]은 부분 시계열인 S(1)과 S(2)에서 공유하고 있으므로, 계절조정 결과가 안정적이라면 S(1)과 S(2)의 계절조정 결과는 큰 차이가 없어야 한다.



$S_t(k)$는 k번째 부분 시계열에서 추정된 t월의 계절요인이며, $A_t(k)$는 k번째 부분 시계열에서 추정된 t월의 계절조정계열이라 하자. 이때 k번째 부분 시계열에서 추정된 t월의 계절조정계열 전월비 $MM_t(k)$ 및 계절조정계열 전년비 $YY_t(k)$는 다음과 같다.

$$MM_t(k) = [A_t(k) - A_{t-1}(k)]/A_{t-1}(k)$$
$$YY_t(k) = [A_t(k) - A_{t-12}(k)]/A_{t-12}(k)$$

X-12-ARIMA에서는 계절조정 결과의 안정성을 평가하기 위한 계절요인과 계절조정계열의 전월비 및 전년비의 최대 퍼센트편차(MPD: Maximum Percentage Difference)를 다음과 같이 정의하며, 임계치 0.03 (3%)을 넘는 경우 t월의 계절조정 결과는 신뢰할 수 없다고 가정한다.

$$계절요인\ MPD: S_t^{\max} = \frac{\max_{k \in N_t} S_t(k)\ -\ \min_{k \in N_t} S_t(k)}{\min_{k \in N_t} S_t(k)} > 0.03$$

$$전월비\ MPD: MM_t^{\max} = \max_{k \in N1_t} MM_t(k) - \min_{k \in N1_t} MM_t(k) > 0.03$$

$$전년비\ MPD:\ YY_t^{\max} = \max_{k \in N12_t} YY_t(k) - \min_{k \in N12_t} YY_t(k) > 0.03$$

여기서, N_t={k: k번째 부분 시계열에 포함된 t월}, $N1_t$={k: k번째 부분 시계열에 포함된 t월과 t-1월}, $N12_t$={k: k번째 부분 시계열에 포함된 t월과 t-12월}이다.

이때, S(%)를 신뢰할 수 없는 계절요인($S_t^{\max} > 0.03$인경우)의 백분율, MM(%)는 신뢰할 수 없는 계절조정계열 전월비($MM_t^{\max} > 0.03$인경우)의 백분율이며 YY(%)는 신뢰할 수 없는 계절조정계열 전년비의 백분율이다. Findley와 Monsell 등(1990)은 신뢰할 수 없는 계절요인 및 계절조정계열 전월비 백분율을 이용하여 계절조정결과를 다음과 같이 해석하고 있다.

S(%)	MM(%)	계절조정결과
S(%) ≤ 15.0	MM(%) ≤ 40.0	만족
15.0 < S(%) ≤ 25.0	MM(%) ≤ 40.0	약간 만족
S(%) > 25.0	MM(%) > 40.0	불만족

그러나 S(%)와 MM(%)의 백분율 결과와 계절조정계열의 안정성을 보기 위한 Q-통계량과 일치하지 않은 결과를 얻을 수 있다. 즉, Q-통계량은 만족스러운 결과를 보이나 이동기간분석에서는 신뢰할 수 없는 결과를 보일 수 있다. 이와 반대로 Q-통계량은 계절조정 결과에 대해서 신뢰할 수 없으나 이동기간분석 결과는 신뢰할 수 있는 결과를 보이기도 한다. 이때는 최근 시계열의 결과 또는 수정률 history의 결과에 의해서 판단한다.

X-12-ARIMA는 계절조정모형이 가법모형인 경우나 계절조정 결과가 너무 작은 값이나 음수의 값이 존재하는 경우 이동기간분석의 불안정성 때문에 분석결과가 제공되지 않는다. 또한 계절요인의 범위가 너무 작은 경우에도 분석의 신뢰도가 낮아 이동기간분석 결과가 제공되지 않는다.

(예 4.1) 다음은 1985년 1월부터 2003년 12월까지의 건축허가면적이다. 계절조정을 실시하기 전에, ARIMA(0 1 1)(0 1 1)모형을 이용하여 다음과 같이 사전조정 변수는 최소 AICC 값을 갖도록 선정하여 사전조정을 실시하였다.

사전조정 내용: 1개 요일효과변수, AO1992.Jan, TC2003.Jun, 추석[4]

이때, Q-통계량은 0.57이며 1보다 큰 M값은 M_8로 1.081이다. default로 사용된 계절 filter 대신 3x9, 3x3을 사용하였을 때의 결과는 다음과 같다.
건축허가면적의 경우, 원계열의 변동이 매우 큼으로 3x5 계절 filter 뿐만 아니라 다른 계절 filter를 사용하더라도 안정성은 높지 않음을 알 수 있다. 계절이동평균 기간을 길게 하면, 신뢰할 수 없는 계절조정계열 전월비의 백분율 MM(%)이 낮아짐을 알 수 있다. 그러나 이러한 현상은 항상 이동평균을 길게하면 백분율을 낮게 해주는 것은 아니다(Findley and Monsell, 1990). 한편 Q-통계량은 높게 나타나 계절조정의 질을 하락시킬 수 있다.

	Q-통계량	MM(%)	SS(%)
3x3	0.63	69.8	74.7
3x5(default)	0.57	66.7	83.2
3x9	0.67	59.7	80.4

X-12-ARIMA에서는 “D 9.A Moving seasonality ratio”의 월별 평균값인 GMSR(4.1 X-11의 계절조정 절차 참조)에 의해서 계절 filter를 결정한다. 이때 X-12-ARIMA에서 결산된 D 9.A MSR은 다음과 같다.

	Jan Jul	Feb Aug	Mar Sep	Apr Oct	May Nov	Jun Dec
MSR	1.888	6.187	3.470	3.644	7.816	8.634
	3.828	2.764	5.684	5.971	3.493	1.881

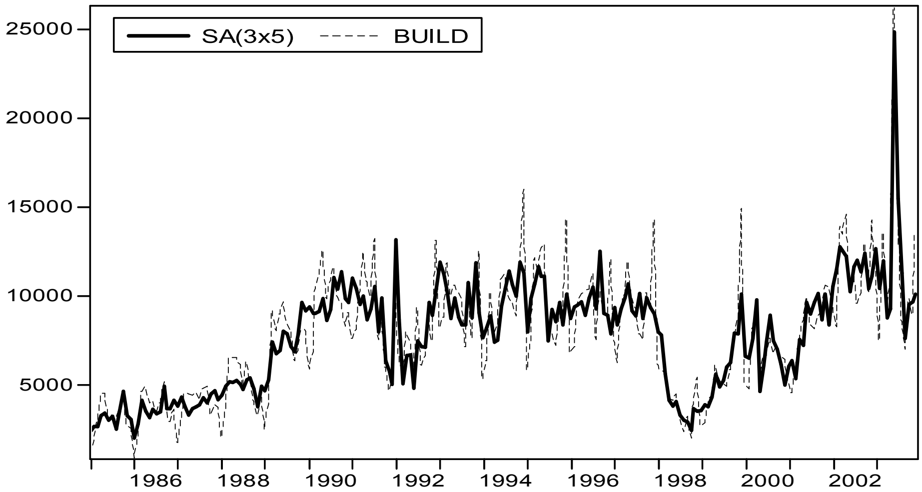
따라서 전체 GMSR은 4.605이므로 3x5 이동평균법(계절 filter)을 적용한다. 위의 MSR 값에 따라 다음과 같이 계절이동평균을 달리 적용할 수 있으며, 결과는 다음과 같다.

x11{seasonalma=(s3x3 s3x5 s3x5 s3x5 s3x9 s3x9
s3x5 s3x5 s3x5 s3x9 s3x5 s3x3) }

	Q-통계량	MM(%)	SS(%)
3x5(월별 달리 적용)	0.59	52.1	62.9

다음 <그림 4.6>은 건축허가면적 원계열과 3x5 계절 filter를 사용한 계절조정계열 그림이다.

<그림 4.6> 건축허가면적 원계열과 계절조정계열(3x5)



4.4.4 수정률(Revision History)

수정률은 시계열 자료를 추가 또는 제외하였을 때, 재추정된 계절조정계열이나 추세요인의 변화율 또는 수정률이 작다면 계절조정의 결과는 안정성이 있다고 할 수 있다. X-12-ARIMA에서 제공되고 있는 수정률은 계절조정계열이나 추세요인의 초기 추정치와 최종 추정치간의 변화율이며, 계절조정계열의 퍼센트 수정률(R1), 계절조정계열의 전월비 수정률(R2), 추세요인의 퍼센트 수정률(R4), 추세요인의 전월비 수정률(R5), 월별 및 연간 수정률, 절대 평균값이 주어진다.

$A_{t_0|t}(t_0 \le t \le T)$는 시계열$(Y_1, Y_2, ..., Y_t)$에 의해 추정된 t_0시점의 관측치 Y_{t_0}에 대한 계절조정계열(또는 추세요인)이라 하자. 그리고 $A_{t|t}$를 시계열 $(Y_1, Y_2, \cdots, Y_t)$을 이용하여 추정된 Y_t의 계절조정계열로 동시 계절조정계열(concurrent) 또는 초기 계절조정계열 이라 하고, $A_{t|T}$를 최종 계절조정계열 또는 가장 최근 자료인 $(Y_1, Y_2, \cdots, Y_T)$를 이용하여 추정한 계절조정계열이라고 하자. 이때 승법 및 가법 계절조정모형의 경우, 계절조정계열의 퍼센트 수정률(R_t)은 다음과 같다.

승법 계절조정모형의 경우: $R_t = \dfrac{A_{t|T} - A_{t|t}}{A_{t|t}} \times 100$

가법 계절조정모형의 경우: $R_t = (A_{t|T} - A_{t|t}) \times 100$

t_0시점의 계절조정계열 Y_{t_0}의 전월비인 $C_{t_0|t}$는 시계열$(Y_1, Y_2, \cdots, Y_t)$으로부터 다음과 같이 정의된다.

$$C_{t_0|t} = \frac{A_{t_0|t} - A_{t_0-1|t}}{A_{t_0-1|t}}$$

이때 계절조정계열의 전월비 수정률(R_t)은 다음과 같다.

$$R_t = C_{t|T} - C_{t|t}$$

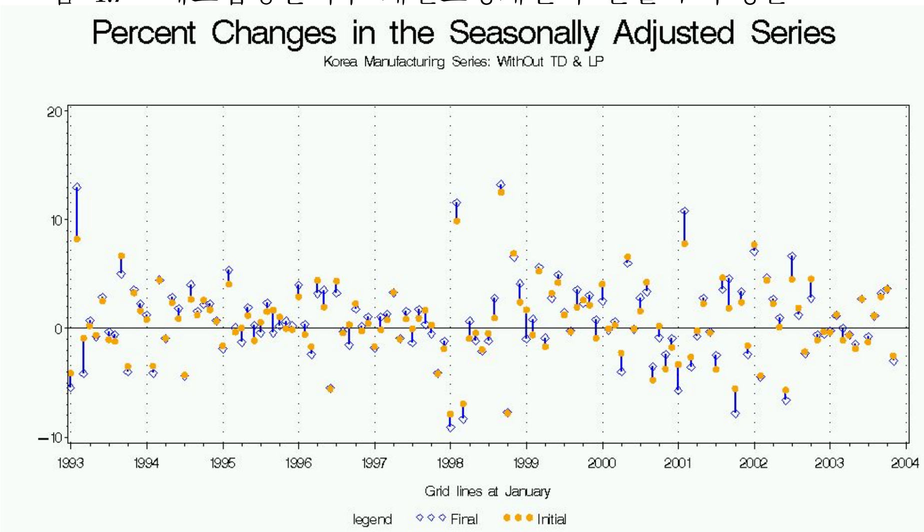
X-12-ARIMA/GRAPH는 계절조정결과의 안정성 검증을 위한 수정률과 모형의 비교를 위한 수정률을 제공한다. 이때 사용되는 계열은 계절조정계열과 추세요인이다.

<그림 4.7>은 제조업생산지수 계절조정 결과의 안정성 검정을 위한 계절조정계열의 초기 추정치와 최종 추정치의 전월비 수정률이다. 그림에서 어떠한 월의 수정률이 다른 월에 비해 큰 것으로 보여지면, 이때 C17와 D9.A를 검토한 후 σ-관리한계와 계절 filter의 길이를 조정함으로써 차이를 축소할 수 있다.

X-12-ARIMA/GRAPH =>

History Graphs/Percent Change in the Seasonal Adjustment

<그림 4.7> 제조업생산지수 계절조정계열의 전월비 수정률

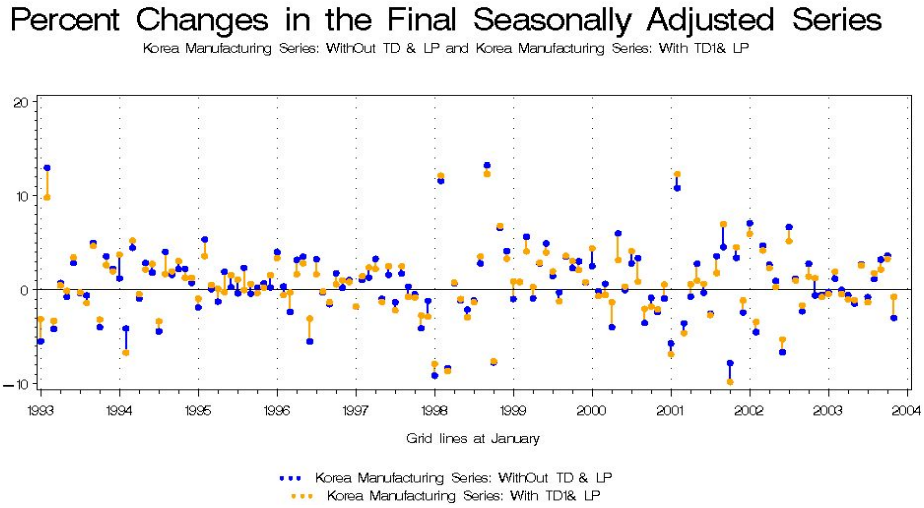


(예 4.2) 다음은 산업생산지수에 요일효과와 윤년효과를 반영하지 않는 RegARIMA 모형(모형 1)과 요일효과와 윤년효과를 반영한 모형(모형 2)의 계절조정 결과이다. 다음 표는 (모형1)과 (모형2)에 대한 계절조정 결과 대한 수정률을 정리한 표이다. 표에 의하면 모형1과 모형2의 계절조정 결과는 안정성이 있는 것으로 보인다. 이때 (모형1)의 안정성을 보기 위한 Q-통계량은 0.46이며 (모형2)는 0.39로 (모형1)보다 낮다. <그림 4.8>은 (모형1)과 (모형2)에 의한 최종 계절조정계열의 전월비 수정률이며, <그림 4.9>은 추세요인의 전월비 수정률이다.

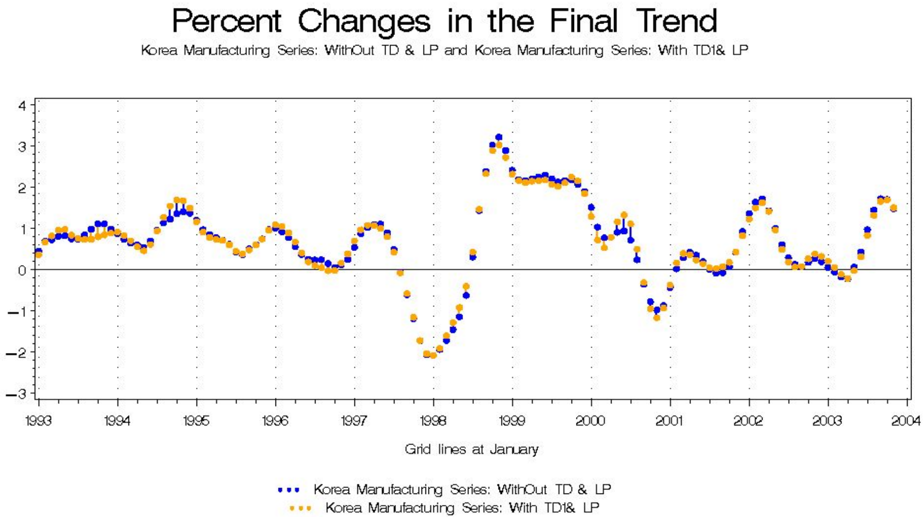
X-12-ARIMA/GRAPH => Comparison Graphs for Two Adjustments /History Graphs/Percent Change in the Seasonal Adjustment

	윤년 및 요일변수 없음(모형1)				윤년 및 요일변수 모형(모형2)			
	R1	R2	R4	R5	R1	R2	R4	R5
MIN	0.00	0.00	0.02	0.00	0.01	0.01	0.02	0.02
Q25	0.30	0.22	0.39	0.18	0.35	0.32	0.38	0.18
MED	0.56	0.62	0.90	0.36	0.66	0.71	0.96	0.41
Q3	1.05	1.19	1.66	0.71	0.98	1.13	1.59	0.73
MAX	2.62	4.73	4.69	2.05	2.68	4.63	4.69	2.03
평균	0.70	0.83	1.10	0.50	0.72	0.85	1.10	0.52

<그림 4.8> 제조업생산지수 최종 계절조정계열의 수정률: 모형1, 모형2



<그림 4.9> 제조업생산지수 최종 추세요인의 수정률: 모형1, 모형2



4.4.5 RegARIMA 모형 비교를 위한 History 분석

요일효과, 명절효과 등의 캘런더 효과와 특이치 등의 사전조정 변수와 ARIMA 모형에 의해 설정되는 RegARIMA 모형의 적합성은 RegARIMA 모형의 우도통계량(Likelihood statistics)인 AIC-통계량과 RegARIMA 모형으로부터 산출되는 out-of-sample의 예측오차 자승합(SSFE: Sum of

Squared Forecast Error)에 의해서 할 수 있다. 시계열의 길이(또는 표본 크기) T가 충분히 긴 시계열에 의해서 구한 로그 우도함수를 L_T, 추정된 모수의 수를 n_p라고 하면, AIC 통계량은 다음과 같이 구할 수 있다.

$$AIC_T = -2L_T + 2n_p$$

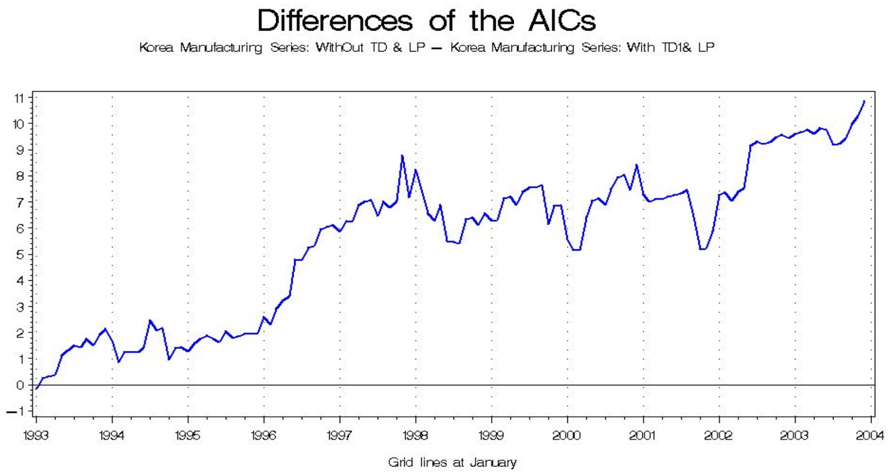
그러나 시계열 길이 T가 충분히 길지 않은 경우에는 수정된 AIC 통계량인 AICC 통계량을 사용한다. 이때 RegARIMA 모형의 선정기준은 AICC (또는 AIC) 통계량이 가장 작은 모형을 선정한다. X-12-ARIMA의 검증통계량은 AIC-통계량보다 AICC-통계량을 선호하고 있으며, TRAMO/SEATS에서는 BIC-통계량을 사용하고 있다.

$$AICC_T = -2L_T + 2n_p\left(\frac{T}{T-n_p-1}\right)$$

RegARIMA 모형 선정을 위한 AIC 차이 history는 $t_0 \le n \le T$에 대해서, 시계열 $(Y_1, Y_2, \ldots, Y_n)$을 이용하여 (모형1)과 (모형2)의 AIC 통계량을 구한 후, 이들 모형의 AIC 차이를 구한다.

$$AIC_n^{1,2} = AIC_n^{\text{모형}1} - AIC_n^{\text{모형}2}$$

<그림 4.10> 제조업생산지수의 AIC 차이 history: (모형1, 모형2)



X-12-ARIMA/GRAPH =>
Comparison Graphs for Two Adjustments/History Graphs/AIC

<그림 4.10>은 (모형1)과 (모형2)에 대한 제조업 생산지수의 AIC 차이 history 그림이다. 그림은 양의 AIC 통계량 차이값 ($AIC_n^{모형1} - AIC_n^{모형2}$)을 갖고 있어, (모형2)의 AIC 통계량이 (모형1)의 AIC 통계량보다 작다. 따라서 (모형2)의 RegARIMA 모형이 선호된다.

한편, Out-Of-Sample의 예측오차 자승합 history는 AIC 통계량 값들이 비슷할 때 또는 차분이나 특이치가 서로 다른 RegARIMA 모형을 비교할 때 유용하게 이용할 수 있다(Hood, 2000).

$t_0 \le t \le T-h$에 대해서, $\hat{Y}_{t+h|t}$를 시계열 자료 $(Y_1, Y_2, ..., Y_t)$를 이용하여 추정된 (t+h)시점의 관측치인 Y_{t+h}의 예측치라고 하자. 이때 Out-Of-Sample의 예측오차 자승합 $SS_{h,M}$은 다음과 같다.

$$SS_{h,M} = \sum_{t=t_0}^{M} (Y_{t+h} - \hat{Y}_{t+h|t})^2, \quad M = t_0, ..., T-h$$

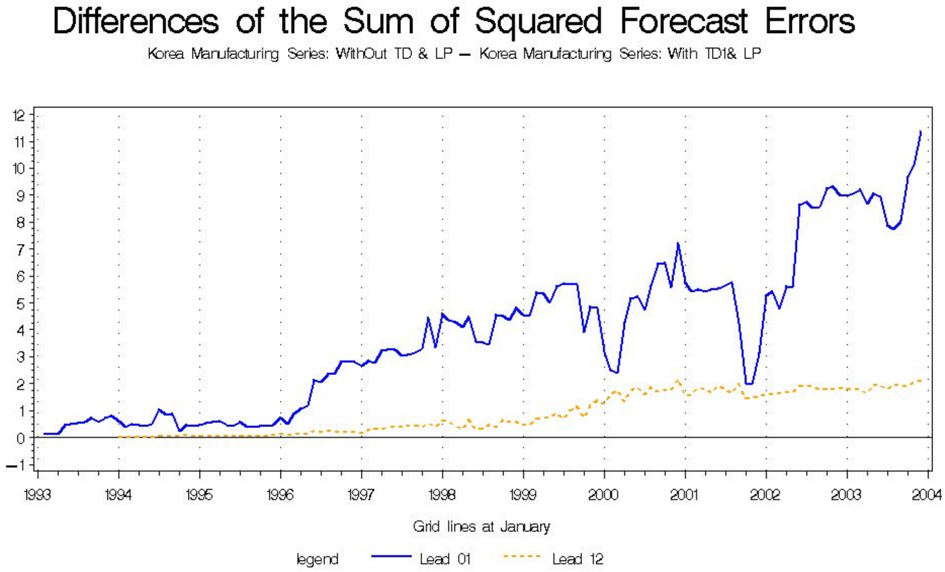
(모형1)과 (모형2)를 비교한다고 가정하면, Out-Of-Sample의 예측오차 자승합을 각각 $SS_{h,M}^{(1)}$과 $SS_{h,M}^{(2)}$으로 표현할 수 있다. X-12-ARIMA/GRAPHICS는 (모형1)과 (모형2)의 예측오차 자승합 차이를 다음과 같이 표준화하여 제공하고 있으며, 선행시차 h는 1시점과 12시점이다.

$$SS_{h,M}^{1,2} = \frac{SS_{h,M}^{(1)} - SS_{h,M}^{(2)}}{SS_{h,T-h}^{(2)} / (T-h-t_0)}$$

<그림 4.11>은 (모형1)과 (모형2)에 대한 제조업 생산지수의 RegARIMA 모형의 1시차 및 12시차 예측오차 자승합 그림이다. 그림을 보면 양수를 나타내고 있어 (모형2)의 예측오차 자승합이 (모형1)보다 작으므로 (모형2)의 RegARIMA 모형이 선호됨을 알 수 있다.

X-12-ARIMA/GRAPH => Comparison Graphs for Two Adjustments
/History Graphs/Sum of Squared Forecast Error

<그림 4.11> 제조업생산지수의 예측오차 자승합 history: (모형1, 모형2)



<참고 4.1> 극단값(Extreme Value) 조정절차

X-12-ARIMA에서는 주변의 값들보다 큰 극단값을 처리하기 위한 방법으로 두가지 방법을 이용된다. 원계열(사전조정 전의 시계열)에서 대부분의 관측치들과는 달리 뚜렷한 차이를 보이는 관측치들을 특이치라고 하며 RegARIMA 모형에 의해서 사전조정을 한다. 또 다른 방법은 X-11 모듈내에서 임의의 SI 요인 값들이 다른 값보다 뚜렷하게 차이가 난다면 극단값으로써 조절하는 방법이다. 즉, 특이치는 RegARIMA 모형에서 다루어지며, 극단값은 X-11 알고리즘에서 다루어진다.

기본적으로 극단값을 조정하기 위한 알고리즘은 다음과 같다.

1. 불규칙요인에서 표준오차를 추정한다.
2. 불규칙요인이 큰 경우, 표준편차와 비교하여 극단값을 조정한다.
   - · 표준편차×1.5배 보다 작으면 대체하지 않음
   - · 표준편차×2.5배 보다 크면 대체
   - · 극단값이 표준편차×1.5배에서 2.5배 사이에 있으면 부분적으로 대체

극단값의 조정은 표 B4부터 시작하며, 표 B9, 표 B17, 표 C17가 계산된다.

표 B4:

Step 1.3의 극단값 대체는 다음과 같이 한다.

ⓐ 계절요인 추정

ⓑ SI로부터 계절요인을 제거함으로써 불규칙요인 추정

ⓒ 이동 표준편차 추정

ⓓ 표준편차에 대한 함수를 기초로 가중치 부여

ⓔ t월과 인근의 4개년도 t월 값을 기초로 극단값 대체

Step 1.3a

Step 1.2 SI에서 3×3 이동평균과 중심화 12-개항 이동평균법을 이용하여 계절요인(S_t)을 계산한다.

Step 1.3b

Step 1.2의 SI 비율을 Step 1.3a에서 계산된 계절요인으로 나눈다.

$$I_t = \frac{SI_t^{(1)}}{S_t}$$

Step 1.3c

1≤t≤12N에 대해서, I_t의 로버스트(robust) 표준오차 σ_t를 구한다. 여기서 N은 연수이다.

Step 1.3d

I_t와 σ_t를 이용하여 가중치 $W_t(0 \leq W_t \leq 1)$를 다음과 같이 구한다.

$$W_t = \begin{cases} 1, & |I_t - 1| \leq 1.5\sigma_t \text{인 경우} \\ \dfrac{2.5\sigma_t - |I_t - 1|}{\sigma_t}, & 1.5\sigma_t < |I_t - 1| < 2.5\sigma_t \text{인 경우} \\ 0, & 2.5\sigma_t \leq |I_t - 1| \text{인 경우} \end{cases}$$

Step 1.3e SI의 극단값 대체(표 B4)

SI의 극단값을 대체하기 위한 대체값을 계산한다. $W_t = 1$인 경우, 대체하지 않는다. $W_t < 1$이고 $W_{t\pm 12} = W_{t\pm 24} = 1$인 경우, SI_t는 다음과 같이 대체된다.

$$SI_t^{(rep)} = \frac{1}{4 + W_t}[SI_{t-24} + SI_{t-12} + W_t SI_t + SI_{t+12} + SI_{t+24}]$$

표 B9:

Step 2.3a에서 계절요인을 추정하기 위하여 3×5 이동평균법을 사용하는 것 이외는 표 B4 절차와 비슷하다.

Step 2.3a

Step 2.2의 계절・불규칙요인(SI 비율)에 3×5 이동평균법과 중심화 12-개항 이동평균법을 적용하여 계절요인을 계산한다.

Step 2.3b

Step 2.2의 SI 비율을 Step 2.3a의 계절요인으로 나눈다.

$$I_t = \frac{SI_t^{(2)}}{S_t}$$

Step 2.3c
$1 \le t \le 12N$에 대해서, Step 2.3b에서 구한 I_t에 대한 로버스트 표준오차 σ_t를 구한다.

Step 2.3d
I_t와 σ_t를 이용하여, Step 1.3d와 같이 가중치 $W_t(0 \le W_t \le 1)$를 구한다.

Step 2.3e SI_t 대체값(표 B4)
Step 1.3e에서와 같이 가중평균에 의해 SI_t의 극단값을 대체한다.

표 B17과 C17:
Step 2.7a
Step 1.3c, 2.3c와 같이 $1 \le t \le 12N$에 대해서, I_t의 로버스트 표준오차 σ_t를 구한다.

Step 2.7b 불규칙요인에 대한 가중치(표 B17, C17)
Step 1.3d, 2.3d와 같이 I_t와 σ_t를 이용하여 가중치 $W_t(0 \le W_t \le 1)$를 구한다.

<참고 4.2> 전체 이동계절성 비율(GMSR) 추정 방법

GMSR은 다음과 같이 계산한다.

① 계절 및 불규칙요인은 가장 최근 12월 자료를 입력한 후 구한다. 표 D9의 SI 비율을 7-개항 이동평균하여 계절요인(S)을 추정한다. 이때, 불규칙요인은 SI/S로 추정한다.

② 월간 MSR(표 D9A)의 계산

i=1,2,⋯,12(4) i월(분기)에 대해서, 계절요인과 불규칙요인의 평균 변화율을 다음과 같이 계산한다.

$$\overline{S_i} = \frac{1}{N_i - 1}\sum_{t=2}^{N_i} | S_{i,t} / S_{i,t-1} - 1|, \qquad \overline{I_i} = \frac{1}{N_i - 1}\sum_{t=2}^{N_i} | I_{i,t} / I_{i,t-1} - 1|$$

여기서 N_i는 i월(분기)의 관측치 수이다. 이때 i월(분기)의 MSR은 다음과 같다.

$$MSR_i = \frac{\overline{I_i}}{\overline{S_i}}$$

③ GMSR의 계산

월별 MSR에서 다음과 같이 GMSR를 구한다.

$$GMSR = \frac{\sum_i (N_i \times \overline{I_i})}{\sum_i (N_i \times \overline{S_i})}$$

<참고 4.3> X-12-ARIMA의 산출과정

다음은 시계열의 분해과정을 각 단계별로 승법형을 중심으로 식으로 나타내었다. 원계열(O)은 추세·순환(C)과 계절(S), 요일변동요인(D)와 불규칙요인(I'')으로 구성되어 있다.

$$O = CSI''D \;;$$

$$D = D_p D_r \;;$$

D_p=사전요일조정요인 ;

D_r=잔여요일변동요인 ;

불규칙요인(I'')은 영구사전조정요인(P′),임시사전조정요인(P″), 부활절효과(H), 특이항(E)과 잔여불규칙요인(I)으로 구성된다. 특이항은 2.5σ(표준편차)밖에 떨어지는 큰 값에 대해 불규칙 값으로 정의한다.

$$I'' = P'P''EHI \;;$$

$$I' = P''EHI \;;$$

단, $E = |I' - 100.0| > 2.5\sigma_I'$

여기서 I'는 명절효과가 존재할때 프로그램의 첫 실행에서 $I' = EHI$, 두번째 실행에서 $I' = EI$로 정의한다.

A부분. 사전조정

이 부분은 원계열에 적용할 수 있는 여러 사전조정을 수행하며, 부활절효과와 요일변동요인, 계절요인의 보다 유효한 추정치를 산출하도록 한다. 사전조정이 필요없으면 바로 B부분이 계산된다. 또한 A부분에서는 요일가중치와 최종월별요일변동요인을 다음에 설명한 것처럼 A1에서 A6와 B1에서 C18의 단계의 프로그램의 첫 실행에서 구할 수 있다. 부활절효과는 A1에서 A9와 B1에서 D13의 단계의 프로그램의 첫 실행에서 추정할 수 있다. 그러나 우리나라에서는 뚜렷한 요일변동요인이 나타나고 있지 않아 적용하고 있지 않으며, 명절효과 역시 우리나라는 설날과 추석의 효

과를 추정해야하므로 부활절효과 추정과정이 불필요하다.

A1. 원계열($O = CSI''D$)

A2. 영구사전월요인(P')

계열의 수준을 영구적으로 조정하고 변화시킬 필요가 있을 때 입력되는 계열로, 일단 조정된 새 계열은 영구변형된다.

A3. 영구사전월조정된 계열($O/P' = CSI'D$)

A1을 A2로 나눈다.

A4. 임시사전월요인(P'')

계열의 수준을 임시로 조정하고 화시킬 필요가 있을때 입력되는 계열로, 이것은 최종계절조정계열, 요일변동 등이 제거된 계열에 다시 적용된다.

A5. 사전조정된 계열($O/P'P'' = CSI'D$)

A1 또는 A3(영구사전조정을 한경우)를 A4로 나눈다.

A15. ARIMA 추정 모형 : 예측(forecasting)

계열을 연장시키기 위해 원계열, 또는 사전조정을 했다면 사전조정계열에 순차적으로 내재된 5개의 표준모형 중에서 채택기준에 만족하는 모형을 자동적으로 선정한다. ARIMA모형의 함수값으로 특이항이 자동적으로 대체되도록 할 수 있다. 즉, 특이항의 사전조정은 추정된 ARIMA 모형에서 잔차가 2.5σ안에 존재하는지 검정하여, 2.5σ밖에 있다면 특이항으로 간주하여 ARIMA모형의 함수값으로 대체한다. 자료가 첨가될 때마다 생기는 모형의 변화를 피하고자 모형의 선택은 12월까지의 update된 자료를 가지고 선택한다. 같은 모형이 예측을 위해 최근년도까지 사용되나, 그 모수는 전체계열을 사용하여 재추정된다.

A16. ARIMA 추정모형 : 후방예측(backcasting)

사용자가 제공한 모형옵션으로 계열을 backcast하여 1년을 연장할 수

있다. 단, 후방예측은 5년 이상 6년 이하의 계열에서만 산출할 수 있다.

B부분. 잠정 추세・순환・계절・불규칙요인의 산출

B1. 원계열 또는 사전조정계열($CSI'D_r$)

이 원계열은 앞에서 설명한 사전조정요인에 의해 조정된 계열이다. 계절성의 유무에 관한 F 검정이 B1에 적용되고 유의수준 0.1%에서 계절성의 유무의 검정결과는 F2.I에 출력된다.

B2. 잠정 추세・순환($M_c\,[CSI'D_r] = C_1$)

B1을 중심화 12개월 이동평균하여 추세・순환의 추정치를 산출한다. 경우에 따라 사용자가 중심화 24개월 이동평균을 할 수 있다.

B3. 잠정 계절・불규칙요인 : SI비율($CSI'D_r/C_1 = SI'D_r$)

B1를 B2로 나누어 계절・불규칙 비율을 계산한다.

B4. 특이항이 수정된 계절・불규칙요인

잠정 계절요인을 추정하기 위하여 SI비율에 따라 B3의 계절・불규칙요인을 월별로 가중5개항 이동평균한 후 중심화 12개월 이동평균한다. 계열양단에 생긴 각각 6개의 결항치는 처음과 마지막의 이동평균값을 6번 반복대체하여 연간합이 1200이 되도록 조정하여 계절요인을 산출한다($M_s\,[SI'D_r] = S$).

계절・불규칙요인을 계절요인으로 나누어 불규칙요인의 추정치를 산출한다($SI'D_r\,/\,S = I'D_r$). 불규칙요인의 추정치를 5년씩 이동하여 표준편차(σ)를 계산하여 $|I' - 100.0| > 2.5\sigma$인 5년간의 중앙년도의 불규칙요인을 특이항으로 간주하여 제거한 후 다시 이동하여 5년간의 표준편차(σ)를 계산한다. 이렇게 구한 표준편차(σ)로 특이항 조정방법에 따라 수정한다.

계열양단 2년의 불규칙요인은 끝에서 3년째의 표준편차(σ)를 사용하여 가중치를 부여한다. B3를 특이항이 수정된 불규칙요인을 적용하여 특이항이 수정된 계절·불규칙요인을 산출한다.

B5. 계절요인($M_s\,[SI^w D_r] \;=\; S_1$)

특이항이 수정된 계절·불규칙요인(B4)의 각월에 각각 가중5 개항이동평균과 중심화 12개월 이동평균을 한 후, 연간 합이 1200이 되도록 조정하여 잠정계절요인을 산출한다. B2단계의서 추세·순환에 대해 중심화 12개월 이동평균하여 생긴 계열양단의 6개의 결항치를 대체하려고, 해당 월에 가장 가까운 유용한 요인을 반복한다.

B6. 계절조정된 계열($CSI' D_r/S_1 \;=\; CI' D_r$)

원계열(B1)을 계절요인(B5)로 나누어 산출한다.

B7. 추세·순환요인($M_c\,[CI' D_r] \;=\; C_2$)

B6의 $\overline{I}/\overline{C}$비율로 선택한 헨더슨 이동평균을 하여 추세·순환요인을 산출한다. 추세·순환의 특이항의 영향은 특이항 조정에 의해 감소시킬 수 있다.

B8. 잠정 계절·불규칙요인($CSI' D_r/C_2 = SI' D_r$)

B7의 추세·순환으로 B3와 같이 계산한다.

B9. 특이항이 수정된 계절·불규칙요인

B8의 계절·불규칙요인을 사용하고 가중7개항이동평균으로 계절요인을 추정하여 B4와 같은 방법으로 계산한다.

B10.계절요인($M_s[SI^w D_r] = S_2$)

특이항이 수정된 계절·불규칙요인(B9)에 가중7개항이동평균을 한 후, 이를 중심화 12개월 이동평균을 하여 연간 합이 1200이 되도록 조정하여 계절요인을 산출한다.

B11.계절조정계열($CSI'D_r/S_2 = CI'D_r$)

B1을 계절요인(B10)으로 나누어 산출한다.

B13.불규칙요인($CI'D_r/C_2 = I'D_r$)

계절조정계열(B11)을 추세・순환요인(B7)으로 나누어 불규칙요인을 산출한다.

B17.불규칙요인의 특이항 수정 가중치

B13의 불규칙요인에 대해 B4와 마찬가지로 가중치를 부여한다.

B20.특이항

불규칙요인(B13)과 특이항 수정가중치(B17)를 이용하여 특이항 $E_t = I'/(1+W(I'-1))$을 산출한다.

C부분. 잠정 추세・순환・계절・불규칙요인의 산출

C1. 원계열($CSI'[1.0 + w(I' - 1.0)]/I' = CSI^w$)

이 원계열은 원계열(B1)에서 불규칙의 특이항이 수정된 계열로 계절요인(B10)에 계절조정계열(B11)을 곱하고 특이항(B20)을 나누어 산출한다.

C2. 잠정 추세・순환요인($M_c[CSI^w] = C_3$)

C1으로 B2와 같이 계산한다.

C4. 수정된 계절・불규칙요인($CSI^w/C_3 = SI^w$)

C1을 C2로 나누어 산출한다.

C5. 계절요인($M_s[SI^w] = S_3$)

C4를 사용하여 B5처럼 계산한다.

C6. 계절조정계열($CSI^w/S_3 = CI^w$)

C1을 C5로 나누어 잠정계절조정계열을 산출한다.

C7. 추세・순환요인($M_c[CI^w] = C_4$)

C6에서 계산된 $\overline{I}/\overline{C}$비율에 의해 헨더슨 이동평균을 하여 추세・순환요인을 산출한다.

C9. 수정된 계절・불규칙요인($CSI^w/S_3 = CI^w$)

원계열(C1)을 추세・순환요인(C7)으로 나누어 계절・불규칙요인을 산출한다.

C10.계절요인($M_s[SI^w] = S_4$)

C9의 수정된 계절・불규칙요인을 사용하여 B10과 같은 방법으로 계산한다.

C11.계절조정계열($CSI'D_r/S_4 = CI'D_r$)

B1을 C10으로 나누어 산출한다.

C13.불규칙요인($CI'D_r/C_4 = I'D_r$)

계절조정계열(C11)을 추세・순환요인(C7)으로 나누어 산출한다.

C17.불규칙의 특이항수정 최종가중치($[I'] = w$)

C3을 이용하여 B17과 같이 계산한다.

C20.특이항

불규칙요인(C13)과 특이항 수정가중치(C17)로 이용하여 특이항 $E_t = I'/(1 + W(I' - 1))$을 산출한다.

D부분. 최종 추세・순환・계절・불규칙요인의 산출

D1. 원계열($CSI'[1.0 + w(I' - 1.0)]/I' = CSI^w$)

C1과 같은 방법으로 계산한다.

D2. 잠정 추세・순환요인($M_c[CSI^w] = C_5$)

D1을 이용하여 B2와 같은 방법으로 계산한다.

D4. 수정된 계절・불규칙요인($CSI^w / C_5 = SI^w$)

원계열(D1)을 추세・순환요인(D2)로 나누어 산출한다.

D5. 계절요인($M_s[SI^w] = S_5$)

D4를 이용하여 B5와 같은 방법으로 산출한다.

D6. 계절조정계열($CSI^w / S_5 = CI^w$)

원계열(D1)을 계절요인(D5)로 나눈다.

D7. 추세・순환요인($M_C[CI^w] = C_6$)

D6을 사용하여 B7과 같은 방법으로 산출한다.

D8. 최종 수정된 계절・불규칙요인($CSI' / C_6 = SI'$)

원계열(A1 또는 A3)을 추세・순환요인(D7)으로 나누어 산출한다. 안정적인 계절성의 유무에 관한 F검정과 비모수검정이 출력된다.

D9. 최종 수정된 계절・불규칙요인($CSI^w / C_6 = SI^w$)

특이항이 수정된 계절・불규칙요인을 산출하고자 D1을 D7으로 나누어 D8과는 다른 결과치를 산출한다. 각 월에 대해 불규칙($\overline{I}$)과 계절($\overline{S}$)의 비($\overline{I} / \overline{S}$ = MSR : 이동계절성비)의 연간평균변화율을 산출한다. 여기서 S는 D8과 D9의 계절・불규칙요인을 비가중 7개항 이동평균한 값이고, I는 계절・불규칙요인에서 S를 나눈 값이다.

D10.계절요인($M_s[SI^w] = S_6$)

D9에서 선택된 월별가중이동평균항수를 D8에 적용하여 최종계절요인을 산출한다.

D10A.1년을 예측한 계절요인

1년의 계절요인 예측치를 다음과 같이 산출한다.

(1) ARIMA의 추정치

(2) ARIMA가 적용안되었다면, $S_{n+1} = S_n + 1/2[S_n - S_{n-1}]$으로 최종계절요인을 추정한다.

D11.최종계절조정계열($CSI'/S_6 = CI'$)

B1을 D10으로 나누어 최종계절조정계열을 산출한다. 잔차의 계절성에 관한 F검정의 결과가 출력된다.

D11A.연간 합을 조정한 최종계절조정계열

A1또는 B1의 연간 합과 D11의 연간 합이 일치하도록 조정된 계열을 제공한다.

D12.최종 추세・순환요인($M_c[CI^w] = C_7$)

D1을 D10으로 나눈 후, D7에서 계산된 $\overline{I}/\overline{C}$에 의해 헨더슨 이동평균을 하여 최종 추세・순환요인을 산출한다.

D13.최종 불규칙요인($CI'/C_7 = I'$)

최종계절조정계열(D11)을 최종추세・순환요인(D12)으로 나누어 산출하고, 전체계열, 각 년도, 각월에 대한 표준편차를 계산한다.

E부분. 특이항이 수정된 계열

E1. 특이항이 수정된 원계열

특이항의 최종 가중치(C17) 중에서 가중치가 0인 원계열(B1)값을 추세・순환요인(D12),계절요인(D10),사전월조정요인(A2)의 곱으로 수정하여 대체한다. 즉, W=0.0일때, I'은 100.0과 같으므로 $CSI''D = CSPD$ 이다.

E2. 특이항이 수정된 계절조정계열

특이항의 최종 가중치(C17)에서 가중치가 0인 최종계절조정계열(D11) 값을 최종추세·순환요인(D12)의 값으로 대체한다. 즉, W=0.0일 때, I′은 100.0과 같으므로 $CI' = C$ 이다.

E3. 특이항이 수정된 불규칙요인

특이항의 최종 가중치(C17)에서 가중치가 0인 최종불규칙요인(D13)의 값을 100.0으로 대체한다. 즉, W=0.0일 때, I′은 100.0과 같다.

E4. 원계열과 최종계절조정계열의 연간비율

최종계절조정계열(D11)과 원계열(A1)의 비와 수정조정계열(E2)와 수정 원계열(E1)의 연간 비율을 각각 구한다.

E5. 원계열의 전월비

원계열(A1)의 전월비를 계산한다.

E6. 계절조정계열의 전월비

최종계절조정계열(D11)의 전월비를 계산한다.

F부분. MCD이동평균과 특성치 - 통계적 평가과정

F1. MCD 이동평균값($M_{MCD}[CI'] = C_{MCD}$)

MCD(Months for Cyclical Dominance)란, 최종 추세·순환요인(D12)과 최종불규칙요인(D13)의 상대적 크기로 계산된다. 추세·순환변동은 일정기간동안 누적적(변화방향의 평균연속기간이 길다)으로 변화하는 성격을 지니고 있는데 반하여, 불규칙변동은 극히 단기적으로 회복(변화방향의 평균지속기간이 짧다)됨에 따라 두 변동요인의 변화폭을 계산할 때, 월 간격을 크게 할수록 추세·순환요인은 커지나 불규칙요인은 거의 변화가 없게 된다. 즉, 월 간격이 길어질수록 추세·순환변동요인의 크기가 불규칙의 크기를 능가하게 되는데 이렇게 능가하는 최초의 월 간격($\overline{I}/\overline{C} > 1$)이 MCD 이동평균기간이다. 따라서 MCD 이동평균은 경제시계열의 변동에서 단기적인 상하운동을 나타내는 불규칙요인을 제거하고 안정적이고 장기적인 시계열의 변

동방향을 신속히 알고자 하는 최종계절조정계열(D11)의 이동평균기간을 의미하므로 계절조정계열을 MCD 이동평균기간으로 이동평균하면 불규칙요인을 제거한 추세・순환변동을 알 수 있게 된다.

최종계절조정계열(D11)을 MCD 단순이동평균을 한다. 단, MCD가 짝수이면, 중심화 MCD 이동평균을 한다.

F2A.구성요인들의 월 간격 변화율 절대평균

월 간격(s=1,2,⋯,12)에 의한 다음의 시계열의 각 구성요인들의 변화율절대평균을 계산한다.

<u>구성요인</u>

A1($\overline{O_s}$: 원계열)

D11($\overline{CI_s}$: 최종계절조정계열)

D13($\overline{I_s}$: 최종불규칙계열)

D12($\overline{C_s}$: 최종추세・순환)

D10($\overline{S_s}$: 최종계절요인)

A2($\overline{P_s}$: 월별사전조정요인)

C18($\overline{TD_s}$: 최종요일조종요인)

E1($\overline{O_s}^M$: 수정된 원계열)

E2($\overline{CI_s}^M$: 수정된 계절조정계열)

E3($\overline{I_s}^M$: 수정불규칙요인)

위의 계열에 대해, 월 간격 s에 대해 절대변화율을 평균한다.

F2B.원계열(B1)의 월 간격 변화율에 관한 각 구성요인들의 기여도

월 간격 s=1,2,⋯,12에 대해 원계열의 변화율에 대한 각 구성요인들의 상대적 기여도를 다음 식에 의해 계산한다.

$$\overline{O_s}^2 \cong \overline{I_s}^2 + \overline{C_s}^2 + \overline{S_s}^2 + \overline{P_s}^2 \quad (s=1,2,\cdots,12)$$

그러나 각 변화율의 제곱합이 반드시 $\overline{O_s}^2$과 일치하지는 않으므로 $\overline{O_s}'^2 \cong \overline{I_s}^2 + \overline{C_s}^2 + \overline{S_s}^2 + \overline{P_s}^2$ (s=1,2,…,12)로 대체하여 각 구성요인들의 상대적 기여도를 계산한다. 이는 각 구성요소의 변화율에의 상대적 중요도를 나타내는 것이다.

D13의 기여도 : $\overline{I_s}'^2 / \overline{O_s}'^2$ (s=1,2,…,12)

D12의 기여도 : $\overline{C_s}'^2 / \overline{O_s}'^2$ (s=1,2,…,12)

D10의 기여도 : $\overline{S_s}'^2 / \overline{O_s}'^2$ (s=1,2,…,12)

A2의 기여도 : $\overline{P_s}'^2 / \overline{O_s}'^2$ (s=1,2,…,12)

또한 $\overline{O_s}'^2$과 $\overline{O_s}^2$이 어느 정도 근사하는지를 나타내는 척도로서 $\overline{O_s}'^2 / \overline{O_s}^2$ (s=1,2,…,12)을 계산한다.

F2C.각 구성요인들에 대한 변화율의 평균과 표준편차

월 간격 s=1,2,…,12에 대해 각 구성요인들과 MCD에서의 변화율의 평균과 표준편차를 계산한다.

F2D.연(RUN)의 평균

연(RUN)이란, 같은 방향으로 연속적인 전월비의 월수로서, 이의 평균을 구한다. 변동이 없는 경우는 전월변동과 같은 방향으로 월수를 첨가한다. 이 부분은 다음 계열에서 계산된다.

D11(CI : 최종계절조정계열)
D13(I : 최종불규칙계열)
D12(C : 최종추세・순환)
F1(MCD : MCD이동평균)

F2E.MCD 측정을 위한 $\overline{I}/\overline{C}$ 율

MCD 값은 $\overline{I_s}/\overline{C_s} < 1.0$인 월간격의 값이므로 MCD 추정을 위해 먼저 $\overline{I_s}/\overline{C_s}$ (s=1,2,⋯,12)를 계산한다.

F2F.분산이 안정화된 원계열의 각 구성요인들의 기여도

가법형과 log가법형의 선형 추세과 승법형의 지수추세를 제거하여 정상계열로 만든다. 이렇게 안정화된 원계열의 변화율에 대한 각 구성요인들의 상대 기여도를 F2B와 같이 계산한다.

F2G.불규칙요인의 자기상관

월 간격 s=1,2,⋯,14에 따라 불규칙요인의 자기상관을 계산한다.

F2H. $\overline{I}/\overline{C}$율과 $\overline{I}/\overline{S}$율

최종 추세・순환요인(D12)에서 계산된 $\overline{I}/\overline{C}$율과 최종계절요인(D10)에서 계산된 $\overline{I}/\overline{S}$율이다.

F2I.계절성평가에 대한 통계량과 검정확률유의수준

① 원계열(B1)에 대한 계절성 존재에 관한 F검정

원계열(B1)을 중심화 12개월 이동평균하여 추세・순환요인(C)를 산출하여 이를 원계열에서 나누어 계절・불규칙요인(SI′)을 구하여 이를 분산분석법을 이용하여 계절성 존재에 관한 F검정을 하는 것이다.

② 최종계절・불규칙요인(D8)의 안정 계절성 존재에 관한 F검정

최종계절・불규칙요인을 계절요인과 불규칙요인으로 분리하여 각 분산의 구성비인 F통계량으로 계절요인의 유의성을 일원분산분석법으로 검정한다.

③ 최종계절・불규칙요인(D8)의 안정계절성 존재에 관한 비모수 검정 : Kruskall-Wallis 검정

④ 최종계절・불규칙요인(D8)의 이동계절성존재에 관한 검정

이동계절성 검정은 최종계절・불규칙요인(D8)의 이원분산분석에 근거한다. 이것은 계절 폭의 점진적인 변화에 의해 특성지워진 이동계절성의 유무에 관해 검정한다. 이동계절성이 존재한다는 것은 계열의 계절요인이 연간계절성에 의하여 지배를 받게 되므로 계절요인을 X-11-ARIMA방법으로 산출하는 것은 타당하지 않다고 판단된다.

G부분. 도표

G0. 차분과 변수변환 및 사전조정된 원계열(B1)에 대한 스펙트럼 그림

G1.차분과 변수변환된 계절조정계열(E2)에 대한 스펙트럼 그림

G2. 수정된 불규칙요인(E3)에 대한 스펙트럼 그림

<참고 4.4> X-12-ARIMA/GRAPH

SAS에서 pull-down방식으로 이용하기 위해서는 “initx12g.sas"를 SAS 프로그램상에서 실행하면 된다. 그러나 program 방식을 이용하기 위해서는 "x12gmac.sas"을 실행하기 전에 다음과 같이 metafile(x12g.*)이 "C:\X12\GRAPHICS\" folder내에 있어야 한다.

C:\X12\GRAPHICS\x12g.mls
C:\X12\GRAPHICS\x12g.gls

여기서 x12g.mls에는 분석하고자 하는 원계열 명을 지정하며, x12g.gls에는 분석하고자하는 그림의 명령어(keyword)와 X-12-ARIMA output 계열명을 기술한다(표1 참조, Hood, 2001). output 계열명들은 X-12-ARIMA 실행시 Options의 Graphics를 click하면 자동으로 *.spc 이름의 *.gmt 확장자를 갖는 파일이 생성된다. 이때 x12gmac.sas를 SAS 프로그램에서 실행하면 x12g.mls에서 요구한 시계열에 대해서, x12g.gls의 그림이 그려진다. 다음은 X-12-ARIMA\GRAPH의 keyword를 이용한 x12g.gls의 예이다.

overlay: ori sa trn
spectrum: spcori
history: aic fct

그러나 모형비교를 위한 history keyword를 사용하기 위해서는 추가적으로 다음과 같은 사전작업이 필요하다. 먼저 *.spc에서 "history {estimates=(*)}" 명령을 함으로써 history 그래프의 계열을 생성해야 한다. graphic metafile인 *.mls에 동일한 line에 2개의 다른 history 결과 data명을 다음의 ②처럼 입력한다.

<history keyword를 사용하기 위한 사전 작업 단계>

① X-12-ARIMA의 spc파일인 model.spc(요일변수 있는 모형) model_wo.spc(요일변수 없는 모형)에서 history 명령을 각각 실행

② test.mls를 다음과 같이 c:\x12\graphic에 작성

model model_wo

③ test.gls 를 다음과 같이 c:\x12\graphic에 작성

history: aichst fcthst

④ SAS 매그로프로그램인 “x12gmac.sas”에서 "infile=test"로 수정후 실행

< Program 방식의 Keyword와 내용 >

keyword	keyword 내용	output 주요 계열명	계열 내용
overlay	2개 이상 시계열 중복	ori / sa oadori adjori/trn 등	원계열/계절조정된 계열 특이치가 조정된 원계열 사전조정된 원계열/추세요인 등
seas	월(분기)별 시계열 그래프	si sf	월(분기)별 SI 비율 월(분기)별 계절요인
spectrum	스펙트럼 그래프	spcori spcsa spcirr spcosa	원계열의 스펙트럼 계절조정된 계열의 스펙트럼 불규칙요인의 스펙트럼 spcori + spcsa 중복 그래프
cmpnent	각 요인의 그래프	cal/td/hol ls/ao ori/sa/irr/trn	캘런더 요인/요일요인/명절요인 LS, Ramps, TC/가법특이치 기타
tvalue	t-값의 그래프	ao ls tc	
forecast	예측치 그래프	fct ftr	원계열과 예측치 변수변환된 원계열과 예측치
history	History 그래프	aichst/fcthst	AIC의 history/예측오차의 history

주) /는 기술의 편의상 사용한 것임

Ⅴ. 총합계열의 계절조정

5.1 총합계열과 계절조정

어떠한 시계열은 2개 또는 그 이상의 하부 시계열(sub-series) 또는 구성계열(component series)에 의해 결합될 수 있으며, 이러한 시계열을 총합시계열 또는 합성시계열이라 한다. 여기에서는 용어의 통일을 위하여 총합계열이라 한다. 이러한 총합계열은 산업생산지수의 경우, 광업생산지수, 제조업생산지수, 전기・가스・수도업생산지수를 합한 형태이며, 광업생산지수, 제조업생산지수, 전기・가스・수도업생산지수를 총합계열인 산업생산지수의 구성계열이라 한다.

총합계열(aggregate series)은 개별 시계열에 가중치를 부여하여 가・감・승・제(addition・subtraction・multiplication・division)의 사칙연산에 의해 생성된 시계열이다. 즉, j=1, 2, …, K에 대하여, j번째 시계열(Y_{jt})의 가중치가 ω_j이고 K개의 구성계열에 의해 승법 결합되는 총합계열 Y_t를 수식으로 표현하면 다음과 같다.

$$Y_t = \sum_{j=1}^{K} \omega_j Y_{jt} \tag{5.1}$$

총합계열에 대한 계절조정방법은 직접법(direct seasonal adjustment: DA)과 간접법(indirect seasonal adjustment: IA)이 있다. 직접법은 계절조정을 하지 않은 개별 구성시계열들을 사칙연산에 의해 합한 총합계열을 직접 계절조정하는 방법이다. 간접법은 개별 시계열을 계절조정한 후, 계절조정한 개별 시계열을 사칙연산에 의해 결합하여 계절조정된 총합계열을 구하는 방법이다(Hood와 Findley, 2001).

총합계열을 Y_t, 개별 시계열을 X_t와 Z_t라 가정하면, 시계열 구성요인을 다음과 같이 비계절요인과 계절요인으로 분리 표현할 수 있으며, 설명의 편의상 가법적 결합을 가정한다.

$$\begin{aligned} Y_t &= SA(Y_t) + SF(Y_t) \\ X_t &= SA(X_t) + SF(X_t) \\ Z_t &= SA(Z_t) + SF(Z_t) \end{aligned}$$

여기서 $SA(Y_t), SA(X_t), SA(Z_t)$는 추세・순환요인, 불규칙요인을 포함하고

있는 계절조정계열이며, $SF(Y_t)$, $SF(X_t)$, $SF(Z_t)$는 계절요인이다. 이때 총합계열 Y_t의 계절조정계열은 다음과 같이 표현할 수 있다.

	총합계열 산식	총합계열의 계절조정		예
		직접법	간접법	
가법	$Y_t = X_t + Z_t$	$SA(Y_t)$	$SA(X_t)+SA(Z_t)$	산업생산지수
감법	$Y_t = X_t - Z_t$	$SA(Y_t)$	$SA(X_t)-SA(Z_t)$	상품수지
승・제법	$Y_t = X_t / Z_t$	$SA(Y_t)$	$SA(X_t)/SA(Z_t)$	실업률

주) 상품 수지=상품수출-상품수입, 실업률=실업자/경제활동인구

(예 5.1) 다음 시계열은 1995년 1월부터 2005년 12월까지의 상품수지(=상품 수출-상품 수입)에 대한 계절조정결과이다. 다음 <표 5.1>의 결과는 2005년 1월부터 12월까지 원계열, 상품 수출・입 계절조정계열 및 직・간접법에 의한 상품수지 계절조정계열이다. 간접법 계절조정방법은 <참고 5.1>를 참고할 수 있다.

<표 5.1> 상품수지의 원계열과 계절조정계열

(단위: 백만달러)

'05년	원 계 열			계절조정계열			
	수 출	수 입	상품수지	수 출	수 입	직접법	간접법
1월	23378.8	18887.2	4491.6	24089.9	19161.7	4643.4	4928.2
2	19515.1	17925.1	1590.0	21149.9	19165.1	1977.9	1984.8
3	25156.3	22082.2	3074.1	23764.2	20469.4	3155.3	3294.8
4	22993	20662.9	2330.1	22573.6	20112.6	2481.0	2461.0
5	23287.9	20677.5	2610.4	23364.4	21139.1	2439.0	2225.3
6	24346.9	20646.9	3700.0	24123.8	21156.1	2971.3	2967.6
7	24140	20973.8	3166.2	24308.4	21328.2	3198.2	2980.2
8	23070.5	21484.7	1585.8	24818.7	22311.5	2205.5	2507.3
9	25151.1	22320.2	2830.9	25189.5	22858.6	2415.5	2331.0
10	25427.7	22317.5	3110.2	24477.1	21610.6	2946.5	2866.5
11	26846.6	23668.5	3178.1	25770.4	23085.0	2828.9	2685.4
12	25681.7	23876.1	1805.6	25022.4	22979.9	2197.2	2042.4

직접법에 의한 상품수지의 계절조정계열(결과표: *.D11)은 (수출-수입)과 일치하지 않으나, 간접법에 의한 상품수지의 계절조정계열(결과표:

*.ISA)는 (수출-수입)계열과 일치한다. 원계열의 연간 합과 계절조정계열의 연간 합을 일치시킨 수출과 수입의 계절조정계열은 *.SAA에 저장된다. 이때 총합계열인 상품수지의 계절조정계열(수출 계절조정계열-수입 계절조정계열)과는 일치하지 않는다. 이는 상품수지의 연간 합을 일치시키기 위하여 계절조정계열을 조정하였기 때문이다.

이론적으로 총합계열의 계절조정계열은 다음 식 (5.2)와 같이 개별 시계열의 계절조정계열 합과 일치하여야 한다.

$$SA(Y_t) \;=\; SA(X_t) \;+\; SA(Z_t) \tag{5.2}$$

그러나 일반적으로 (5.2)의 식은 계절조정방법에 비선형성(non-linearity)이 포함되어 있으므로 만족하지 않을 수 있다. 즉, 2개 이상의 구성계열을 결합한 총합 시계열의 계절조정계열은 (예 5.1)과 같이 계절조정방법을 직접법 또는 간접법에 의해서 했느냐에 따라 그 결과는 달리 나타난다.

직접법에 의한 계절조정 결과는 총합계열 자체를 계절조정한 $SA(Y_t)$와 개별 구성계열을 각각 계절조정한 후 결합한 $SA(X_t)+SA(Z_t)$는 차이를 보일 수 있다. 그러나 간접법의 경우, 개별 구성계열을 계절조정한 후, 개별 계절조정계열을 결합($SA(X_t)+SA(Z_t)$)하여 총합 계절조정계열을 만들게 되므로 일치하게 된다. 이러한 차이로 총합계열에 대한 각 구성계열의 전월비 기여도 계산에 있어, 직접법의 경우 전월비 기여도의 합이 총합계열의 전월비와 일치하지 않게 된다((예 5.2) 참조).

구성계열 k 계절조정계열($SA(k)_t$)의 총합 계절조정계열에 대한 전월비 기여도($SA^c(k)$)를 다음과 같이 정의할 수 있다.

$$SA^c(k_t) = \frac{SA(k_t) \;-\; SA(k_{t-1})}{SA(Y_{t-1})} \cdot \omega_k \times 100(\%), \quad k = 1,2,\ldots,K$$

여기서 ω_k는 k계열의 가중치이다.

(예 5.2) 다음은 1995년 1월부터 2004년 12월까지 소비재판매액지수의 계절조정결과이다. 사전조정은 default 옵션에 의해 특이치만 구하였으며, ARIMA 모형은 AUTOMDL 명령에 따라 자동 선정하였다. 이때 가중치는 내구재(DG) 0.2543, 준내구재(DSG) 0.2403, 비내구재(NDG) 0.5054이다.

각 계열의 계절조정계열과 전월비(<표 5.2>) 및 기여도(<표 5.3>)는 다음과 같다. <표 5.3>의 기여도를 보면, 간접법의 기여도 합은 전월비와 일치하나 직접법의 기여도는 전월비와 일치하지 않음을 알 수 있다.

<표 5.2> 소비재 계절조정계열 및 전월비

	계 절 조 정 계 열					전월비		
	DG	SDG	NDG	간접법	직접법	DG	SDG	NDG
04.1월	103.7	114.3	123.4	116.2	115.6	-5.2	0.0	5.5
2월	110.2	117.5	115.6	114.7	114.0	6.3	2.8	-6.3
3월	105.0	110.4	118.3	113.0	113.3	-4.7	-6.0	2.3
4월	106.8	115.5	116.7	113.9	113.8	1.6	4.6	-1.3
5월	101.1	117.3	115.9	112.5	113.2	-5.3	1.5	-0.7
6월	107.0	117.6	119.3	115.8	116.5	5.9	0.3	2.9
7월	109.6	115.9	117.0	114.9	114.7	2.5	-1.5	-2.0
8월	109.7	110.2	115.0	112.5	112.9	0.0	-4.9	-1.7
9월	111.2	115.8	118.2	115.9	115.8	1.4	5.0	2.8
10월	111.8	118.0	114.4	114.6	114.5	0.5	1.9	-3.3
11월	113.6	116.6	114.7	114.9	114.6	1.7	-1.2	0.3
12월	117.1	119.1	116.6	117.3	117.1	3.1	2.1	1.7

주) DG: 내구재 SDG: 준내구재 NDG: 비내구재

<표 5.3> 소비재 구성계열의 기여도(%p)

	기여도(간접법)					기여도(직접법)				
	전월비	DG	SDG	NDG	합	전월비	DG	SDG	NDG	합
04. 1월	1.6	-1.3	0.0	2.8	1.6	1.0	0.3	0.3	0.4	1.0
2월	-1.3	1.4	0.7	-3.4	-1.3	-1.4	-0.4	-0.6	-1.0	-2.0
3월	-1.5	-1.2	-1.5	1.2	-1.5	-0.6	-0.2	0.0	0.3	0.1
4월	0.8	0.4	1.1	-0.7	0.8	0.4	0.1	0.5	0.5	1.1
5월	-1.2	-1.3	0.4	-0.4	-1.2	-0.5	-0.1	-0.4	-0.4	-1.0
6월	2.9	1.3	0.1	1.5	2.9	2.9	0.7	0.9	1.5	3.2
7월	-0.8	0.6	-0.4	-1.0	-0.8	-1.5	-0.4	-0.8	-1.9	-3.1
8월	-2.1	0.0	-1.2	-0.9	-2.0	-1.6	-0.4	-0.3	0.0	-0.7
9월	3.0	0.3	1.2	1.4	3.0	2.6	0.7	1.1	1.9	3.6
10월	-1.1	0.1	0.5	-1.7	-1.1	-1.1	-0.3	-0.9	-1.6	-2.8
11월	0.3	0.4	-0.3	0.1	0.2	0.1	0.0	0.3	0.5	0.8
12월	2.1	0.8	0.5	0.8	2.1	2.2	0.6	0.4	0.9	1.9

어떠한 시계열의 구성계열이 뚜렷하게 다른 계절요인 형태를 보이고 계절조정이 잘 되었다면, 일반적으로 간접법 계절조정에 의한 해석이 좋

은 결과를 얻을 수 있다. 간접법 계절조정 결과는 구성계열과 총합계열의 계절조정 결과 간 해석이 일치하기 때문에 많은 사용자에게 선호되는 방법이다(U.S. Census Bureau, 2004). 그러나 간접법에 의한 계절조정은 많은 작업시간이 소요된다는 점과 함께 시계열 분류가 2가지 이상인 경우, 각 분류에 의한 총합계열의 계절조정 결과가 달리 나타날 수 있는 단점이 있다. 예를 들어, 국내총생산(GDP)의 경우 생산측면과 지출측면의 GDP가 작성된다. 이때 원계열의 GDP는 생산측면과 지출측면의 합이 일치하나, 계절조정계열은 생산측면과 지출측면의 계절요인이 다르기 때문에 생산측면의 계절조정계열 합과 지출측면의 계절조정계열 합이 일치하지 않게 된다. 즉, 생산측면의 GDP 계절조정계열과 지출측면의 GDP 계절조정계열이 일치하지 않는다. 그러나 공통의 최하 단위 계열을 이용하여 다른 형태로 분류하는 경우에는 동일한 결과를 얻을 수 있다(예 5.3 참조).

X-12-ARIMA에서는 간접법에 의한 계절조정은 “COMPOSITE" 명령에 의해서 수행되며, 결과는 직접법에 의한 계절조정 결과와 간접법에 의한 계절조정 결과를 동시에 제공해준다(자세한 내용은 PART B의 SERIES 명령과 COMPOSITE 명령을 참조).

5.2 두 개의 분류를 갖는 시계열의 계절조정

남・녀, 취업자・실업자로 분류하는 경제활동인구와 같이, 두 개의 분류(two way classification)를 갖는 시계열을 직접법에 의해 계절조정한다고 가정하자. 이때 이들 분류를 <표 5.4>와 같이 나타낼 수 있으며, 직접법에 의한 계절조정 결과는 다음과 같은 가정을 생각할 수 있다.

<표 5.4> 경제활동인구 분류

	남	여	합
취업자	a	b	C
실업자	d	e	F
합	G	H	I

가정 ⅰ) 취업자 계절조정계열(C)은 취업자 남자 계절조정계열(a)과 취업자 여자 계절조정계열(b)의 합과 일치하지 않을 수 있다.

가정 ⅱ) 남자 계절조정계열(G)과 여자 계절조정계열(H)의 합은 전체 경제활동인구의 계절조정계열(I)과 일치하지 않을 수 있다.
가정 ⅲ) 남자 취업자 계절조정계열(C)과 남자 실업자 계절조정계열(F)의 합은 전체 경제활동인구 계절조정계열(I)과 일치하지 않을 수 있다.
가정 ⅳ) 각 개별 구성계열인 남자 취업자(a), 여자 취업자(b), 남자 실업자(d), 여자 실업자(e) 계절조정계열은 전체 경제활동인구(I) 계절조정계열과 일치하지 않을 수 있다.

호주 통계청(ABS)에서는 시계열 값이 양의 값을 갖는 경우, 비율 추정법 또는 반복 비율 적합법(Raking ration estimation 또는 Iterative proportional fitting), 제약 최소법(Constrained minimization) 등의 Calibration 방법을 이용하여 이러한 불일치성 문제를 해결하고 있다(Cannon과 Gemma, 2000). 캐나다 통계청에서는 직접법에 의해 추정된 총합계열의 합을 일치시키기 위한 Raking법은 원래 추정된 계절조정계열의 전월비를 크게 변화시키지 않는 경우에 사용할 것을 권장하고 있다(Statistics Canada, 2003).

5.2.1 비율 추정법

비율 추정법은 개별 구성계열에 대한 계절조정계열의 합과 총합계열에 대한 계절조정계열 간에 차이가 있다면, 그 차이만큼 조정해 주는 방법이다. 예를 들어 <표 5.4>의 행 합이 전체 합과 일치하지 않는 경우 ($G+H \neq I$), $I-(G+H)$ 만큼 차이가 발생하므로 남자 계절조정계열(G)에는 I/(G+H), 여자 계절조정계열(H)에는 I/(G+H) 비율만큼 조정해 준다.

이와 같은 원리에 따라 가정 ⅲ)의 경우, 전체 행과 열, 그리고 열과 행 원소(cell)들의 합을 전체 합과 일치시키기 위하여 다음과 같이 반복하여 비율을 조정해 준다.

STEP 1: 열 합(C+F)과 전체 합(I)이 일치하도록 열 합의 비율을 조정해 준다.

$$\hat{C}= C\times\frac{I}{(C+F)}, \qquad \hat{F}= F\times\frac{I}{(C+F)}$$

STEP 2: 행 합(G+H)과 전체 합(I)이 일치하도록 행 합의 비율을 조정해 준다.

$$\hat{G}= G\times\frac{I}{(G+H)},\qquad \hat{H}= H\times\frac{I}{(G+H)}$$

STEP 3: 열 원소의 합(a+d)과 열 합($\hat{G}$)이 일치하도록 열 원소의 비율을 조정해 준다.

$$\hat{a}^{(1)}= a\times\frac{\hat{G}}{(a+d)},\qquad \hat{d}^{(1)}= d\times\frac{\hat{G}}{(a+d)}$$
$$\hat{b}^{(1)}= b\times\frac{\hat{H}}{(b+e)},\qquad \hat{e}^{(1)}= e\times\frac{\hat{H}}{(b+e)}$$

STEP 4: 행 원소의 합($\hat{a}^{(1)}+\hat{b}^{(1)}$)과 행 합($\hat{C}$)이 일치하도록 행 원소의 비율을 조정해 준다.

$$\hat{a}^{(2)}= \hat{a}^{(1)}\times\frac{\hat{C}}{(\hat{a}^{(1)}+\hat{b}^{(1)})},\qquad \hat{b}^{(2)}= \hat{b}^{(1)}\times\frac{\hat{C}}{(\hat{a}^{(1)}+\hat{b}^{(1)})}$$
$$\hat{d}^{(2)}= \hat{d}^{(1)}\times\frac{\hat{F}}{(\hat{d}^{(1)}+\hat{e}^{(1)})},\qquad \hat{e}^{(2)}= \hat{e}^{(1)}\times\frac{\hat{F}}{(\hat{d}^{(1)}+\hat{e}^{(1)})}$$

(예 5.3) 다음은 1999년 6월부터 2005년 5월까지의 경제활동인구를 남·여, 취업자·실업자로 분류하여 계절조정한 결과이다. 계절조정하기 위하여 원계열을 log 변환하였으며, 사전조정과 계절조정은 default 옵션을 사용하였다.

다음 <표 5.5>는 2005년 5월의 원계열과 직접법 계절조정계열(D11A)의 값이다. 계절조정계열에 있어 남자 취업자는 1,339.0만명, 여자 취업자는 951.6만명으로 이들을 합하면 2,290.6만명이다. 그러나 남·여 취업자 합에 대한 계절조정 결과는 2,291.0만명으로 각각 계절조정한 계열의 합과는 차이가 있다. 한편 남자 실업자와 여자 실업자 합에 대한 계절조정 결과는 88.0만명이나 남자 실업자와 여자 실업자에 대한 각각 계절조정계열의 합은 88.1만명이다.

<표 5.5> 2005년 5월 경제활동인구 원계열과 계절조정계열(직접법)

원계열(단위: 천명)				계절조정계열			
	남	여	합		남	여	합
취업자	13496.0	9702.6	23198.6	취업자	13389.9	9516.3	22910.2
실업자	517.5	334.8	852.3	실업자	542.8	338.0	880.4
합	14013.5	10037.4	24051.0	합	13932.9	9848.1	23785.5

경제활동인구(2,378.6만명)를 고정시키고, 각 구성계열의 계절조정 합, 행과 열의 계절조정 계열을 비율 추정법에 의해 행과 열의 합을 일치시키면 다음과 같은 결과를 얻을 수 있다. 즉, 남자, 여자 취업자에 대한 계절조정 결과의 합은 취업자의 계절조정 결과와 일치한다. 남자 취업자와 실업자 계절조정 결과의 합은 남자 경제활동인구 계절조정 결과와 일치한다. 또한 취업자와 실업자의 계절조정 결과는 경제활동인구와 일치한다.

(단위: 천명)

	남	여	합
취업자	13393.1	9512.5	22905.6
실업자	542.4	337.5	879.9
합	13935.6	9849.9	23785.5

이때 최근 1년의 계절조정계열과 비율 추정법에 의해 추정한 계열의 전월비(<표 5.6-1>과 <표 5.6-2>)를 보면, 큰 차이가 없어 Statistics Canada (2003)에서 권장하는 바와 같이 비율법에 의해 조정된 결과를 사용해도 큰 문제는 없어 보인다.

<표 5.6-1> 경제활동인구 전월비(%)

	2004년							2005년				
	6월	7월	8월	9월	10월	11월	12월	1월	2월	3월	4월	5월
경활인구	0.2[1)]	-0.1	-0.2	0.8	0.1	0.0	0.4	-0.2	0.1	0.4	0.1	0.4
	0.2[2)]	-0.1	-0.2	0.8	0.1	0.0	0.4	-0.2	0.1	0.4	0.1	0.4
실업자	0.0	1.9	-0.1	0.3	-1.3	0.4	0.9	0.3	-0.2	-0.1	1.7	-1.4
	0.0	1.9	-0.1	0.3	-1.2	0.3	1.0	0.1	-0.1	-0.1	1.6	-1.4
취업자	0.2	-0.2	-0.2	0.9	0.1	0.1	0.3	0.0	0.1	0.3	0.1	0.5
	0.2	-0.2	-0.2	0.8	0.2	0.0	0.4	-0.3	0.1	0.4	0.1	0.5

주 1) 계절조정계열(D11A) 전월비, 2) 비율 추정법으로 조정된 계절조정계열 전월비

<표 5.6-2> 경제활동인구 전월비(%)

	2004년							2005년				
	6월	7월	8월	9월	10월	11월	12월	1월	2월	3월	4월	5월
남	0.2	-0.2	-0.1	0.6	0.2	0.1	0.1	-0.1	0.1	0.5	0.2	0.2
	0.2	-0.2	-0.1	0.5	0.3	0.1	0.0	0.0	0.1	0.5	0.2	0.2
여	0.3	0.0	-0.3	1.3	-0.2	0.0	1.1	-0.7	0.1	0.2	0.0	0.6
	0.3	-0.1	-0.3	1.2	-0.1	0.0	1.0	-0.6	0.1	0.2	0.0	0.6

(예 5.4) 다음은 (예 5.3)를 간접법에 의해 계절조정하는 프로그램 예이다. 계절조정하기 위한 SPC 파일 이름을 다음과 같이 정의하고 각 cell의 SPC 파일을 작성한다. 또한 열과 행의 통합 SPC 파일을 작성한다. 즉, 남ㆍ녀 취업자 합(EAPE)를 구하기 위한 통합 SPC 파일인 COMP_EAPE 등을 작성한다(<참고 5.1> 참조).

	남	여	합
취업자	EAPEM	EAPEF	EAPE
실업자	EAPUM	EAPUF	EAPU
합	EAPM	EAPF	EAP

EAPEM.SPC	COMP_EAPE
series{ Title=" EAPEM" start=1999.06 period=12 decimals=1 COMPTYPE=ADD COMPWT=1 span=(1999.06,2005.5) file="c:\comp_sa05\eapem.txt" format="tramo" } transform{function=log} automdl { } outlier {types=all} x11{ force=totals save=(d11 saa) }	composite{TITLE=" EAP EMP." name="EAPE" decimals=1 save=(isf isa) modelspan=(1999.06,2005.05)} x11{ force=totals save=(d10 d11 d12) }

간접법에 의해 취업자, 실업자, 남자, 여자, 전체 경제활동인구의 계절조정계열을 추정하기 위해서는 행과 열의 합에 대한 다음과 같은 5개의 meta 파일이 필요하다.

전 체 EAP.MTA	남 자 EAPM.MTA	여 자 EAPF.MTA	취 업 자 EAPE.MTA	실 업 자 EAPU.MTA
EAPEM	APEM	EAPEF	APEM	EAPUM
EAPEF	EAPUM	EAPUF	EAPEF	EAPUF
EAPUM	COMP_EAPM	COMP_EAPF	COMP_EAPE	COMP_EAPU
EAPUF				
COMP_EAP				

다음 <표 5.7>은 간접법에 의한 계절조정 결과의 일부('05.1월~5월)이다. 간접법의 경우, 하위 계열인 남ㆍ녀 취업자, 남ㆍ녀 실업자에 대한 계절조정 결과를 결합하여 상위 계열인 남자, 여자, 취업자, 실업자, 전체 계절조정계열을 작성하므로 직접법에서 발생하였던 분류에 따른 계절조정계열의 불일치 현상은 나타나지 않는다.

<표 5.7> 경제활동인구 계절조정계열(간접법)

(단위: 천명)

	취업자		실업자		남 자	여 자	취업자	실업자	경 활 인 구
	남 자	여 자	남 자	여 자					
'05.1	13224.6	9441.8	557.3	328.5	13782.0	9770.3	22666.4	885.8	23552.2
2	13247.5	9440.2	551.0	328.1	13798.5	9768.2	22687.6	879.1	23566.7
3	13317.1	9468.5	554.5	322.2	13871.6	9790.6	22785.6	876.6	23662.2
4	13345.4	9454.0	556.9	335.6	13902.3	9789.6	22799.4	892.5	23691.9
5	13389.9	9516.3	542.8	338.0	13932.7	9854.2	22906.1	880.8	23786.9

5.2.2 최소 제약법

최소 제약법은 라그랑지 승수(Lagrange multiliers)를 이용하여 행렬의 원소들이 행과 열의 합과 일치하도록 하는 제약식 하에서 거리함수(distance function)가 최소가 되도록 하는 방법이다. 제약식인 거리함수는

행렬의 원소 값이 변하나, 추정된 원소는 원래의 원소와의 차이가 최소가 되도록 한다. 이때 거리함수는 최소자승함수(Least square distance function) 등을 사용한다.

예를 들어 가정 iii)의 경우, 모든 원소가 행과 열의 합 및 전체 합과 동시에 일치하도록 하기 위하여 먼저 모든 원소 및 합에 승수 가중치 α_i를 곱해준다.

	남	여	합
취업자	α_1*a	α_2*b	α_3*C
실업자	α_4*d	α_5*e	α_6*F
합	α_7*G	α_8*H	α_9*I

이때 $\alpha_i = 1$이면, 조정된 값이 조정하지 않은 원래의 값과 같게 된다. 따라서 계절조정계열의 변동을 최소화하기 위하여 α_i는 1에 가깝도록 최소자승함수를 다음과 같이 설정하고, 행과 열은 다음과 같은 제약 조건을 갖는다.

최소화: $(\alpha_1 - 1)^2 + (\alpha_2 - 1)^2 + \ldots + (\alpha_8 - 1)^2 + (\alpha_9 - 1)^2$ (5.3)

제약조건: $\alpha_1{*}a + \alpha_2{*}b - \alpha_3{*}C = 0$ (5.4)

$\alpha_4{*}d + \alpha_5{*}e - \alpha_6{*}F = 0$ (5.5)

$\alpha_7{*}G + \alpha_8{*}H - \alpha_9{*}I = 0$ (5.6)

$\alpha_1{*}a + \alpha_4{*}d - \alpha_7{*}G = 0$ (5.7)

$\alpha_2{*}b + \alpha_5{*}e - \alpha_8{*}H = 0$ (5.8)

$\alpha_3{*}C + \alpha_6{*}F - \alpha_9{*}I = 0$ (5.9)

각 열의 값을 합하면 식 (5.7)과 같이 되므로 6개의 제약식 중 하나는 중복이 되므로 식 (5.9)를 제거한다. 식 (5.3)를 최적화하기 위하여 식 (5.3)~(5.8)를 α_i에 대하여 편미분을 한다. 이때 각 제약조건에 보조변수(auxilliary variables) λ_i를 적용한 후 λ_i와 α_i에 대해서 편미분하고, 가우시안 제거법(Gaussian elimination)에 의하여 해를 구한다. 편미분 결과는 다음과 같다.

α_1	α_2	α_3	α_4	α_5	α_6	α_7	α_8	α_9	λ_1	λ_2	λ_3	λ_4	λ_5	RHS
1	0	0	0	0	0	0	0	0	a	0	0	a	0	1
0	1	0	0	0	0	0	0	0	b	0	0	0	b	1
0	0	1	0	0	0	0	0	0	−C	0	0	0	0	1
0	0	0	1	0	0	0	0	0	0	d	0	d	0	1
0	0	0	0	1	0	0	0	0	0	e	0	0	e	1
0	0	0	0	0	1	0	0	0	0	−F	0	0	0	1
0	0	0	0	0	0	1	0	0	0	0	G	−G	0	1
0	0	0	0	0	0	0	1	0	0	0	H	0	−H	1
0	0	0	0	0	0	0	0	1	0	0	−I	0	0	1
a	b	−C	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	d	e	−F	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	G	H	−I	0	0	0	0	0	0
a	0	0	d	0	0	−G	0	0	0	0	0	0	0	0
0	b	0	0	e	0	0	−H	0	0	0	0	0	0	0

조정된 원소 값은 가우시안 제거법에 의해 추정된 $\hat{\alpha}_i$를 원래의 원소 값에 곱해줌으로써 구할 수 있다. 그러나 총합(I)는 고정되었다고 가정한다면, 다음과 같이 승수 가중치를 부여하여 제약조건을 설정할 수 있다

	남	여	합
취업자	α_1*a	α_2*b	α_3*C
실업자	α_4*d	α_5*e	α_6*F
합	α_7*G	α_8*H	I

5.3 총합계열의 계절조정결과 진단

일반적으로 개별 구성계열의 계절 형태(seasonal pattern)가 뚜렷하게 다르고 계절조정 결과가 적절하다면, 대부분 시계열의 경우 간접법에 의한 총합 계절조정 결과는 직접법보다 우수하다. 그러나 개별 구성계열이 noise를 갖고 있고 계절요인이 비슷한 경우에 간접법에 의한 계절조정방법을 이용하면 noise가 상쇄되므로 정보를 상실하게 된다. 따라서 이러한 경우에는 개별 시계열이 갖고 있는 정보를 유지하기 위하여 간접법보다는 직접법에 의한 계절조정이 보다 우수한 결과를 얻을 수 있다(Hood와

Findley, 2001).

직접법과 간접법 계절조정을 비교하기 위한 방법 중 하나는 평활도를 보는 것이다. X-11-ARIMA(Dagum, 1988)에서는 총합계열의 계절조정방법을 평활도(smoothness) 정도에 의해서 결정하고 있으며, R_1과 R_2 두 가지 통계량을 제공하고 있다. 계절조정계열을 $\widehat{SA}_t$라 하면, R_1 평활도는 다음과 같이 계절조정계열에 대한 1차 차분 자승합으로 나타낸다.

$$R_1 = \sum_{t=2}^{T} (\widehat{SA}_t - \widehat{SA}_{t-1})^2$$

이때 R_1이 크다면, 평활이 덜 되었다고 할 수 있다. R_2는 R_1의 평활도에서 단기주기의 순환성을 제거하기 위하여 13-항 핸더슨(H) 필터를 기초로 다음과 같이 정의하고 있다(핸더슨 필터는 4.2.다 참조).

$$R_2 = \sum_{t=1}^{T} (\widehat{SA}_t - H\widehat{SA}_t)^2 = \sum_{t=1}^{T} [(I - H)\widehat{SA}_t]^2$$

일반적으로 두 통계량은 거의 비슷한 결과를 얻는다. 그러나 총합계열이 단기주기의 순환 변동에 영향을 크게 받으면, 계절조정계열의 평활성은 R_2에서 더 잘 나타난다. X-12-ARIMA에서는 전체 기간 및 최근 3년의 R_1과 R_2 뿐만 아니라 R_1과 R_2의 차이백분율(Δ)을 제공해 준다.

$$\Delta = \frac{R^{direct} - R^{indirect}}{R^{direct}} \times 100(\%) \qquad (5.10)$$

이때 Δ가 양수이면, 간접법이 선호된다.

한편, Hood와 Findley(2001)는 다음의 조건을 만족할 때 총합계열의 계절조정 결과가 우수하다고 평가하고 있다.

- 시계열 수준의 편의가 작아야 한다. 즉, 원계열과 계절조정계열간의 수준이 비슷하여야 한다.
- 계절조정 결과에 잔차 계절성(residual seasonality)이 없어야 한다.
- 계절조정 결과는 안정적(stability)이여야 한다.

잔차 계절성을 탐색하기 위한 통계량으로 F-검정 통계량을 사용한다. 그러나 시계열의 길이가 충분히 길다면 그래프적인 방법인 스펙트럼 분석도 잔차 계절성과 요일효과(trading day)을 진단하기 위한 방법이다.

강한 계절성을 갖는 월간자료의 경우, 스펙트럼 그림을 그려보면 시계열이 매년 12개월, 6개월 또는 3개월의 주기를 갖고 상승과 하락을 반복하므로 계절주기에서 높은 진폭(amplitude)을 갖는다. 즉, 최종 계절조정계열의 k/12($1 \leq k \leq 6$) 주기에서 정점을 보인다면, 계절정점을 갖는다고 볼 수 있으며, 잔차 계절성이 있음을 보여준다. 분기자료의 경우는 1/4 주기, 반기자료의 경우는 1/2 주기에서 계절정점을 갖는다. X-12-ARIMA에서는 이러한 잔차 계절성의 존재를 보기 위하여 계절조정계열과 불규칙요인의 스펙트럼 그림을 제공해 주고 있다(스펙트럼분석은 4.4.나 참조).

X-12-ARIMA에서 계절조정의 안정성을 보기 위하여 기간이동 분석(sliding spans)과 수정율 분석(revision history)을 제공해 주고 있다. 기간이동 분석은 시계열 분석 기간을 기간이 중복 되도록 4개의 부분 시계열로 재구성한 후, 각 부분 시계열에 대한 계절조정 결과를 분석한다(기간이동분석은 4.4.다 참조).

한편 수정율은 시계열이 새로 추가되었을 때, 초기 추정치와 최종 추정치간의 차이를 수정율이라 하고 이를 분석한다. 수정율이 크다면, 계절조정의 불안정성을 나타낸다(수정율은 4.4.라 참조).

X-11-ARIMA와 X-12-ARIMA는 직접법 또는 간접법에 의한 계절조정 시, M과 Q-통계량을 제공해 준다. 이들 통계량은 사용자에게 계절조정계열이 갖고 있는 어떠한 잠재적인 문제점을 보여주기 위한 것이지, 직접법 혹은 간접법 계절조정방법을 선정하기 위한 것은 아니다. 즉, M과 Q-통계량은 직접법 또는 간접법 계절조정방법을 선정하기 위하여 사용하기에는 어렵다.

5.4 간접법에 의한 총합계열의 계절조정 예

소비재판매액지수(CG)는 다음과 같이 내구재(DG), 준내구재(SDG), 비내구재(NDG)을 합한 지수이다.

$$CG=0.2543*DG + 0.2403*SDG + 0.5054*NDG$$

간접법에 의한 소비재판매액지수의 계절조정방법은 다음과 같다. 먼저, 소비재판매액지수의 하위 구성계열인 DG, SDG, NDG에 대한 계절조정을 위한 *.SPC 화일을 작성한다. 이때 하위 구성계열의 결합방법과 가중치 등은 각 계열에 대해서(즉, 각각의 SPC 파일) SERIES { comptype=· compwt=· } 명령에서 다음과 같이 정해준다.

```
series{  Title=" Durable Goods Series"
         start=1995.01
           ~~~~~~
          COMPTYPE=ADD
          COMPWT=0.2543 }
```

구성계열의 계절조정계열을 합하여 총합계열의 계절조정계열을 생성하기 위한 spc 파일(여기에서는 COMP_CG.SPC)에는 COMPOSITE 명령을 이용한다. COMPOSITE 명령은 간접법으로 계절조정을 하기 위한 명령이므로 반드시 있어야 한다.

개별 시계열에 대한 계절조정과 이들 명령을 통합하기 위한 meta file (CG.MTA)을 다음과 같이 작성한다.

```
DG
SDG
NDG
COMP_CG
```

X-12-ARIMA에서 실행은 CG.MTA를 실행한다.

(예 5.5) 다음은 1995년 1월부터 2004년 12월까지 소비재판매액지수의 계절조정 결과이다. 사전조정은 default 옵션에 의해 특이치만 구하였으며, ARIMA 모형은 AUTOMDL에 따라 자동 선정하였다.

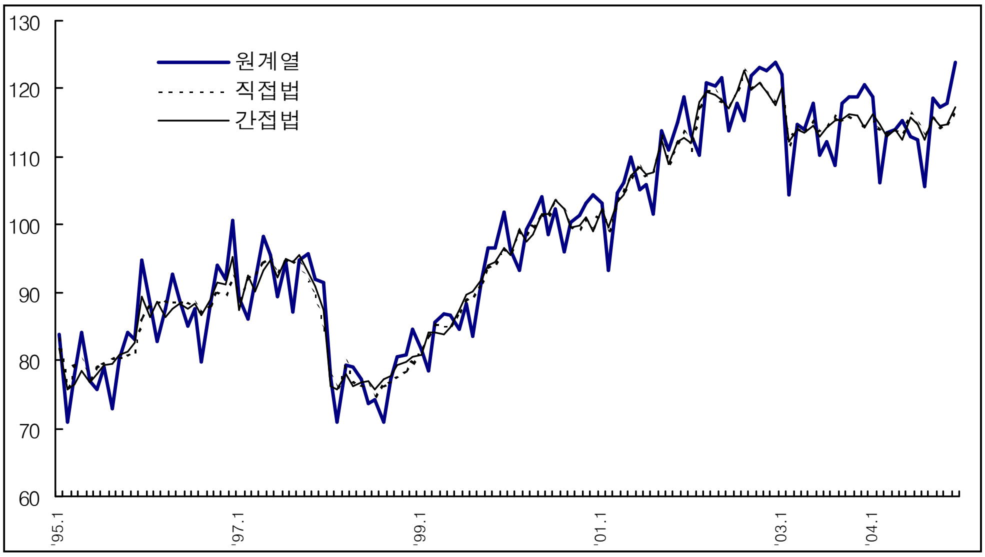
간접법에 의한 계절조정 결과에 대한 잔차 계절성을 보기 위하여 최종 간접법 계절조정계열인 D11 표 또는 D11.A 표의 F-검정 결과를 보면 잔차 계절성이 없음을 알 수 있다. X-12-ARIMA에서는 직접법과 간접법에 의한 계절조정 계열를 제공해 줌으로 반드시 “Indirect Seasoanlly Adjusted data(D11 또는 D11.A)"를 이용하여야 한다.

" Test for the presence of residual seasonality.
No evidence of residual seasonality in the entire series at the 1 per cent level. F = 0.32
No evidence of residual seasonality in the last 3 years at the 1 per cent level. F = 0.57
No evidence of residual seasonality in the last 3 years at the 5 per cent level. "

잔차 계절성의 존재 여부는 간접법 계절조정계열과 불규칙요인에 대한 스펙트럼 결과(OUTPUT의 G.1과 G.2)에 의해서도 확인 가능하다.

다음 <그림 5.1>은 직접법과 간접법에 의해 산출된 계절조정계열이며, <표 5.8>은 내구재, 준내구재, 비내구재와 직접법 및 간접법에 의한 소비재판매액지수의 계절조정계열에 대한 전월비이다. 가중 합은 내구재, 준내구재, 비내구재의 전월비에 각각 해당 가중치를 부여하여 계산한 총합 전월비이다.

<그림 5.1> 소비재의 원계열과 계절조정계열



<표 5.8> 결과를 보면, 간접법에 의한 총합계열의 전월비가 구성계열을 전월비를 합한 가중 합과 유사함으로써, 직접법에 의한 전월비보다는 간접법에 의한 전월비가 구성 계열의 전월비를 더 잘 반영하고 있음을 보여주고 있다. 그러나 개별지수를 총합한 계열의 전월비는 간접법의 전월

비와 일치한다.

<표 5.8> 소비재 전월비(%)

	내구재 (0.2543)주	준내구재 (0.2403)	비내구재 (0.5054)	직접법	간접법	가중 합
1월	-4.1	0.0	5.5	1.0	1.6	1.7
2월	3.8	2.8	-6.3	-1.4	-1.3	-1.5
3월	-3.4	-6.0	2.3	-0.6	-1.5	-1.2
4월	2.1	4.6	-1.3	0.5	0.8	1.0
5월	-4.4	1.5	-0.7	-0.5	-1.3	-1.1
6월	4.4	0.3	2.9	2.9	3.0	2.7
7월	3.0	-1.5	-2.0	-1.5	-0.8	-0.6
8월	0.0	-4.9	-1.7	-1.6	-2.1	-2.1
9월	0.3	5.0	2.8	2.6	3.0	2.7
10월	1.9	1.9	-3.3	-1.1	-1.1	-0.7
11월	0.6	-1.2	0.3	0.1	0.3	0.0
12월	1.5	2.1	1.7	2.2	2.1	1.8

주) 내구재, 준내구재, 비내구재의 소비재에 대한 가중치임

5.5 경상작업을 위한 총합 계절조정계열의 추정 방법

총합계열 Y_t는 구성계열 X_t와 Z_t에 각각 가중치 ω_X와 ω_Z를 부여하여 다음과 같이 결합된다고 가정하자.

$$Y_t = \omega_X \times X_t + \omega_Z \times Z_t \tag{5.11}$$

X-12-ARIMA에서 간접법에 의한 총합계열의 계절조정요인과 사전조정요인은 다음과 같은 방법에 의해서 계산된다.

먼저 구성계열의 사전조정계열(*.B1)과 계절조정계열(*.D11)을 산출하고, 총합계열의 사전조정계열과 사전조정계열은 다음과 같이 구한다.

총합 사전조정계열: $Y_t^* = \omega_X \times X_t^* + \omega_Z \times Z_t^*$ (5.12)

총합 계절조정계열: $SA(Y_t) = \omega_X \times SA(X_t) + \omega_Z \times SA(Z_t)$ (5.13)

이때, 총합계열의 사전조정요인 $\widehat{PR}(Y_t)$과 계절조정요인 $\widehat{SF}(Y_t)$은 다음과 같이 추정할 수 있다.

$$\widehat{PR}(Y_t) \ = Y_t / \ Y_t^*, \qquad \widehat{SF}(Y_t) = Y_t^* / \ SA(Y_t) \tag{5.14}$$

새로운 원계열 Y_{t+1}가 주어졌을 때, 경상적으로 총합계열의 계절조정계열을 추정하기 위한 방법은 2가지 방법을 생각해 볼 수 있다; 각 구성계열의 계절조정계열을 합하는 방법과 총합 원계열에서 총합계열의 계절조정요인과 사전조정요인을 제거하는 방법이다.

각 구성계열의 계절조정계열을 합하여 총합 계절조정계열을 구하는 방법은 먼저 각각의 구성계열에 대한 계절조정계열을 추정한 후, 구성계열의 계절조정계열을 식 (5.13)과 같이 결합한다.

총합 원계열에서 총합계열의 계절조정요인($\widehat{SF}$)과 사전조정요인($\widehat{PR}$)을 제거하여 총합 계절조정계열($\widehat{SA}$)을 구하는 방법은 다음과 같다. 먼저, 식 (5.14)와 같이 총합계열의 사전조정요인과 계절조정요인을 추정한 후, 총합계열에서 식 (5.15)와 같이 총합계열의 사전조정요인과 계절조정요인을 제거한다.

$$\widehat{SA}(Y_{t+1}) \ = \ \frac{Y_{t+1}}{\widehat{PR}(Y_{t+1}) \times \widehat{SF}(Y_{t+1})} \tag{5.15}$$

X-12-ARIMA에서는 간접법 계절조정에 의한 계절요인(SF), 사전조정요인(PR) 및 계절조정계열을 다음과 같이 출력(OUTPUT)과 파일형태로 저장(SAVE)한다. 여기서 D16과 *.IAF는 계절요인(SF)와 사전조정요인(PR)이 결합된 조정요인이다.

	계 절 요 인(SF)	사전조정요인(PR)	계절조정계열(SA)
출 력	D16(Indirect Combined adj. factors)		D11
	D10	D18	
저 장	*.IAF[2]		*.ISA
	*.ISF[1]	*.ICA[2]	

주 1) COMPOSITE 명령에서 SAVE=(ISF) 지정
2) 실행 시, Graphics를 지정하면 자동으로 GRAPHICS 폴더에 저장

(예 5.6) <표 5.9>는 계절조정 시, RegARIMA 모형에 명절효과와 요일효

과 등을 고려한 내구재(DG), 준내구재(SDG), 비내구재(NDG)에 대한 계절조정계열(*.D11)이며, 총합계열인 소비재(CG)는 X-12-ARIMA에서 간접법에 의해 제공되는 계절조정계열(SAVE은 *.ISA, OUTPUT은 indirect SA D11)이다. 계산 합은 다음과 같이 내구재, 준내구재, 비내구재 계절조정계열에 가중치를 부여한 후, 합한 결과이다.

$$CG = 0.2543 \times DG + 0.2403 \times SDG + 0.5054 \times NDG$$

계산 합의 결과는 X-12-ARIMA에서 제공된 소비재 계절조정계열(*.isa)과 일치하고 있음을 알 수 있다.

<표 5.9> 소비재 구성계열 및 총합 계절조정계열: 합

	내구재 (0.2543)[주]	준내구재 (0.2403)	비내구재 (0.5054)	소비재 (*.isa)	계산 합
'04.1월	106.1	114.6	118.6	114.5	114.5
2월	106.7	116.7	119.4	115.5	115.5
3월	104.8	111.2	118.7	113.4	113.4
4월	106.4	115.7	117.3	114.1	114.1
5월	102.8	116.3	116.7	113.1	113.1
6월	107.6	117.7	118.7	115.6	115.6
7월	109.9	115.8	116.7	114.7	114.7
8월	109.5	111.0	115.1	112.7	112.7
9월	109.6	116.2	118.1	115.5	115.5
10월	111.9	117.0	114.4	114.4	114.4
11월	112.0	117.4	114.7	114.7	114.7
12월	115.7	119.1	116.5	116.9	116.9

주) ( )는 가중치

다음은 X-12-ARIMA에서 제공되는 간접법에 의한 총합계열의 계절요인과 사전조정요인인 "*.isf"와 "*.ica" 계열을 이용하여 총합계열인 소비재의 계절조정계열을 계산한 결과이다.

<표 5.10>의 6번째 열(Y/SF*PR)은 원계열에서 계절요인과 사전조정요인을 제거하여 총합계열의 계절조정계열을 구한 결과이다. 계산 결과, X-12-ARIMA에서 제공된 계절조정계열(*.isa)과 일치함을 볼 수 있다.

<표 5.10> 계절 및 사전조정요인에 의한 총합 계절조정계열

	소비재 (Y)	SF 계열 (*.isf)	PR 계열 (*.ica)	SF*PR	$\frac{Y}{SF*PR}$	SA계열 (*.isa)
'04.1월	118.9	1.0190	1.0194	1.0389	114.5	114.5
2월	106.2	0.9299	0.9880	0.9188	115.5	115.5
3월	113.6	1.0024	0.9998	1.0022	113.4	113.4
4월	113.9	0.9978	0.9999	0.9976	114.1	114.1
5월	115.4	1.0203	1.0004	1.0206	113.1	113.1
6월	112.9	0.9761	1.0000	0.9760	115.6	115.6
7월	112.5	0.9805	1.0000	0.9805	114.7	114.7
8월	105.7	0.9378	1.0000	0.9377	112.7	112.7
9월	118.6	1.0239	1.0030	1.0269	115.5	115.5
10월	117.4	1.0285	0.9976	1.0260	114.4	114.4
11월	117.9	1.0287	0.9998	1.0285	114.7	114.7
12월	123.8	1.0594	0.9998	1.0592	116.9	116.9

<참고 5.1> 간접법에 의한 상품수지 계절조정 프로그램

EXPERT.SPC, IMPORT.SPC는 수출과 수입을 계절조정하기 위한 spc 화일이며, COMP_GOODB.SPC는 수출과 수입차이에 의해서 상품수지를 계절조정하기 위한 spc 화일이다. 이때 실행은 GOODB.MTA를 click하면 된다.

프로그램의 X11 명령은 수출, 수입, 상품수지 배별 시계열에 대한 계절조정 결과에 적용된다. 따라서 COMP_GOODB.SPC에 적용된 X11 명령은 상품수지를 직접법에 의해서 계절조정된 결과를 제공해 주며, 간접법에 의한 결과는 COMPOSITE 명령에 의해서 생성된다.

filename: EXPERT.SPC	IMPORT.SPC
series{	series{
Title=" EXPERT"	Title=" IMPORT "
start=1995.01	start=1995.01
period=12	period=12
decimals=1	decimals=1
COMPTYPE=add	COMPTYPE=sub
COMPWT=1.0	COMPWT=1.0
span=(1995.01, 2005.12)	span=(1995.01, 2005.12)
file="c:₩comp_sa05₩expert.txt"	file="c:₩comp_sa05₩import.txt"
format="tramo" }	format="tramo" }
transform{function=log}	transform{function=log}
automdl { }	automdl { }
outlier { types=all }	outlier { types=all }
x11{ force=totals	x11{ force=totals
save=(d10 d11 saa) }	save=(d10 d11 saa) }

filename: COMP_GOODB.SPC	GOODB.MTA
composite { TITLE=" GOODS BALANCE"	expert
name="comp_good_b"	import
decimals=1	comp_goodb
save=(isf isa)	
modelspan=(1995.jan,)	
spectrumstart=1985.jan }	
transform{ function=none}	
x11{ save=(d10 d11)	
force=totals }	

PART B:

X-12-ARIMA 프로그램의 사용법

ARIMA

내 용

RegARIMA 모형의 ARIMA 부분을 설정한다. regression 명령이 없으면 순수한 ARIMA 모형을 설정한다. 모형의 ARIMA 부분은 승법계절요인과 결측 시차를 포함할 수 있다. 반복추정을 하기 위해 ar, ma 옵션을 이용하여 AR과 MA 모수의 초기치를 개별적으로 지정할 수 있다. 또한 일부 모수 값은 초기 값으로 고정하고서 나머지 모수 값을 추정할 수 있다.

일반형식

arima { model = ([2 3] 1 1)(0 1 1)12
title = "ARIMA Model"
ar = (0.3f, -0.14)
ma = (-0.7, 0.85f)
}

명 령

ar 비계절·계절형의 자기회귀 모수에 대한 초기 값을 지정한다. ar 옵션을 사용할 때에는 모형에 있는 모든 AR 모수를 지정하거나 결측치로 지정해야 한다. 결측치의 경우는 0.1로 간주한다. 예를 들어 두 개의 AR모수가 있다면, ar=(0.7,) 로 표시하며, 이는 (0.7, 0.1)과 같다. 세 개의 AR 모수의 경우 ar=(0.8, ,-0.4) 또는 ar=(0.8, 0.1, -0.4)로 써야 한다. 만일 첫 번째 모수의 초기 값을 고정시키고 나머지 모수를 추정하고자 하면 다음과 같이 "f"를 사용하여 ar=(0.7f, 0.1)로 써야 한다.

ma 이동평균모형의 모수에 대한 사항으로 ar 옵션과 같다.

model 모형의 ARIMA 부분을 설정한다. 비계절 ARIMA 모형의 경우는 (p d q)로 기술하며, 이때 p는 자기회귀모형의 AR(p), q는 이동평균모형의 MA(q), d는 차분을 나타낸다. 계절 ARIMA 모형이 있는 경우는 (p d q)(P D Q)로 기술하며, 이때 P는 계절 자기회귀모형의 AR(P), Q는 계절 이동평균모형의 MA(Q), D는 계절차분이다. 두 개 이상의 ARIMA 요인을 설정할 수 있으며, ARIMA 요인은 default(series 명령의 period나 start 옵션에 따른다)와는 다른 계절주기를 줄 수 있다. ARIMA 요인인 (p d q)는 공백과 콤마 로 구별할 수 있으며, (0 1 1)과 (0, 1, 1)은 같다. 결측 시차는 model=([2 3] 0 0)으로 기술할 수 있으며, 이때 모형은 $(1-\phi_2 B^2-\phi_3 B^3)Z_t=a_t$ 이다.

title ARIMA 모형 분석에 대한 제목이다.

print & save ARIMA 명령에서 이용가능한 결과표는 없다.

기타명령

file 선행의 X-12-ARIMA 실행결과로부터 저장된 선정한 모형 내용을 포함한 파일 이름이다. 이 화일은 estimate 명령에서 model 표를 저장함으로써 만들 수 있다. file 명령을 사용하면 model, ma, ar 명령을 쓸 수 없다.

fix file명령의 지정된 모형에서 발견된 계수의 추정 값을 초기화하여 사용하지 않고 고정시킬 것인지를 지정한다.
fix=both는 default로 회귀모수와 ARIMA 추정모수를 지정된 값으로 고정하고, fix=arma이면 ARIMA 추정모수만 지정된 값으로 고정하며, fix=none이면 어떤 추정모수도 고정하지 않는다. 또한, 이 명령은 estimate 명령에 지정된 parms을 무시한다.

참　　고

arima 명령은 pickmdl, automdl 명령과 함께 사용할 수 없다. arima 명령의 ar 옵션은 estimate 명령에서 file 옵션을 사용한 경우에는 사용할 수 없다.

모형의 인수는 ARIMA 모형의 요인(factor) 수만큼 쓸 수 있으나, 모형의 AR과 MA모수 및 차분차수는 총 25개를 넘을 수 없다. AR과 MA 모수의 최대 시차는 24이며, ARIMA 요인의 최대차분은 3이다.

일반적으로 ARIMA 모형의 요인은 계절주기가 S인 경우는(p d q)s로 표시한다. 따라서 $(0\ 1\ 1)_6$은 $(1-B^6)$에 의해 차분되며, 모형은 $(1-\Theta_6 B^6)MA$가 된다. 계절주기가 없는 첫 번째 ARIMA 요인은 비계절형임을 나타내며, 두 번째 ARIMA 요인은 series 명령의 계절주기를 갖는 계절형 모형임을 나타낸다. 예를 들어, series 명령에서 period=12라고 하였다면(또는 start=year.month) model=(0 1 1)(0 1 1)과 model=$(0\ 1\ 1)(0\ 1\ 1)_{12}$이다. 만일 ARIMA 요인을 추가한다면, 계절주기를 명확히 기술하지 않는 한 비계절형 모형으로 간주한다.

단위근을 갖지 않는(with roots inside the unit circle: 정상계열) MA 다항식을 만드는 경우 사용자는 MA 모수의 초기 값을 지정하지 말아야 한다. 이때 초기 값을 지정하면, 프로그램은 실행을 멈추고 '초기 값을 재지정하라'는 에러 메시지를 출력한 후, 프로그램으로 다시 돌아간다. 단위근(on the unit circle: 비정상계열)을 갖는 MA 다항식을 만드는 초기 값은 계열이 추정되지 않는 비가역성(non-reversibility) 다항식인 경우에만 지정할 수 있다. 즉, 모수가 추정을 하지 않거나 다항식이 있는 모수를 추정하는 동안 고정된 것으로 지정하려면 초기 값을 지정할 수 있다. 예를 들어 단지 MA모수만을 첫 번째 계절 MA모수로 가진 모형인 경우, ma=(1.0f)는 항상 지정할 수 있으며, 추정을 하지 않는 경우라면 ma=(1.0)을 지정할 수 있고, ma=(1.1)은 지정할 수 없다.

AR 다항식을 우도함수에 사용하면(estimate 명령문의 exact=arma), 사용자는 비정상(nonstationary) AR 다항식을 만드는 AR모수에 대한 초기 값을 지정하지 말아야 한다.

예 제

(예 1) 1차 차분을 갖고 MA모수의 시차가 1인 비계절형 ARIMA 모형인 $(1-B)y_t=(1-\theta B)a_t$를 추정하는 경우를 가정하자.

```
series {title = "Quarterly Grope Harvest" start =1950.1
        period = 4
        data =( ·············· )}
arima {model = (0 1 1)}
estimate {          }
```

(예 2) 월별 시계열 y_t를 log 변환한 계절형 ARIMA 모형 $(1-\phi_1 B-\phi_2 B^2)(1-B)(1-B^{12})y_t=(1-\Theta_{12}B^{12})a_t$의 (2 1 0)(0 1 1)인 모형 추정을 가정하자.

```
series {title = "Monthly sales"     start =1976.Jan
        data =(138 128 . . . 297) }
transform {function = log}
arima {model = (2 1 0) (0 1 1)}
estimate {          }
```

(예 3) 고정된 계절효과, 추세상수와 회귀잔차에 대한 ARIMA(0,1,1) 모형을 갖는 RegARIMA 모형의 추정을 가정하자. 이때 모형은 $(1-B)(y_{t-}\sum\beta_i M-c\cdot t)=(1-\theta B)a_t$, 여기서 M는 고정된 계절회귀효과 변수이다.

```
series {title = "Monthly sales"     start =1976.Jan
        data =(138 128 . . . 297) }
transform {function = log}
regression {variables = (seasonal const)}
arima {model = (0 1 1) }
estimate {          }
```

(예 4) 1차 차분한 시계열이 AR(2) 모형이나 1차 시차의 AR 모수가 결측된 모형인 $(1-\phi_2 B^2)(1-B)y_t=a_t$의 추정을 가정하자.

```
series {title = "Annual olive Harvest"     start =1950
        data = (251   271   . . . 240)}
arima {model = ([2] 1 0)}
estimate {          }
```

(예 5) 1차 계절차분하고 상수추세와 회귀오차 z_t를 가진 ARIMA 모형 $(1-B^{12})z_t = (1-\Theta B^{12})a_t$을 추정하자.

```
series {title = "Monthly sales"     start =1976.Jan
        data =(138 128 . . . 297) }
transform {function = log}
regression {variables = const  }
arima {model = (0 1 1)12  }
estimate {          }
```

(예 6) 회귀오차 Z_t의 모형식이 3개의 ARIMA 요인을 가지고 있는 다음과 같은 모형이라 하자.

$$(1-\phi_1 B)(1-\phi_3 B^3)(1-B)Z_t = (1-\Theta B^{12})a_t$$

여기서 제3의 ARIMA 요인인 $(1-\phi_3 B^3)$는 3개월을 묶어 분기를 나타내는 분기 자기상관이다. 모형설정 시, 분기요인의 주기는 반드시 주어져야 한다

```
series {title = "Monthly sales"     start =1976.Jan
        data =(138 128 . . . 297) }
regression {variables = (const seasonal) }
arima {model = (1 1 0)(1 0 0)3(0 0 1) }
estimate {          }
```

(예 7) 회귀오차 Z_t의 모형 (0,1,1)(0,1,1)인

$(1-B)(1-B^{12})Z_t = (1-\theta B)(1-1.0\ B^{12})a_t$

로 계절 MA의 모수가 1.0으로 고정된 경우를 가정하자.

```
series {title = "Monthly sales"      start =1976.Jan
          data =(138 128 . . . 297) }
transform {function = log}
arima {model = (1 1 0) (1 1 0)12
         ma = (   ,1.0f)}
estimate {          }
```

AUTOMDL

내 용

RegARIMA의 ARIMA 부분은 TRAMO(Gomez와 Maravall, 2000)에 사용된 자동 모형 선정방볍에 의해서 추정된다. 사용자는 모형에 사용하기 위한 ARMA 모형의 최대 차수와 차분을 지정할 수 있으며, 기타 여러 가지 선정 기준을 이용할 수 있다.

일반형식

```
automdl { maxorder = (3 1)
          maxdiff = (1 1) or diff = (1 0)
          acceptdefault = no
          checkmu = yes
          ljungboxlimit = 0.99
          print = (none bestfivemdl autochoice)
          savelog = automodel
        }
```

명 령

acceptdefault 모형의 잔차에 대한 Ljung-Box Q 통계량이 모형을 채택할 수 있는 수준이라면, default 모형을 선정할 수 있는지 검토한다(acceptdefault=yes). 이때 월별자료는 24시차, 분기자료는 16시차를 이용한다. default 모형을 채택할 수 있다면, 더 이상 모형의 식별이나 차분을 하지 않는다. default는 acceptdefault=yes이다.

checkmu 자동 모형 선정과정에서 상수항에 대한 유의수준을 검토할 것인지(checkmu=yes) 혹은 regression spc에서 사용자가 지정할 것인지(checkmu=no)를 지정한다. default는 checkmu=yes이다.

diff 자동 ARIMA 모형 식별에 사용되는 일반 차분 차수와 계절 차분 차수를 지정한다. 두 값을 반드시 지정해야 하며, default 값은 없다. 일반 차분 차수는 0, 1 그리고 2, 계절 차분 차수는 0과 1을 사용한다. 같은 spc 명령에 maxdiff를 사용하면, diff 옵션은 무시되며, maxdiff에 지정된 일반 차분과 계절 차분으로 자동 식별이 된다.

ljungboxlimit Ljung-Box Q 통계량의 신뢰계수에 대한 채택 기준치를 지정한다. 최종 모형의 잔차에 대한 Ljung-Box Q 통계량이 ljungboxlimit 보다 크다면 모형은 기각된다. 이때 특이치 임계치가 축소되고, 모형과 특이치 식별은 축소된 임계치(reducecv 옵션)에 의해서 재실시된다. default는 ljungboxlimit=0.95이다.

maxdiff 자동 차분 차수 결정을 위한 일반 및 계절 차분의 최대 차수를 지정한다. 일반 차분의 사용할 수 있는 최대 차수는 1과 2이며, 계절 차분의 차수는 1이다. default는 maxdiff=(2 1)이다.

maxorder 자동 ARIMA 모형 식별 중에 검토하게 될 일반 및 계절 ARMA 다항식의 최대 차수를 지정한다. 일반 ARMA 모형의 최대 차수는 0보다 커야 하며 4보다는 작다. 한편, 계절 ARMA 모형의 최대 차수는 1혹은 2이다. default는 maxorder=(2 1)이다.

mixed 자동 모형 식별 과정에서 일반 AR 및 MA 항 또는 계절 AR 및 MA 항을 갖는 ARIMA 모형을 고려할 것인지 지정한다(mixed=yes). mixed=no이라면, 혼합모형은 고려되지 않는다. 예를 들어, mixed=no일때 ARIMA (0 1 1)(1 1 0) 모형은 고려되나, ARIMA (1 1 1)(0 1 1) 모형은 비계절 부분의 MA 항이 혼합되어 있기 때문에 고려되지 않는다. default는 mixed=yes이다.

print AUTOMDL에서 이용할 수 있는 표는 다음과 같으며 save 옵션은 이용할 수 없다. default로 header, autochoice와 unitroottest 표가 인쇄된다.

<AUTOMDL: 이용가능한 출력표>

이 름	단축	내 용
autochoice	ach	자동 모형 선정 절차에서 선정된 모형
autochoicemdl	amd	ARIMA 모형 차수 선정 중 고려된 모형의 요약 결과
header	hdr	자동 모형화의 결과 header
unitroottest	urt	차분 선정
unitroottestmdl	urm	차분 차수 식별 중 추정된 모형의 요약 결과
bestfivemdl	b5m	ARMA 모형 차수 선정 중 고려된 5개 최적 모형의 요약

savelog log 파일에 저장되는 진단 통계량을 지정하며, 통계량은 다음과 같다.

이 름	단축	내 용
autodiff	adf	자동 모형 식별 과정의 차분 선정
automodel	amd	자동 모형 식별 과정의 ARIMA 모형 선정
bestfivemdl	b5m	ARMA 모형 차수 선정 중 고려된 5개 최적 모형의 요약

기타 명령

armalimit 모형의 최종 검정을 위해 추정된 ARMA 모형의 계수에 대한 t-통계량의 기준치를 지정한다. 최대 차수의 ARMA 모형의 계수가 이 값보다 작은 t-값을 갖는다면, ARMA모형의 차수를 줄인다. armalimit의 주어진 값은 상수항에 대한 최종 검정을 위해 사용된다. 상수항의 t-값이 armalimit보다 작다면, 회귀변수에서 상수항을 제외한다. armalimit 값은 0보다 커야 하며, default는 armalimit=1.0이다.

urfinal 최종 단위근 검정을 위한 기준치이다. 최종 모형의 AR 근의 크기가 이 값보다 작다면, 단위근을 가정하며 AR 다항식 차수를 1씩 줄이고 차분의 차수를 늘린다. urfinal 값은 1보다 커야 하며, default는 urfinal=1.05이다.

reducecv 식별된 모형이 채택할 수 없는 Ljung-Box Q 통계량을 갖고 있을 때, 특이치의 임계치를 줄이는 비율을 지정한다. 이 값은 0과 1사이의 값을 취하며, 자동 특이치 식별을 하는 경우에만 사용할 수 있다. 축소된 임계치는 $(1-reducecv)\times CV$로 계산되며, CV는 원래의 임계치이다. default는 reducecv=0.14268이다.

참 고

automdl 명령은 pickmdl, arima 명령이나 estimate에서 file 변수를 사용하는 경우 함께 사용할 수 없다.

Version 0.3에 적용된 ARIMA 모형 자동 선정 방법은 Gomez와 Maravall(1997)에 의해서 개발된 TRAMO의 계절조정 프로그램에 채용하고 있는 Hannen-Rissanen의 반복 계산법(Recursive calculation)에 의한 ARIMA 모형 자동 선정 방법을 기초로 하고 있다. ARIMA 모형 자동 선정 방법은 TRAMO에서의 절차와 비슷하나, X-12-ARIMA는 모형추정을 위한 프로그램, RegARIMA 옵션, 변수변환, 특이치 식별과 모형 진단방법 등이 TRAMO 방법과는 다소 다르다. X-12-ARIMA/0.3의 ARIMA 모형 선정절차는 다음의 5단계에 의해서 이루어 진다.

- default 모형 추정: default 모형 추정, 초기 특이치 식별과 회귀계수 검정, 잔차 진단을 실시
- 차분 차수 식별: 모형에 필요한 차분의 차수를 결정하기 위한 단위근 검정(Unit root test) 실시
- ARIMA 모형 식별: 반복절차(Iterative procedure)에 의해 ARIMA 모수를 결정
- default 모형과 식별된 모형 비교: 식별된 모형과 default 모형 비교
- 최종모형 검토: 최종모형 적절성을 검토

사용자는 두 번째 단계인 차분 차수(differencing orders)의 식별단계는 선택적으로 사용할 수 있다. 즉, 사용자는 diff 변수를 이용하여 차분 차수를 지정할 수 있다.

(1) default 모형 추정

특이치를 자동 선정하는 첫 번째 단계는 default 모형을 추정하는 것이다. default 모형은 일명 “airline”모형인 ARIMA(0 1 1)(0 1 1)s이다.

default 모형은 사용자가 regression 명령에서 요일효과, 명절효과 등을 검정하고자 하는 경우 이들 효과의 유의성에 대한 초기 검토를 하기 위해 사용된다. regression의 aictest 옵션은 AICC로 알려진 AIC 최소 통계량에 의해 요일효과와 명절효과에 대한 회귀변수의 유의성을 검증하는데 이용된다.

특이치가 식별되면 요일효과, 명절효과 및 상수항이 유의적인지 검토한다. 이때 검정은 1.96의 임계치를 갖는 t-검정을 실시한다. default 모형의 회귀모형이 결정된 후에는 선정된 모형의 잔차에 대한 다음 사항을 진단한다.

· 잔차에 대한 Ljung-Box Q 통계량(월별자료는 24시차, 분기자료는 16시차)
· Ljung-Box Q 통계량의 신뢰계수
· RegARIMA 모형에 대한 잔차 평균의 t-값
· 잔차의 표준오차

(2) 차분 차수의 식별

“선형화된 계열(원계열)”에 대해서 적절한 차분 차수를 식별한다. 차분 차수의 식별은 먼저 선형화된 계열을 차분하고 ARMA 모형의 적합과 단위근 검정을 실시함으로써 이루어진다. 차분한 ARMA 모형은 Hannen-Rissanen의 반복계산법(Hannan과 Rissanen 1982, Gomez와 Maravall 2000)에 의해서 추정된다. 차분 차수는 다음 단계에 의해 추정한다.

STEP 1: Hannen-Rissanen 방법을 이용하여 선형화된 계열에 ARIMA (2 0 0)(1 0 0)s 모형을 적합하고, 적합된 모형의 AR 근을 추정한다. 근의 값이 1.042보다 작다면, 단위근을 갖는다고 볼 수 있으며, (일반 또는 계절) 근에 따라 차분 차수를 1씩 증가시킨다.
Hannen-Rissanen 과정에서 단위원 내의 근을 갖는 모형을 추정하였다면, X-12-ARIMA는 정확한 MLE 방법에 의해서 ARIMA 모형의 근을 재추정하며, 앞에서 설명한 검정과정을 재실시한다.

STEP 2: STEP 1에서 차분 차수가 결정되었다면, STEP 2에서 처음부터 선형화된 계열을 차분한다. 차분된 계열에 ARMA (1 1)(1 1)s 모형을 적합하며, AR 모수가 1에 가까운지를 검토한다. AR 모형에서 추정된 계수가 1에 가깝다면, AR과 MA 다항식의 공통요인을 제거할 수 있는지 검토한다. 제거할 수 없다면, 선형화된 계열은 새로운 차분 차수에 의해 차분을 한다. ARMA 모형이 다시 적합되며, 차분을 더 할 수 있는지 검토한다. 이 과정은 더 이상 차분을 할 수 없을 때 까지 반복한다. 차분 차수가 결정되면, 차분된 계열의 평균 항에 대한 t-통계량이 표본 평균(차분이 식별되지 않는 경우)이나 RegARIMA에 상수항을 포함한 모형에 의해서 계산된다. 이 때 검정의 임계치는 계열의 관측치에 의존한다.

(3) ARMA 모형의 차수 식별

적절한 차분 차수가 결정되면, ARMA 모형의 모수에 대한 차수를 추정하게 된다. X-12-ARIMA 출력결과에는 다음과 같은 BIC(Bayesian Information Criterion) 통계량이 주어진다. 여기서 N은 모형의 일반 및 계절 차분을 실시한 후의 관측치 수이며, $\widehat{L_N}$은 N개의 관측치에 의해 계산된 log 우도함수의 최대값, n_p는 백색잡음 분산을 포함한 모형에서 추정된 모수의 수이다.

$$BIC_N = -2\widehat{L_N} + n_p \log N$$

X-12-ARIMA에서는 자동 모형 선정 절차에 사용된 최소 BIC2 통계량은 다음과 같다.

$$BIC2_N = (-2\widehat{L_N} + n_p \log N)/N$$

ARIMA 모형을 자동 선정하기 위하여, 먼저 계절 ARIMA 모형에 대한 초기모형을 결정한다. 초기 모형을 결정하기 위하여 ($0 \le P$, $Q \le m_s$)에 대해서 (3 d 0)(P D Q)s의 모든 ARIMA 모형에 대한 BIC2를 계산한다. 여기서 m_s는 계절 AR 및 MA의 최대 차수이며, default는 1, 최대는 2까지 사용할 수 있다. 이때 BIC2가 최소가 되는 P와 Q를 선택한다.

추정된 P와 Q를 이용하여, 비계절 ARIMA 모형에 대한 최적 모형을 식별한다. ($0 \le p$, $q \le m_r$)대해서 (p d q)(P D Q)s의 모든 ARIMA 모형에 대한 BCI2를 계산한다. 여기서 m_r은 비계절 AR 및 MA의 최대 차수이며,

default는 2, 최대 차수는 3까지 사용할 수 있다. 이때 BIC2가 최소가 되는 p와 q를 선택한다.

추정된 p와 q를 이용하여, 계절모형의 차수를 다시 결정한다. $0 \leq P, Q \leq m_s$에 대해서 (p d q)(P D Q)s의 모든 ARIMA 모형에 대한 BIC2를 계산한다. 이때 BIC2가 최소가 되는 P와 Q를 선택한다.

(4) default 모형과 식별된 모형의 비교

식별된 모형이 default 모형이 아닌 경우, 자동으로 선정된 모형의 잔차 진단 결과와 default 모형의 잔차 진단 결과를 비교한다. Q_A를 자동으로 선정된 모형에 의한 Ljung-Box Q 통계량(월별자료는 24시차, 분기는 16시차)에 대한 신뢰계수, Q_D는 default 모형에 의한 Ljung-Box Q 통계량에 대한 신뢰계수, RSE_A는 자동으로 선정된 모형에 대한 잔차 표준 오차, RSE_D는 default 모형에 대한 잔차 표준 오차이다. 다음과 같은 경우에는 default 모형(D)이 자동으로 선정된 모형(A)보다 선호된다.

· default 모형에서 자동으로 식별된 특이치의 수가 자동으로 선정된 모형에서 식별된 특이치의 수보다 작거나 같은 경우이면서(AND)
· $Q_A<0.95$ 이면서(AND) $Q_D<0.75$ 이고 $RSE_D<RSE_A$, 혹은(OR)
· $Q_A>0.95$ 이면서(AND) $Q_D<0.95$(첫 번째 pass에서만 적용), 혹은(OR)
· $Q_A<0.95$ 이면서(AND) $Q_D<0.75$ 이고
$Q_D<Q_A$ 이면서 $RSE_D<RSE_A*1.013$, 혹은(OR)
· $Q_A \geq 0.95$ 이면서(AND) $Q_D<0.95$ 이고 $RSE_D<RSE_A*1.013$, 혹은
· 자동선정 모형이 (1 0 1)(0 1 1)s 이거나 (1 0 0)(0 1 1)s이고
$\phi_1 \geq 0.82$, 혹은
· 자동선정 모형이 (0 1 1)(1 0 1)s 이거나 (0 1 1)(1 0 0)s이고
$\Phi_1 \geq 0.65$.

선호된 모형을 채택할 수 있는지를 보기 위한 검정통계량은 Ljung-Box Q 통계량의 신뢰계수이다. 이 값이 0.975보다 크다면, 자동 특이치 선정의 임계치 값은 reducecv 값을 기초로 다음과 같이 낮춘다.

$$CV_r = (1 - reducecv) \times CV$$

여기서 CV는 원래의 임계치, CV_r은 낮추어진 특이치의 임계치이다. 낮추

어진 임계치가 2.8이하이면 임계치로 허용하지 않는다. 새로운 임계치에 의해 자동 모형 재식별과 특이치 재식별이 다시 실행된다.

재추정된 모형과 이전에 선호된 모형을 비교한다. 이때 Ljung-Box Q 통계량이 다시 계산되며, 신뢰계수가 0.990보다 크다면 검정은 실시되지 않는다. 채택할 수 있는 Ljung-Box Q 통계량을 갖는 모형이 선정되지 않는다면, 프로그램은 (3 d 1)(0 D 1)s인 ARIMA 모형을 잠정모형으로 선정하고 특이치를 식별한다.

마지막으로 요일효과, 명절 및 상수항에 대한 t-통계량이 검토되며, 이때 t-값의 임계치는 1.96이다. 요일효과의 경우, 적어도 하나의 회귀계수는 1.96보다 큰 임계치를 가져야 한다.

(5) 최종 모형 검토

최종모형의 적정성에 대한 검정을 실시한다. 일반 및 계절 차분 차수가 맞는지를 보기 위하여 AR 다항식의 단위근을 검토한다. 구해진 AR 근이 1.05보다 작거나 같으면 AR 단위근이 있다. 단위근이 존재한다면, AR 다항식의 차수를 줄이고 차분의 차수를 늘린다. 이때, 모형을 재추정하고, 재추정된 모형에 대한 진단을 실시한다.

다음으로 비계절 MA 다항식에 대한 단위근을 검정한다. 비계절 MA 계수의 합이 0.999인지를 검토한다. 합이 0.999이면, 일반 차분 차수를 하나씩 줄여가며, MA 다항식의 차수도 하나씩 줄여간다. 이후 상수항을 RegARIMA 모형에 포함하고 모형을 재추정한다. 그리고 재추정된 모형에 대한 진단을 실시한다.

RegARIMA 모형에 상수항이 포함되지 않았다면, 모형의 잔차 평균에 대한 t-통계량이 유의적인지(즉 2.5보다 큰지) 확인한다. t-통계량이 유의적이면 회귀변수에 상수항을 포함한다.

유의적이지 않은 ARIMA 모수에 대한 검정이 식별된 모형을 단순화하기 위하여 실시된다. ARIMA 계수에 대한 t-통계량이 계산되며, 다음 기준에 의해 가장 높은 일반 및 계절 AR 및 MA 차수에 대한 계수의 유의성 검정이 실시된다.

· 가장 큰 차수의 일반 및 계절 AR과 MA 계수의 t-통계량이 지정한 **armalimit** 값보다 큰 모형은 피한다.
· 추정된 계수의 절대값이 0.15(관측치가 150 이하인 경우), 0.10(관측치가 150 이상인 경우) 보다 커야 한다.

CHECK

내 용

추정된 모형에 대한 잔차 검증을 위한 통계량을 산출하는 명령이다. 잔차 검증을 위한 통계량은 표준오차와 관련된 잔차의 표본 ACF와 PACF, Ljung-Box Q-통계량과 p-value, 잔차에 대한 통계량, 표준화된 잔차의 히스토그램 등이 있다.

일반형식

check { **maxlag** = estimate
print = (none +histogram +acf +pacf)
save = (acf)
savelog = normalitytest
}

명 령

maxlag 잔차 표본 ACF, PACF에 요구되는 시차의 수, 월자료의 default는 36, 분기는 12이다.

savelog 이용할 수 있는 진단 통계량은 다음과 같다.

이 름	단축	내 용
normalitytest	nrm	RegARIMA 모형 잔차에 대한 정규성 검정의 결과 (첨도, 왜도, Geary's a 통계량)
Ljungboxq	lbq	Ljung-Box Q 통계량에 대한 유의적인 시차

print & save 이용할 수 있는 결과표는 다음과 같다. acf, acfplot, histogram, normalitytest 결과표는 default로 주어진다.

<Check: 이용 가능한 출력표>

이 름	단축	저장	내 용
acf	acf	+	각 시차에서 계산된 Ljung-Box Q-통계량과 표준오차를 갖는 잔차의 자기상관함수
acfplot	acp	·	±2 표준오차 한계 내의 잔차자기상관함수의 plot
pacf	pcf	+	표준오차를 갖는 잔차의 편자기상관함수
pacfplot	pcp	·	±2 표준오차 한계 내의 잔차 편자기상관함수의 plot
acfsquared	ac2	+	각 시차에서 계산된 Ljung-Box Q-통계량과 표준오차를 갖는 잔차자승의 자기상관함수
acfsquaredplot	ap2	·	±2 표준오차 한계 내의 잔차자승의 편자기상관함수 plot
normalitytest	nrm	·	모형 잔차의 정규성 검정을 위한 Geary's a와 첨도 검정, 잔차의 왜도 검정
specresidual	spr	+	RegARIMA 모형 잔차의 스펙트럼 그림
histogram	hst	·	표준화된 잔차의 히스토그램, 잔차에 대한 통계량[최소값, 최대값, 중앙값, 표준편차, 잔차 표준편차에 대한 robust추정(1.48×{편차의 절대값에 대한 중앙값})]

참 고

check은 추정된 모형의 잔차에 대한 통계량을 산출하는 명령으로, 기본적으로 estimate 명령이 없으면 check 명령에 의해 모형이 추정된다. 모형이 옳다는 귀무가설 하에서 Ljung-Box Q-통계량은 근사적으로 χ^2-분포가 되는데, 이때 자유도는 (시차- 추정된 AR과 MA 모수의 수)이다. 자유도는 출력 시에 나타나며, 자유도가 '0'인 Q-통계량과 p-값은 무시된다.

X-12-ARIMA는 RegARIMA 모형의 잔차(ϵ_i)에 대한 정규성 검정을 위하여 2가지의 검정통계량을 제공한다. 하나는 다음과 같은 Geary의 a-통계량이다.

$$a = \frac{\frac{1}{n}\sum_{i=1}^{n}\left|\varepsilon_i - \bar{\varepsilon}\right|}{\sqrt{\frac{1}{n}\sum_{i=1}^{n}(\varepsilon_i - \bar{\varepsilon})^2}}$$

다른 하나는 다음과 같은 첨도(kurtosis)를 이용한 통계량이다.

$$b_2 = \frac{m_4}{m_2^2} = \frac{n\sum(\varepsilon_i - \bar{\varepsilon})^4}{(\sum\varepsilon_i - \bar{\varepsilon})^2)^2}$$

이때 b_2가 3보다 크거나 작으면 정규분포에서 벗어났다고 볼 수 있다.

X-12-ARIMA는 1% 유의수준에서 정규성 검정을 실시하며, 이들 "통계량 중 하나가 유의적이다"라면 표준화된 잔차가 표준정규분포를 따르지 않음을 나타낸다. RegARIMA 모형이 잘 적합되면 통상적으로 정규성의 문제를 야기시키지 않는다. 그러나 어떠한 효과가 모형에 잘 반영되지 않을 때 정규성 검정은 유의적으로 나타난다.

예 제

(예 1) 모형의 잔차를 분석할 수 있는 가능한 모든 통계량을 출력한다. 잔차 표본자기상관함수와 편자기상관함수는 lag=36 (default)으로 계산하고, estimate 명령없이 모형을 추정한다

```
series { title ="Monthly   Retail  Sales"   start  =  1964.jan
    file = "sales1.dat"      }
regression { variables = (td ao1967.jun
                          ls1971.jun  easter[14])   }
arima { model = (0 1 1)(0 1 1) }
check { print = (all) }
```

(예 2) (예 1)과 동일한 시계열과 모형으로 잔차에 대한 모든 통계량을 산출한다. 이때, 잔차 ACF는 lag=24로 구하고, 잔차 PACF의 plot과 table은 구하지 않는다.

```
series { title ="Monthly  Retail Sales"  start = 1964.jan
         file = "sales1.dat"      }
regression { variables = (td ao1967.jun
                          ls1971.jun  easter[14])   }
arima { model = (0 1 1)(0 1 1) }
check { print = (all  -PACF  -PACFplot)  maxlag = 24  }
```

COMPOSITE

내 용

이 명령은 총합계열의 계절조정계열을 직접법과 간접법으로 얻기 위한 것이다. 총합 계절조정계열을 구하기 위하여, metafile 내의 화일들은 총합계열의 구성계열로 정의되어야 하며, 총합계열이 어떻게 결합되어 있는가를 지정하여야 한다(각 구성계열의 series 명령에서 comptype와 compwt 옵션에 의해서 알 수 있다). comptype와 compwt 옵션은 각 series 명령에서 사용한다.

일반형식

composite { title = "Total one family housing starts"
name = "hs1ft"
decimals = 2
modelspan = (1985.jan,)
print = (brief +aggtest)
save = (indseasonal)
savelog = (indtest)
}

명 령

decimals 계절조정 결과를 출력할 때, 소수점 자릿수를 지정한다. 값은 0과 5사이의 정수이며, default는 0이다.

diffspectrum diffspectrum=no이면 원계열이나 계절조정계열에 일차 차분이 적용되지 않은 계열에 의해 스펙트럼이 생성된다. diffspectrum=yes이면 일차 차분이 적용된다.

name 총합계열 이름으로 8자 이내로 따옴표 안에 기술하며, 매 페이지마다 출력된다.

modelspan RegARIMA 모형의 계수를 결정하기 위해 사용되는 총합계열의 분석기간을 지정한다. 이 옵션은 최근의 시계열이 사전조정을 위한 회귀모형이나 과거의 시계열이 전방 예측에 영향을 미치지 않도록 하기 위하여 유용하게 사용할 수 있다. modelspan 옵션은 분석하고자 하는 기간의 시작과 끝이 있어야 한다. 월별 자료의 경우, RegARIMA 모형을 1968년 1월부터 분석하는 시계열의 마지막 기간까지 이용하고자 한다면, modelspan=(1968.1,) 으로 기술하며, 여기서 콤마(,)는 반드시 있어야 한다. 모형의 분석기간은 series 명령에서 기술한 기간 내에 있어야 하며, 시작 월(분기)은 마지막 월(분기)보다 선행해야 한다.

saveprecision save 옵션으로 결과표를 각각의 파일로 저장할 때, 저장하고자 하는 소숫점 자리수이다. saveprecision의 default는 15이다. 예) saveprecision=10

spectrumtype 스펙트럼 그림에서 사용된 스펙트럼 추정치의 형태를 지정한다. spectrumtype=periodogram이면, 시계열이 주기표가 계산되고 그림이 그려진다. default는 시계열의 AR 모형의 스펙트럼이 생성된다(spectrumtype=arspec).

startspectrum 총합계열, 직접 및 간접법에 의한 계절조정계열, 수정된 불규칙 계열에 대한 스펙트럼 분석을 위해 사용되는 시계열의 시작 시점을 기술한다. startspectrum=*year.seasonal period*로 표현한다. 월자료의 스펙트럼을 위한 default 시작시점은 시계열이 8년 이상이면 최근 8년 동안의 시계열이다. 분기의 default 시작 시점은 총합계열의 시작 시점이다.
예) startspectrum=1987.Jan.

title 총합계열의 계절조정에 대한 제목으로 79자 이내로 따옴표 안에 기술한다.

print & save 선택 가능한 출력표는 다음과 같다.

<Composite: 이용 가능한 출력표>

이 름	단축	저장	내 용
adjcompositesrs	b1	+	총합 사전조정계열
compositers	cms	+	
compositeplot	cmp	+	사전조정된 총합계열의 plot
header	hdr	+	간접 계절조정계열의 header
indajustfac	iaf	+	간접계절조정계열의 최종 혼합 조정요인
indaoutlier	iao	+	간접계절조정계열의 AO 특이치
indcalendar	ica	+	간접계절조정계열의 캘런더 요인
indlevelshift	ils	+	간접계절조정계열의 수준변화 요인
indtest	itt	·	총합계열의 계절조정 적합성 test
indunmodsi	id8	+	간접 계절조정계열의 수정하지 않은 최종 si-비율 (차분)
indftestd 8	idf	·	간접 계절조정계열의 안정성과 이동계절성을 위한 F-test
indreplacsi	id9	+	간접 계절조정계열의 si-ratios(차분) 극한값의 최종 대체값
indmovseasrat	ims	·	간접 계절조정계열의 이동계절성 비율
indseasadj	isa	+	최종 간접 계절조정계열
indseasonal	isf	+	간접 계절조정계열의 최종 계절요인
indseasonalplot	isp	·	월(분기)별 간접 계절요인 plots
indadjsatot	iaa	+	최종 간접 계절조정계열로 연간 합이 원계열의 연간 합과 일치
indresidualseasf	irf	·	잔차 계절성의 F-test
indseasadjplot	iap	·	최종 간접 계절조정계열의 plot
indtrend	itn	+	간접 계절조정계열의 최종 추세-순환
indtrendplot	itp	·	간접 계절조정계열의 최종 추세-순환 plot
indirregular	iir	+	간접 계절조정계열의 최종 불규칙요인
indirregularplot	iip	·	간접 계절조정의 최종 불규칙요인의 plot
indmodoriginal	ie1	+	간접 계절조정계열의 극한값을 수정한 원계열
indmodsadj	ie2	+	간접 계절조정계열의 극한값을 수정한 계절조정계열
indmodirr	ie3	+	간접 계절조정계열의 극한값을 수정한 불규칙요인
indmcdmovavg	if1	+	최종 간접 계절조정계열의 MCD 이동평균
indyrtotals	ie4	·	간접 계절조정계열과 원계열에 대한 연간합 비율
indsachange	ie6	·	간접 계절조정계열의 퍼센트 변동(차분)

<Composite: 이용 가능한 출력표(계속)>

이 름	단축	저장	내 용
indtrendchanges	ie7	+	간접 계절조정계열의 최종 추세요인 펜센트 변화(차분)
indrobustsa	iee	+	극단값이 수정된 최종 간접 계절조정계열
indx11diag	if2	+	간접 계절조정계열의 계절조정 진단 요약
indqstat	if3	·	간접 계절조정계열의 Q-통계량
indrevsachanges	i6a	+	수정한 년간 합을 가진 간접 계절조정계열의 퍼센트 변동
origchanges	ie5	+	원계열의 퍼센트 변화(차분)
origwindsaplot	ie0	·	간접 계절조정에 의한 총합계열 plot
ratioplotindsa	ir2	·	원계열의 전월비(전분기) plot
ratioplotorig	ir1	·	간접 계절조정계열의 전월비(전분기) plot

참　　고

COMPOSITE 명령을 갖는 입력 spc 화일은 총합계열을 이루고 있는 구성계열에 대한 spc 화일과 연계해서 사용하여야 한다. 구성계열의 spc 화일 이름은 metafile 내에 있어야 하며, 총합계열은 마지막에 기술한다. sries 명령의 comptype 옵션은 총합계열이 구성계열을 어떻게 결합하였는지를 나타낸다.

총합계열의 조정은 총합계열을 직접 계절조정하거나 구성계열들을 계절조정한 후, 이들을 합하는 간접 계절조정방법이 있다. 간접 계절조정은 각 구성계열을 계절조정하거나 spc 화일에 따라 계절조정을 하지 않은 계열을 comptype에 의해 합성을 한다. 직접 계절조정은 총합계열의 spc 화일에 의해서 실시한다. 직접 계절조정 결과의 출력은 x11 명령의 print나 save 옵션을 사용한다.

직접 또는 간접 계절조정의 이동기간분석(sliding span analysis)을 하고자 한다면, 이동기간분석 옵션을 각 구성계열의 계절조정 옵션에 지정해야 한다. 만약, 계절 filter 길이가 각 구성계열마다 서로 다르다면, 이를 같도록 하기 위해 사용자는 slidingspans 명령의 length 옵션을 사용해야 한다.

계절조정계열의 수정분석(revision analysis)을 총합계열에 대해서 지정

했을 때, 총합계열의 직접 혹은 간접 계절조정의 수정분석이 실시된다. 수정분석은 각 구성계열에 대해서 지정해야 한다.

차분된 계절조정계열과 불규칙요인의 스펙트럼분석은 계절과 요인주기에서 정점을 갖는지 자동적으로 찾는다. 정점들이 발견되면 경고메시지와 plot를 출력한다.

총합 계절조정 단계

(1 단계) 각 구성계열에 대해 계절조정을 위한 spc 화일을 만든다. 예를 들어, 미국의 1인 가구(c1fths) 구성계열은 다음과 같이 4개 지역으로 구성되며 spc 파일 이름은 다음과 같다.

북동부: cne1hs.spc, 중서부: cmw1hs.spc,
남 부: cso1hs.spc, 서 부: cwt1hs.spc

이때 각 구성계열은 3×9 계절이동평균법으로 계절조정하며, 총합계열(c1fths.spc)은 단순 합으로 합성한다. 북동부 계열의 계절조정 spc 화일(cne1hs.spc)은 아래와 같으며, 중서부, 남부, 서부의 1인 가구에 대한 각각의 spc 화일도 만든다.

```
series { title = "Northest ONE-FAMILY Housing Starts"
           file = "cne1hs.ori"   name = "CNE1HS"   format = "2R"
           comptype = add    }
X11 { seasonalma = (s3x9)
        title = ("Component for Composite Adjustment"
                  "of Total U.S. 1-Famliy Housing Starts") }
```

(2 단계) 총합계열을 계절조정하기 위한 spc 화일을 만든다. 계열의 직접계절조정방법은 승법으로 3×9 계절이동평균법을 사용한다. 직접조정의 계절인자(D10)와 간접법에 의한 계절인자(isf)를 c1fths.*에 저장한다.

```
composite { title = "TOTAL ONE-FAMILY Housing Starts"
              name = "C1FTHS"     save = (indseasonal) }
X11 { seasonalma = (s3x9)
        title = "Composite adj. of 1-Family housing starts"
        save = (D10 isf)  }
```

(3 단계) 구성계열의 spc 화일(cne1hs.spc, cmw1hs.spc, cso1hs.spc, cwt1hs.spc)과 총합계열의 spc 화일(c1fths.spc)에 대한 metafile을 생성하며, 생성된 metafile을 hs1ftot.mta로 저장한다.

<hs1ftot.mta의 metfile 구성 형태>

cne1hs
cmw1hs
cso1hs
cwt1hs
c1fths ← *총합계열의 spc 화일*

(4 단계) 위의 예제를 X-12-ARIMA로 실행하기 위하여 hs1ftot.mta 파일을 실행한다. 실행 결과는 각 구성계열에 대한 spc 이름과 총합계열의 spc 이름을 따른다;
cne1hs.out, …, cwt1hs.out, c1fths.out에 저장된다

```
x12a -m hs1ftot.mta
```

ESTIMATE

내 용

regression과 arima 명령에 의해서 RegARIMA 모형을 추정한다. 다양한 추정옵션을 사용할 수 있으며, 추정결과는 AR, MA와 회귀 모수에 대한 점 추정치와 추정된 표준오차, 분산(σ^2)의 최우추정치, 각 회귀모수의 t-통계량, 회귀효과와 관계되는 모수들의 결합 유의성(joint significance) 평가를 위한 χ^2-통계량, 모형 선택을 위한 우도함수(기본통계량) 등을 포함한다. 회귀모수의 χ^2-통계량은 고정된 계절효과, 요일효과, 사용자가 정의한 회귀효과에 의해 산출된다.

일반형식

estimate { tol = 1.0e-5
maxiter = (200)
exact = arma
outofsample = yes
print = (none +model +estimate +lkstats)
save = (model)
savelog = (aic bic)
}

명 령

exact 추정, 우도함수 계산, 예측을 위해 조건부 또는 정확한 우도함수를 이용할 것 인지에 대한 명령이다. default인 exact=arma는 AR과 MA 모수를 정확한 우도함수를 이용하여 구하고, exact=ma는 AR 모수는 조건부이지만 MA 모수는 정확한 우도함수를 이용한다. exact=none은 AR과 MA 모수가 모두 조건부 우도함수를 이용한다.

maxiter ARIMA 모형의 반복추정 시, 최대 반복 횟수이다. 회귀변수를 가진 RegARIMA 모형에서 반복적 일반화 최소자승법(IGLS; Iterative Generalized Least Square)에 의한 반복이 ARMA 총 반복 횟수까지 적용되지만, 회귀변수가 없는 RegARIMA 모형에서는 ARMA 반복 추정 횟수가 최대 횟수이다. default는 maxiter= 500 이다.

outofsample 최근 3년 예측오차(진단 통계)의 평균 크기를 계산하기 위해 사용된 예측오차의 종류를 결정한다. outofsample=yes 이면, out-of-sample의 예측오차가 사용된다. 이것은 모형을 추정하기 위해 사용된 시계열에서 예측기간 내에 있는 시계열을 제외하고 (가장 최근 3년 동안에 대한) 1년을 예측한다. outofsample=no이라면, within-sample 예측오차가 사용된다. default는 outofsample=no이다.

tol 비선형 추정에 대한 수렴한계(convergence tolerance)이다. 로그-우도함수의 절대 변화는 반복 추정치의 수렴성을 보기 위하여 tol과 비교한다. 회귀변수를 가진 모형에서는 새로운 AR과 MA를 재추정하기 위한 IGLS 반복 추정법의 수렴성을 검토하기 위하여 tol이 사용된다. IGLS 방법을 사용하지 않는 회귀변수가 없는 모형에서 tol은 AR과 MA 모수를 추정하기 위하여 사용된 비선형 반복법의 수렴성을 검토하기 위하여 사용된다. default 값은 tol = 1.0e-5.

print & save 출력표는 default 출력표와 이외의 출력표가 있다.

<Estimate: default 출력표>

이 름	단축	저장	내 용
options	opt	·	추정옵션의 header
model	mdl	+	print를 사용하면, 모형의 간략한 설명을 인쇄하고, save를 사용하면 지정된 ARMA 모수 초기 값을 사용하여 추정된 모형과 관련된 regression과 arima 명령이 포함된 화일을 생성
estimates	est	+	표준오차를 가진 회귀와 ARMA 모수 추정
averagefcsterr	afc		최근 3년에 대한 예측오차의 평균크기
lkstats	lks	+	exact=arma (default 옵션)를 사용하면, 최종 모수를 로그-우도함수로 추정하고, 모형선정 통계량(AIC, F-corrected AIC, Hannan-Quinn, BIC) 출력

<Estimate: 이외의 출력표>

이 름	단축	저장	내 용
iterations	itr	+	log-우도 값과 모수, 함수 계산과 반복 추정에 대한 자세한 내용을 출력
iterationerrors	ite	·	수렴 실패 등에 관한 반복 추정 에러 메시지
regcmatrix	rcm	+	print 옵션을 사용하면, 회귀모수의 상관행렬(correlation matrix)을 추정하고, save 옵션을 사용하면 회귀모수의 공분산 행렬(covariance matrix)을 추정
armacmatrix	acm	+	print 옵션을 사용하면, ARMA 모수의 상관행렬(correlation matrix)을 추정하고, save 옵션을 사용하면, ARMA 모수의 공분산 행렬을 추정
lformulas	lkf	·	로그-우도함수와 모형 선정 기준을 계산하기 위한 공식
roots	rts	+	추정된 모형에서 자기회귀와 이동평균 연산자의 제곱근
regressioneffects	ref	+	$X\hat{\beta}$ (회귀변수 행렬에 추정된 회귀계수의 벡터를 곱한 것)
residuals	rsd	+	날짜 또는 관측치 수와 관련된 모형 잔차

기타 명령

file X-12-ARIMA 실행에서 모형을 설정하기 위한 파일의 이름을 지정한다. 파일은 save=model 혹은 save=mdl로 함으로써 생성된다. 파일이름은 “ ”으로 하며, 파일이 현재의 폴더에 없는 경우에는 경로를 지정해 주어야 한다. file 옵션을 사용하면, arima 명령의 model, ma, ar 옵션과 regression 명령의 variable, user, b 옵션 뿐만 아니라 autmdl 명령, pickmdl 명령을 사용할 수 없다.

fix file 옵션의 모형 파일에 의해 추정된 어떠한 계수를 다음 단계의 추정을 위한 초기 값으로 사용하는 대신 고정된 값으로 사용할 것 인지를 기술한다. fix=all이면, 회귀모수와 arma 모수 추정시 model 파일의 값을 고정해서 사용한다. fix=arma이면, ARMA 모수만 고정, fix=none이면, 고정된 모수는 없다. default는 fix=nochange이며, model 파일에서 고정된 값으로 보관되며 다른 계수를 재추정한다.

참　고

대체적으로 X-12-ARIMA의 추정 결과는 “standard(또는 정상성)" 가정하에 충분히 긴 시계열에서 유효하다. 정확한 우도함수를 사용한다면, 모형 선정을 위한 통계량으로 우도함수를 계산한다.

반복 추정치가 어떠한 값에 수렴하면 추정 결과를 출력한다. 만약 추정치가 수렴하지 않으면, 마지막 반복 추정치 결과를 보여주고 이를 출력한다. 이때 이 값들을 모수 추정치로서 사용하지 않고, 반복실행이 끝났을 때 얻어진 모수를 시작 값으로 사용한다.

tol 값을 너무 크거나 작게 지정하면 않된다. tol 값을 너무 크게 지정하면 MLE의 참값에서 크게 벗어난 추정치가 결정되고, 반대로 너무 작으면 불필요한 반복을 많이 하게 되거나 또는 추정결과의 정도가 낮게 된다. default tol은 10^{-5}가 적당하다. Double precision의 tol 값을 컴퓨터가 갖는 한계 정도(machine precision) 수준으로 낮게 정하면(PCs와 Sun4 computers인 경우 대략 10^{-14}) 에러가 발생하므로 주의해야 한다.

회귀변수를 가진 모형에서 두 번째 수렴 한계는 각각의 IGLS 내에서 ARMA 반복에 의해 추정치를 결정하는데 필요하다. 여기서는 두 번의 IGLS 반복에서 수렴 한계를 100×tol 값으로 정하고, 그 후에는 원래의 tol 값으로 재선정된다. 비교적 큰 변동은 초기 IGLS 반복추정 방법에 의하여 회귀모수가 만들어지기 때문에, 처음의 tol 값 범위 내에서 ARMA 모수를 결정할 필요는 없다. default tol 값은 10^{-5}이며, 두 번의 IGLS 반복을 하게 되면 ARMA 수렴한계는 10^{-3}이 되고, 그 후의 tol 값은 다시 10^{-5} 된다. 또한 회귀변수를 가진 모형에서 한계 값은 매번 IGLS 반복에서 사용 가능한 ARMA 반복의 최대회수로 정해야 하며, 한계 값은 40이다.

print = roots 옵션은 추정된 모형의 모든 AR과 MA의 table을 산출하며, 이 table에는 각 근에 대해 modulus와 [-0.5, 0.5] 구간에서의 분포를 제공한다. 1보다 큰 modulus의 근은 단위원 밖에 있으며, 안정적인 AR 또는 가역적인 MA 모형을 나타낸다.

우도함수가 AR 모수에 대해서 조건부로 정의(exact=ma 또는 exact=none)되었을 때, AR 근은 단위원 상 또는 단위원 내(modulus ≤ 1)에 있게 된다. 이때 MA 근은 가역성 때문에 추정되므로 단위근 밖(modulus >1)의 MA 근은 발생하지 않는다. 단위근 상(modulus=1)에 있는 MA 근

은 round-off error 내에서 추정할 수 있거나, 모수를 추정하는 동안 고정된 것으로 지정된 경우에만 추정할 수 있다.

예 제

(예 1) M_{it}를 월별 고정된 계절조정효과에 대한 회귀변수라 하자. 이때, $(1-B)(y_t - \sum_{i=1}^{11} \beta_i M_i t) = (1-\theta B)a_t$의 회귀계수는 일반화 최소자승법에 의해서 추정한다. MA 모수 값 Θ는 0.25로 고정된 값이며, 모형의 잔차는 *.rsd라는 확장자로 spc 파일 이름으로 현재 폴더에 저장된다.

```
series { title = "Monthly  sales"     start=1976.jan
         Data = (138 128 ⋯ 297)    }
regression {  variables=seasonal  }
arima  { model = (0,1,1)    ma=(0.25f)  }
estimate { save = residuals  }
```

(예 2) 10^{-4} 의 tol, default 보다 낮은 수렴 기준치와 최대 허용반복수는 100회의 조건을 갖고 다음의 계절모형을 추정한다.

$$(1-\phi B)(1-B)(1-B^{12})z_t = (1-\Theta B^{12})a_t$$

모형에 회귀변수가 없으므로, tol과 maxiter 옵션을 ϕ와 Θ를 추정하기 위한 비선형 ARMA 반복법에 적용할 수 있다. 모수 추정을 위해 사용된 우도함수는 정확한 MA와 조건부 AR이다. print 옵션에서 우도함수 값과 모수 값을 매 반복 시 인쇄하며 추정된 AR과 MA의 근을 인쇄한다.

```
series { title = "Monthly  sales"     start=1976.1
         Data = (138 128 ⋯ 297)  }
transform {  function = log }
regression { variables = (td ao1999.01)  }
arima  { model = (1,1,0)(0,1,1) }
estimate { tol=1e-4  maxiter = 100  exact = ma
            save=mdl  print=(iterations roots) }
```

(예 3) (예 2)의 file 옵션에 의해 저장된 RegARIMA 모형의 추정치를 이용한 것 이외에는 (예 2)와 같다. 모든 모수 추정치는 모형 파일에 저장된 값을 고정 값으로 이용한다.

```
series { title = "Monthly   Inventory"    start=1976.12
         Data = (1209   834   ⋯ 1002)  }
transform  { function = log    }
arima  { model = (1,1,0)(0,1,1)  }
estimate { file = "test.mdl"
           fix=all  }
```

(예 4) 3개 자료 추가되었으며, 이들 자료가 특이치인지 확인하는 것 이외에는 (예 3)과 같다. (예 2)와 (예 3)의 마지막 자료의 월은 1999년 12월이다. RegARIMA 모형의 모수는 (예 2)의 기간에서 얻어진 값을 이용한다.

```
series { title = "Monthly   Inventory"    start=1976.12
         Data = (1209   834   ⋯ 1002   1425 901 1375)  }
transform  { function = log    }
arima  { model = (1,1,0)(0,1,1)  }
estimate { file = "test.mdl" fix=all  }
outlier { span=(2000.01, ) }
```

FORECAST

내 용

추정된 모형을 사용하여 series 명령에서 주어진 시계열에 대해서 미래(과거) 시계열을 예측한다. 변환된 계열의 점 예측치, 예측 표준오차와 원계열에 대한 점 예측치와 예측구간을 출력한다.

일반형식

forecast { maxlead = 12
maxback = 12
probability = 0.95
exclude = 10
print = (none +original +transformed)
save = (variances)
}

명 령

exclude 예측 전에 시계열의 마지막 관측치부터(또는 series 명령의 span 옵션에 의해 지정된 기간의 끝부터) 제외되는 관측치 개수를 지정한다. exclude=0이 default이며, 시계열의 끝부터 예측을 시작한다.

maxback 추정하는 과거 시계열의 수를 지정한다. default는 0이며, 최대 60이다. (최대한계 60은 바꿀 수 있다. 2.6장 참조) 참고 : 15년 이상의 과거 시계열은 추정되지 않는다.

maxlead 추정하는 미래 시계열의 수를 지정한다. default는 12이고 최대는 60이다. (최대한계 60은 바꿀 수 있다. 2.6장 참조)

probability	정규성을 가정하고, 예측구간의 확률범위(coverage probability) 정한다. default probability = 0.95이다. 예) 확률범위가 95%일 때, 변환된 시계열의 예측구간은 '점 예측치±1.96×예측 표준오차'이다.
print & save	선택 가능한 출력표는 다음과 같다.

<forecast: 이용 가능한 출력표>

이 름	단축	저장	내 용
transformed	ftr	+	예측표준오차에 의한 변수변환된 시계열의 예측
variances	fvr	+	변수변환된 시계열의 총 예측오차 분산에 대한 확률분산과 회귀분산
forecasts	fct	+	상한과 하한의 예측구간을 갖는 원계열에 대한 점추정

참 고

예측(forecasting)은 추정된 모형에 의해 실시된다. estimate 명령이 없으면, forecast 명령은 예측 전에 실행되어진 추정량(default)을 사용한다. 예측에 사용되는 모형은 regression과 arima 명령에 의해 지정되며, outlier 명령이 있으면 자동으로 식별된 특이치가 회귀변수로 추가되어 실행된다. 특이치가 시계열의 끝에서 발견될 때, 이 특이치는 모수 추정에 영향을 주어 간접 또는 직접적으로 예측에 영향을 줄 수 있다.

모형에 하나 이상의 이동평균 연산자를 포함하고 있으면, 예측은 추정된 모형의 잔차에 의해 달라질 수 있다. estimate 명령의 exact 옵션은 exact 우도 함수(default) 또는 조건부 우도함수에 의해 계산할지를 결정한다.

예측 표준편차는 모형의 회귀모수 추정에서 발생하는 오차는 고려하지만, AR과 MA 모수의 추정에서 발생하는 오차는 고려하지 않는다.

모형에 사용자가 지정한 회귀변수가 존재한다면, 모든 시점의 시계열 예측 시에 이를 고려해야 한다. 예측구간은 "점 예측치 ± K × 예측표준오차"로 구한다. K는 지정된 포함범위확률에 해당하는 정규분포표의 표준

오차 승수이다.

원계열의 점 예측치와 예측구간 한계는 변수 변환된 시계열에서 점 예측치와 예측구간 한계를 역변환하여 구한다. 변환 계열에는 Box-Cox 또는 logistic 변환과 사전조정요소(regression 명령에 포함된 variables=td에 포함된 월(분기)길이)가 있다.

transform 명령에 사용자가 정의한 사전조정요인을 포함하고 있다면, 예측기간을 역변환하고자 하는 결과에 대해서 적용하여야 한다. 만약 예측기간에 대해서 변환을 하지 않는다면, 예측 기간은 1로 가정된다. 이때, 예측에 대한 사용자 정의의 사전조정효과는 무시된다.

exclude > 0을 사용하는 이유는 모형 예측력을 평가할 목적으로 보류하고 있는 자료의 어떤 시점에 대한 예측치를 만들기 위해서 이다.
X-12-ARIMA는 관측된 자료값이 있는 예측기간의 (관측치 - 점 예측치)인 실제 예측오차를 인쇄하여 비교를 한다. 예측에서 보류된 자료가 모형 추정에서는 보류되지 않기 때문에 exclude > 0 인 경우, 표본 내(within-sample) 비교를 하게 된다. 더욱 더 현실적인 표본 외(out-of-sample) 예측 비교는 모형추정과 예측 시에 자료를 보류함으로써, series 명령의 span옵션을 이용한다(예제 4 참조).

계절조정이 실시된 X-12-ARIMA에서 생성된 미래 또는 과거 예측치는 원계열에 덧붙여지며, 계절조정 절차는 연장된 미래 또는 과거 시계열에 적용된다. REGARIMA 모형이 사용되지만 forecast 명령이 사용되지 않는 계절 조정 명령이라면, 모형으로부터 1년의 미래 예측치가 만들어진다. 미래 예측치의 연장없이 계절조정하는 방법은 maxlead=0이다.

REGARIMA의 사전조정요인인 요일과 특이치, 휴일 효과, 사용자 정의의 회귀효과 등이 원계열에서 사전조정되었다면, 미래와 과거 예측치도 역시 조정이 된다.

경 고

X11 명령에 의해 계절조정을 하면, exclude는 시계열의 끝에서부터 관측치를 제거시키기 위해 사용할 수 없다. 이 경우, exclude는 0이 되며, 경고문이 출력된다.

예 제

(예 1) 월간 시계열 자료로 12개월을 예측하고, 95% 예측구간 (점 예측 ±1.96×표준오차)을 구하는 예로, 모든 옵션을 default로 처리한다. estimate 명령 대신 forecast 명령으로 default 예측 옵션을 갖고 모형추정을 한다. 지수변환과 $m_t/\overline{m}$ (regression 명령의 variables=td 옵션이용)을 곱하여 변환한 시계열에 대해 점 예측치와 예측구간을 구한다.

```
series { title = "Monthly  sales"   start = 1976.jan
         Data = (138 128 ⋯ 297)  }
transform { function = log  }
regression { variables = td  }
arima { model = (0 1 1)(0 1 1)12  }
forecast {   }
```

(예 2) (예 1)의 시계열에 대해 24개월을 예측한다. outlier 명령에 의해 예측하기 위하여 사용된 추정 모형에는 검색된 AO와 LS 특이치를 포함될 것이며, regression 명령에 의해서 요일변수가 지정된다.

```
series { title = "Monthly  sales"   start = 1976.jan
         Data = (138 128 ⋯ 297)  }
transform { function = log  }
regression { variables = td  }
arima { model = (0 1 1)(0 1 1)12  }
estimate  {   }
outlier   {   }
forecast  {maxlead = 24   }
```

(예 3) 시계열 전체를 이용하여 모수 추정을 하고, 예측 시에는 시계열의 끝에서 부터 10개월을 제외한 15개월을 예측한다. 제외된 10개월은 표본 내(within-sample) 예측오차를 산출한다. 또한 90% 포함범위확률(coverage probability)에 의해 예측구간 한계(점 예측치 ± 1.645 × 예측표준오차)를 구한다.

```
series { title = "Monthly  sales"   start = 1976.jan
         Data = (138 128 ⋯ 297)  }
transform { function = log  }
regression { variables = td  }
arima { model = (0 1 1)(0 1 1)12  }
estimate  {   }
orecast  {maxlead = 15
          probability = .90
          exclude = 10       }
```

(예 4) 1992년 3월까지의 시계열에서 series 명령의 span 옵션을 사용하여 시계열의 마지막 관측치로 부터 24개 관측치를 제외한 시계열로 모형추정을 한다. 즉 1990년 3월까지의 시계열로 모수추정을 한다. 또한 원계열에서 제외된 24개 관측치에 대해서 표본 외(out-of-sample) 예측오차를 출력한다.

```
series { title = "Monthly  sales"   start = 1976.jan
         Data = (138 128 ⋯ 297)
         span = (  ,1990.mar)   }
transform { function = log  }
regression { variables = td  }
arima { model = (0 1 1)(0 1 1)12  }
estimate  {   }
forecast  {maxlead = 24  }
```

(예 5) 월 시계열에서 12개월의 미래 시계열과 95%의 예측구간(점예측치 ± 1.96 ×표준오차)을 구한다. 이때, 모든 옵션은 default로 처리하며, 12개월의 과거 시계열을 예측한다. 예측된 미래 및 과거 시계열에 대해 승법계절조정(default)을 하고, 요일효과로 사전조정을 한다.

```
series { title = "Monthly   sales"    start = 1980.jan
            file = "ussales.dat"      }
transform { function = log   }
regression { variables = td   }
arima { model = (0 1 1)(0 1 1)12   }
forecast   {maxback = 12   }
x11   {    }
regadjust { prior = td    }
```

HISTORY

내 용

i) 초기(동시 또는 예측된) 계절조정계열의 수정률(revision), ii) 표본 외(out-of-sample) 예측오차, iii) 우도함수 통계량 등의 historical record을 작성하기 위하여 사용하는 명령이다. 사용자는 i)~iii) 모두 또는 각각에 대해 시작 시점을 지정할 수 있다. 예측오차를 구하면, 사용자는 예측 기간벡터를 지정할 수 있다.
대체로 수정률 분석(revision history analysis)을 하면 프로그램 실행시간이 늘어난다.

일반형식

history { estimates = (sadj fcst trend)
sadjlags = (1, 2, 3, 12)
trendlags = (1, 2, 3)
target = final
start = 1985. jan
fstep = (1 2)
fixmdl = no
fixreg = outlier
endtable = 1990. jan
print = (all -revvalsa)
save = (sar trr fcsterrors)
savelog = (aveabsrevsa aveabsrevtrend) **}**

명 령

estimates RegARIMA 모형 또는 X-11 계절조정에서 결정된 추정치를 분석한다. *예) estimate=(sadj aic)* default는 계절조정계열(sadj)이다. 다음은 estimates 옵션에서 이용할 수 있는 추정치이다.

<history: estimates에서 이용 가능한 명령어>

이 름	설 명
sadj	최종 계절조정계열(총합 계절조정을 실시한 경우, 직접법 계절조정 계열 이용)
sadjchng	최종 계절조정계열의 전월(분기)비
trend	최종 핸더슨 추세요인
trendchng	최종 핸더슨 추세요인의 전월(분기)비
seasonal	최종 계절요인과 추정된 계절요인
aic	RegARIMA모형의 AIC와 최대 로그 우도함수
fcst	예측치와 RegARIMA모형에서 생성된 예측오차 평균제곱. 예측치를 생성하지 않은 경우에는 사용할 수 없다.

endtable 계절조정계열과 추세 추정치의 수정률 분석 결과표와 이들의 전월(분기)비에 대한 최종 계절조정계열과 추세계열을 이용한다. endtable의 값이 지정되지 않았다면, 계열 끝 시점이 사용되거나 sadjlags 혹은 trendlags에서 지정된 가장 긴 시차보다 한 시점이 큰 값이 사용된다. 이 옵션은 우도 추정과 예측치에 대한 수정률 분석에 영향을 주지 않는다. 예) endtable=1990.jan

fixmdl 수정률 분석을 하는 동안 RegARIMA 모형을 재추정할 것인지를 결정한다. fixmdl=yes이면, RegARIMA 모형의 ARIMA 모수와 회귀계수는 전체 기간(modelspan 옵션이 지정되면, 모형 기간 동안)을 이용한 시계열에서 추정된 값이 분석하는 동안 고정된 값으로 사용된다. default인 fixmdl=no이면, RegARIMA 모형의 모수는 시계열의 끝 시점이 변하면, 매번 다시 추정된다.

fstep 예측오차의 수정률 분석에서 분석하게 될 4개 까지의 예측 leads 벡터를 지정한다. 예를 들어 fstep=(1 2 12)는 1-step, 2-step, 12-step 앞선 예측치에 대해 오차분석을 한다. default는 월별자료는 (1 12), 분기 자료는 (1 4)이다. 벡터값은 1이상으로 forecast 명령의 maxlead 옵션의 지정값을 초과할 수 없다.

fixreg RegARIMA 모형 혹은 불규칙요인을 이용한 회귀모형의 회귀계수를 고정한다. 계수들은 series나 composite 명령에 나타나는 전체 기간(model span)에서 구해진 값들이 고정된다. 다른 모든 계수들은 매 분석마다 재추정 된다. 요일효과(td), 휴일효과(holiday), 특이치(outlier), 사용자 정의 회귀효과(user)를 고정할 수 있다. RegARIMA 모형이나 불규칙요인을 이용한 회귀모형이 시계열에 적합되지 않거나 fixmdl=yes를 지정한 경우에는 이 옵션은 무시된다.

sadjlags 차분된 계절조정 결과의 수정률 분석에서 5개 까지 차분된 수정률 벡터를 지정한다. 이 옵션은 estimates 옵션에서 계절조정계열이나 계절조정계열에 대한 전월(분기)비의 수정률 분석이 지정되었을 때 유효하다. default는 수정률 분석을 하지 않음이다.

start 수정률 분석의 시작 시점을 지정한다. 시작 시점을 지정하지 않으면, default로 가장 긴 계절 filter 길이에 따라 시작 시점이 결정된다. default 시작 시점은 가장 긴 filter가 3×3이거나 안정적인 filter인 경우에는 6년, 3×5 filter의 경우 8년, 3×9 filter의 경우 12년이다. 계절조정이 수행되지 않은 경우에는 default로 시계열의 시작 후 8년이다.

target sadjlags 혹은 trendlags에서 지정한 시차에서 계산된 계절조정과 추세에 대한 수정률을 동시 추정치 혹은 최종 추정치에서 유도할 것 인지를 지정한다. default는 target= final.

trendlags sadjlags와 유사하며, 차분된 추세요인의 수정률 분석에 이용된다. 차분벡터는 5개까지 지정할 수 있으며, 이 옵션은 estimates 옵션에서 추세요인이나 추세요인의 전월(분기)비 수정률 분석이 지정되었을 때 유효하다. default는 수정률 분석을 하지 않음이다.

print & save 직접법과 간접법 계절조정 결과에 의한 default 출력표는 다음과 같다. * 실제 인쇄는 한정되어 있다.

<history: default 출력표>

이 름	단축	저장	내 용
outlierhistory	rot	+	(특이치 자동 검색을 하는 경우) 수정률분석을 위해 제거되었거나 유지된 특이치 기록
sarevisions	sar	+	가장 최근 계절조정 추정치에 대한 동시 수정률
sasummary	sas	·	계절조정 수정률의 요약통계
chngrevisions	chr	+	계절조정 자료의 가장 최근 전월(분기)비 추정치에 대한 동시 수정률
chngsummary	chs	+	계절조정계열의 전월(분기)비 수정률에 대한 수정률 요약 통계
indsarevisions	iar	+	가장 최근 계절조정계열의 추정치에 대한 동시 수정률
indsasummary	ias	·	간접 계절조정 수정률에 대한 요약 통계
trendrevisions	trr	+	추세요인의 가장 최근 추정치에 대한 동시 수정률
trendsummary	trs	+	추세요인 수정률에 대한 요약 통계
trendchngrevisions	tcr	+	추세요인의 가장 최근 전월(분기)비 추정치에 대한 동시 수정률
trendchngsummary	tcs	·	추세요인의 가장 최근 전월(분기)비 수정률에 대한 요약 통계
sfrevisions	sfr	+	계절요인과 예측된 계절요인의 가장 최근 추정치에 대한 동시 수정률
sfssummary	ssm	·	계절요인 수정률에 대한 요약 통계
lkhdhistory	fch	+	AICC와 우도 값의 history
fcsterrors	fce	+	예측 오차의 누적 자승합의 수정률

기타명령

outlierwin 특정기간의 수정률 분석을 하는 동안 얼마나 많은 관측치로 특이치를 식별할 것 인가를 지정하는 명령으로, default는 월 자료 12, 분기자료 4이다. 이 옵션은 outlier=auto 명령과 같이 사용하면 효과적이다.

refresh 두 개의 옵션으로 초기 값을 지정하여 RegARIMA 모형의 모수를 추정한다. refresh=yes는 최종 모형에서 계산된 추정 모수를 수정률 분석시 실행하는 RegARIMA 모형의 시작 값(starting value)으로 사용한다. default인 refresh=no는 RegARIMA 모수의 초기치를 전체 계열에서 추정한다.

outlier 수정률 분석을 하는 동안 RegARIMA 모형이 재추정될 때마다 자동으로 특이치를 식별할지를 지정한다. outlier 명령을 사용하지 않으면, 이 옵션은 유효하지 않다.
outlier=keep이면, default로서 수정률 분석을 하는 동안, 전체 계열에서 자동으로 식별한 특이치를 RegARIMA 모형에 반영한다. 추정된 특이치 회귀계수는 RegARIMA 모형의 모수가 재추정될 때마다 재추정된다. outlier=remove이면, 전체 계열에서 자동으로 식별된 특이치는 RegARIMA 모형에서 제외된다. 이때, 특이치 효과들은 추정하지 않고 제외된다. outlier=auto이면, 전체 계열에서 자동으로 인지된 특이치 중에서 수정률 분석의 시작 시점전에 outlierwin 관측치수까지 RegARIMA 모형에 자동으로 포함한다.

참 고

총합계열에 대한 계절조정계열의 수정률 분석을 지정했을 때, 직접법과 간접법에 의한 계절조정계열의 수정률이 제공된다. 수정률 분석에서는 계절조정하지 않은 계열도 각 구성계열을 지정해야 한다. 간접법 계절조정계열(sadj)의 수정률 분석은 간접법 계절조정을 한 경우에 대해서만 이용가능하다.

자동 계절 filter 선정 옵션이 사용되면, 수정률 분석을 하는 동안 자료의 기간이 변경될 때 마다 계절 filter를 다시 재선정한다.

예측 수정률을 위한 시작 시점은 fstep에서 주어진 값에 의존한다. n-step 앞선 예측오차에 대한 시작 시점은 수정분석 시작 시점인 n이다. 예를 들어, fstep=(1 12)이고 start=1992.jan이면, 1-step과 12-step 앞선 예측은 1992년 2월과 1993년 1월이다.

예 제

(예 1) RegARIMA 모형을 이용하여 예측 또는 회귀 특이치 조정을 사용하지 않고, 모든 월에 대해 3×9 계절이동평균을 하는 승법 계절조정 예이다. 계절조정계열의 수정률 분석을 모든 자료에 실행하고, 동시 관측치 이후 2개월을 추정한다.

```
series { title = "sales  of  livestock"   start = 1967.1
         file = "cattle.ori"   }
x11 {seasonalma = s3x9  }
history  { sadjlags = 2 }
```

(예 2) 특이치와 수준변동 사전조정을 위한 회귀변수를 포함한 계절 ARIMA 모형을 이용한다. 지정된 회귀변수는 상수, 요일효과, 2개월의 수준변동(1972년 5월, 1976년 9월)등이며, ARIMA 모형은 (0,1,2)(1,1,0)12이다. 1-step 예측오차에 대해 수정률 분석을 하며, 이때 분석의 시작 시점은 1975년 1월이다.

```
series { title = "Export of leather goods"
         start = 1969.jul    file = "expleth.dat "  }
regression { variables = (const  td  ls1972.may  ls1976.oct ) }
arima { model = (0 1 2)(1 1 0)  }
estimate  {   }
history  { estimates = fcst   fstep = 1   start = 1975.jan   }
```

(예 3) (예 2)와 동일한 RegARIMA 모형에 대해 1-step, 12-step 예측오차에 대해 수정률 분석을 실행하여 저장하는 프로그램이다. 12-step 예측오차가 없는 경우(1975년 2월~12월)에는 ‘0’을 저장한다.

```
series { title = "Export of leather goods"
         start = 1969.jul     file = "expleth.dat "  }
regression { variables = (const  td  ls1972.may  ls1976.oct ) }
arima { model = (0 1 2)(1 1 0)  }
estimate  {   }
history  { estimates = fcst  save = r6   start = 75.jan    }
```

(예 4) RegARIMA 모형에 의해 총합계열이 연장되고 3×3 계절이동평균 갖는 월별 총합 계절조정을 실시한다. RegARIMA 모형은 마지막 12월까지 적합된다. 직접법과 간접법 계절조정계열과 직접법 계절조정에 의한 추세요인의 수정률 분석은 매년 12월에 RegARIMA 모형 모수를 재추정하면서 실시한다. 추정치 각각의 수정률 퍼센트는 별도의 파일에 저장된다.

```
composite { title = "Total Housing Starts in the US"
            modelspan=(, 0.Dec)
          }
regression { variables = td        }
arima { model = (0 1 1)(0 1 1)  }
x11  { seasonalma  = s3x5  }
history { estimates = (sadj trend)
          save = (sar  iar  trr  )   }
```

IDENTIFY

내　용

RegARIMA 모형의 ARIMA 부분을 식별하기 위하여 표본 ACF와 PACF의 그림과 표를 만들기 위한 명령이다. 표본 ACF와 PACF는 diff와 sdiff 옵션에 따라 계절·비계절 차분의 모든 조합에 의해 만들어진다. ACF와 PACF의 표와 그림은 regression 명령이 있으면 회귀 잔차 계열을 차분하며, regression 명령이 없으면 원계열을 차분한다.

일반형식

identify { **diff** = (0, 1)
sdiff = (0, 1)
maxlag = 36
print = (none+acf+acfplot+pacf+pacfplot)
}

명　령

diff　비계절 차분의 차수를 기술한다. 0이면 차분을 하지 않음, 1이면 1차 비계절 차분 (1-B), 2는 2차 비계절 차분 $(1\text{-}B)^2$ 등이다. 기술된 ACF와 PACF는 서술된 비계절 차분의 모든 차수로 만들어지며, sdiff에서 기술된 계절 차분의 모든 차수와 합쳐진다. default는 diff=(0)이다.

maxlag　ACF와 PACF의 그림과 표에 기술되는 시차의 길이를 나타낸다. default는 36이다.

sdiff　계절 차분의 차수를 기술한다. 0이면 차분을 하지 않음, 1이면 1차 계절 차분 $(1-B^s)$, 2는 2차 계절 차분 $(1-B^{2s})$ 등이다. 기술된 ACF와 PACF는 서술된 계절 차분의 모든 차수로 만들어지며, diff에서 기술된 비계절 차분의 모든 차수와 합쳐진다. default는 sdiff=(0)이다.

print & save 선택 가능한 출력표는 다음과 같으며, default로 제공된다.

<Identify: 이용 가능한 출력표>

이 름	단축	저장	내 용
acf	iac	+	시차에 대한 표준오차, Ljung-Box, Q-통계량과 표본 자기상관함수
acfplot	acp	·	±2 표준오차 한계를 갖는 표본 자기상관함수의 그림
pacf	ipc	+	각 시차에 대한 표준오차와 표본 편자기상관함수
pacfplot	pcp	·	±2 표준오차 한계를 갖는 표본 편자기상관함수의 그림

참 고

regression 명령이 있다면, diff와 sdiff에 의해 기술된 차분의 최대 차수에 따라 transform 명령이 수행된 후의 계열과 회귀변수가 차분된다. 차분된 계열은 차분된 회귀 변수에 회귀한다. 추정된 회귀계수$\hat{\beta}_i$는 차분하지 않은 회귀효과($\sum \tilde{\beta}_i x$)를 계산하기 위하여 사용되며, 그때 차분하지 않은 회귀오차계열을 만들기 위하여 차분하지 않은 자료 y_t에서 뺀다 ($\tilde{z}_t = y_t - \sum \tilde{\beta}_i x$). 회귀오차계열과 diff와 sdiff에서 기술된 차분은 ACF와 PACF를 만들기 위하여 사용한다.

ACF와 PACF는 diff와 sdiff에서 기술된 계절·비계절 차분 차수의 모든 조합에 의해서 만들어진다. 예를 들어, diff=(0, 1), sdiff=1이라면, ACF와 PACF는 $(1-B^s)\tilde{z}_t$와 $(1-B)(1-B^s)\tilde{z}_t$, 여기서 $\tilde{z}_t$는 회귀오차계열이며 s는 series 명령의 계절 주기이다. diff=(0, 1, 2)와 sdiff=(0, 1)를 지정했다면, 다음과 같은 6개의 ACF와 PACF가 계산된다.

$$\tilde{z}_t, (1-B)\tilde{z}_t,\ (1-B)^2\tilde{z}_t,\ (1-B^s)\tilde{z}_t,$$
$$(1-B)(1-B^s)\tilde{z}_t,\ (1-B)^2(1-B^s)\tilde{z}_t$$

identify 명령과 estimate 명령이 같이 있는 경우에는 identify 명령이 우선한다. identify 명령은 regression 명령의 정보를 사용하며, arima 명령은 있더라도 무시된다.

특이점(singularity)은 치명적인 오류를 야기시키므로, 차분을 함으로써

사용자가 정의한 회귀변수를 포함한 회귀 변수가 특이점을 만들지 않도록 해야 한다. 특이점을 만들지 않기 위해서는 regression 명령에 variables=(seasonal)을 포함하는 한편, sdiff를 양의 값(즉, 1)으로 지정하면 된다.

ACF와 PACF에 요구되는 시차의 길이를 시계열 길이와 같거나 길게 한다면, ACF와 PACF는 가능한 가장 긴 시차만을 계산한다.

예 제

(예 1) 월별 시계열 Y_t을 log 변환한 $y_t = \log(Y_t)$에 대하여 필요한 차분의 차수를 파악하기 위한 ACF 표를 만들고자 한다. ACF는 y_t, $(1-B)y_t$, $(1-B^{12})y_t$와 $(1-B)(1-B^{12})y_t$를 계산한다. regression 명령은 없으므로 제거할 회귀효과는 없다. ACF는 default인 36시차로 계산한다.

```
series { title = "Monthly Sales"    start = 1976.jan
          data = (138 128 … 297) }
transform { function = log }
identify { diff = (0,1)        sdiff = (0,1)    print = (none+acf) }
```

(예 2) 표본 ACF, PACF를 계산하기 전에 고정된 계절효과를 제거한다. regression 명령에 추세상수 뿐 아니라 고정된 계절변수를 포함한다. identify 명령은 차분된 회귀변수 $(1-B)x_{it}$에 $(1-B)y_t$를 회귀, 차분하지 않은 회귀잔차 $\widetilde{z}_t = y_t - \sum_{i=2}^{12} \widetilde{\beta}_i x_{it}$를 계산함으로써(추세상수항 $\widetilde{\beta}_1 x_{1t}$를 빼지 않는다) 고정된 계절효과를 제거한다. 이때 $\widetilde{z}_t$와 $(1-B)\widetilde{z}_t$의 ACF와 PACF를 계산한다. 상수항은 $(1-B)y_t$에 있어 0이 아닌 평균을 따르므로 선형추세상수(x_{it}=t)이다.

```
series { title = "Monthly Sales"    start = 1976.jan
          data = (138 128 … 297) }
regression { variables = (const seasonal) }
identify { diff = (0, 1) }
```

(예 3) Y_t는 (예 1)과 같은 시계열이며, 1차 계절·비계절차분을 하며 요일 및 부활절 효과를 포함한다. regression 명령이 있기 때문에 identify 명령은 $(1-B)(1-B^{12})x_{it}$에 $(1-B)(1-B^{12})y_t$를 회귀한다. 여기서 x_{it}는 요일과 부활절효과에 대한 회귀변수이며, y_t는 달 길이효과에 의해서 조정된 원계열 Y_t의 로그변환이다. $\tilde{\beta}_i$가 추정된 회귀계수라면, identify 명령은 회귀잔차계열 $(1-B)(1-B^{12})(y_t - \sum \tilde{\beta}_i x_{it})$에 대한 ACF, PACF 그림을 그린다. 30개월의 시차를 갖는 ACF와 PACF를 계산한다.

```
series { title = "Monthly Sales"    start = 1976.jan
         data = (138 128 ··· 297) }
transform { function=log }
regression { variables = (td easter[14]) }
identify { diff = (1)    sdiff = (1)
           print = (none+acfplot+pacfplot)
            maxlag = 30  }
```

(예 4) 1971년 1/4분기에서 수준 변화가 있는 시계열의 모형식별을 위하여 16분기 시차의 ACF와 PACF를 작성한다. 수준변화 효과는 차분된 수준변화 변수에 $(1-B)(1-B^4)y_t$를 회귀함으로써 추정한다. ACF와 PACF는 $\tilde{z}_t$, $(1-B)\tilde{z}_t$, $(1-B^4)\tilde{z}_t$와 $(1-B)(1-B^4)\tilde{z}_t$에 대해서 계산한다. 다음의 RegIMA 모형은 표준진단 검토와 추정을 한다.

$$(1-B)(1-B^4)(y_t - \beta LS71.1_t) = (1-\Theta B)(1-\Theta B^4)a_t$$

진단 결과 모형이 부적절하다고 한다면, 사용자는 새로운 모형을 선정하기 위하여 identify 명령을 사용한다.

```
series{ title ="Quarterly Sales"  start = 1963.1
        period = 4
        data =(56.7 57.7 ··· 68.0) }
regression { variables = (ls1971.1) }
arima { model = (0 1 1)(0 1 1)}
identify { diff = (0, 1)   sdiff = (0, 1)    maxlag = 16 }
estimate { }          check { }
```

OUTLIER

내 용

지정된 모형을 이용하여 가법 특이치(additive outliers: AO), 일시적 변화(temporary change outliers: TC), 수준변화(level shifts: LS) 또는 위의 세 가지 특이치 조합을 자동으로 검색하기 위한 명령이다. AO, TC, LS의 특이치가 자동으로 식별이 되면, 식별된 특이치는 회귀변수로 모형에 설정되고, 설정된 모형은 재추정된다. 이러한 과정은 특이치가 발견되지 않을 때까지 반복된다. 두 개 또는 그 이상의 수준변화가 있는 경우(또는 regression 명령에 의해 수준변화 특이치가 모형에 존재하는 경우), t-통계량은 두 개 또는 그 이상의 연속적인 수준변화에 대한 매번의 검정에서 "임시 수준변화를 채택할 수 없다"라는 귀무가설을 검정하는데 이용된다.

일반형식

outlier { **types** = all
critical = 3.75
method = addall
span = (1983.may, 1992.sep)
lsrun = 0
print = (none+header)
save = (tests)
}

명 령

types 검색할 특이치 형태를 나타낸다. type=ao는 가법특이치, type=ls는 수준변화, type=tc는 일시적 변화, type= all은 가법 특이치, 일시적 변화, 수준변화를 동시에 검색한다. type=none은 특이치 검색은 하지 않으나, 임시 수준변화에 대한 t-통계량은 계산된다. default는 type=(ao ls)이다.

critical 특이치 검색을 위한 임계치(critical value)를 설정하며, 특이치에 대한 t-통계량의 절대값과 비교한다. default 임계치는 특이치 검색을 하고자 하는 분석 기간 내의 관측치 수에 의해 결정되며, 이들 임계치는 3~99개 사이의 관측치 수에 따라 임계치를 보간한(interpolate) 륭(Ljung, 1993)의 수정 근사식에 의해 얻어진다. 아래표는 관측치 수에 따른 t-검정의 임계치이다. 하나의 임계치(critical=3.5)가 주어졌다면, 이 임계치는 모든 유형의 특이치(AO, LS, TS)에 적용된다. 임계치 형식이 *critical=(3.5, 4.0, 4.0)*인 경우, 첫 번째 값인 3.5는 가법특이치, 두 번째 값인 4.0은 수준변화특이치, 세 번째 값인 4.0은 일시적 변화에 대한 임계치이다. *critical=(3.25, ,3.25)*와 같이 수준변화의 임계치가 결측치이면, 관측치 수에 의해 default 임계치가 설정된다. 임계치를 크게 하면 특이치 검색의 민감도를 감소시켜 특이치로 판별되어야 할 관측치의 수가 줄어들 가능성이 있다.

<특이치 식별을 위한 default 임계치>

관측치 개수	임계치	관측치 개수	임계치
1	1.96	48	3.63
2	2.24	72	3.73
3	2.44	96	3.80
4	2.62	120	3.85
5	2.74	144	3.89
6	2.84	168	3.92
7	2.92	192	3.95
8	2.99	216	3.94
9	3.04	240	3.99
10	3.09	264	4.01
11	3.13	288	4.03
12	3.16	312	4.04
24	3.42	336	4.05
36	3.55	360	4.07

method 검색된 특이치를 모형에 어떻게 추가할 것인지를 결정한다. method=addone과 method=addall이 있으며, 자세한 것은 <참고>를 참조한다. default는 method=addone이다.

lsrun 2개 이상의 연속적인 수준변화에 대한 검정을 실시하며, “임시 수준변화를 채택할 수 없다”라는 귀무가설을 검정하기 위한 t-통계량을 계산한다. 수준변화에 대한 검정을 하기 위한 t-통계량은 추정된 모수의 합을 표준오차로 나눈 값으로 계산된다(Ott와 Bell, 1993). 자동적으로 식별된 수준변화와 regression 명령에서 지정된 수준변화가 검정에 사용된다. lsrun은 0부터 5까지의 값을 줄 수 있으며, 0과 1은 임시 수준변화의 t-통계량을 계산하지 않는다. lsrun에 명시한 값이 특이치 검색 모형에 사용된 수준변화의 총 갯수를 초과할 경우, lsrun은 이 총 갯수로 다시 조정된다. Default값는 0으로 임시 수준변화의 t-통계량을 계산하지 않는다. 자세한 사항은 <참고>를 참조한다.

print & save 다음은 이용 가능한 출력표이며, Header와 Temporaryls는 default로 출력된다.

<Outlier: 이용 가능한 출력표>

이 름	단 축	저 장	내 용
header	hdr	·	특이치 검색에 대한 옵션(임계치, 찾고자 하는 특이치의 형태 및 기간 등을 포함)
iterations	oit	+	검색된 특이치, 제외된 특이치, 모형 모수 추정 및 잔차표준편차에 대한 robust와 non-robust 추정을 포함한 특이치 검색의 반복 결과
tests	ots	·	반복된 특이치 검색에서 특이치 종류와 매 시점의 t-통계량
temporaryls	tls	·	임시 수준변화에 t-통계량의 요약
finaltests	fts	+	최종 특이치 검색 반복 작업 중 생성되는 특이치 종류와 매 시점의 t-통계량

span 특이치를 검색하고자 하는 시계열의 기간을 나타낸다. span의 시작과 끝은 분석하고자 하는 시계열의 기간 내에 있어야 한다. span 기간을 series 명령의 modelspan 옵션에 의해 지정하면, 지정된 기간은 modelspan에서 지정한 기간 내에 있어야 한다. 시작은 끝 시점 보다 앞서야 한다. 결측치(span=(1976.jan,))가 있으면, 시계열의 시작이나 끝 시점이 default로 설정된다. 이때 default는 series 명령에서 span이 사용되었다면, span의 시작과 끝 시점은 series 명령에서 주어진 시작과 끝 시점으로 대체된다.

기타 명령

tcrate 일시적 변화 특이치의 감속 정도를 정의한다. 이 값은 0보다 크고 1보다 작아야 한다. Default 값은 tcrate= 0.7**(12/period)이다. 여기서 period는 1년의 관측치 개수이다(위는 월별 시계열인 경우이며, 분기별 시계열인 경우 4를 사용한다). 이 공식은 서로 다른 기간의 계열들이 전기간 동안 동일한 감속 정도를 가짐을 말한다. 이 명령이 regression spec에서 명시되었다면 outlier 명령에서는 사용할 필요가 없다.

참 고

첫 번째 관측치에 의해서는 시계열의 수준을 알 수 없으므로, 수준변화를 추정할 수 없다. 따라서 첫 번째 관측치에 대한 LS 검정통계량은 계산할 수 없다. 또한 마지막 관측치에 대한 LS는 AO와 구별할 수 없으며 두 번째 관측치에 대한 LS는 첫 번째 관측치에 대한 AO와 구별할 수 없다. 따라서 AO가 검색되었더라도 두 번째와 마지막 관측치에 대한 LS 통계량은 계산되지 않는다. 계산되지 않는 LS 통계량은 ‘0’으로 인쇄된다.

유사하게 마지막 관측치에 대한 일시적 변화(TC) 특이치도 AO와 구별할 수 없으므로 AO가 검색되었더라도 마지막 관측치에 대한 TC 통계량은 계산되지 않는다. 계산되지 않는 TC 통계량은 ‘0’으로 인쇄된다.

검색된 특이치를 RegARIMA 모형에 추가하기 위한 방법은 addone 방법과 addall 방법이 있다. addone 방법은 다음과 같다. 특이치 검색이 실시되는 각각의 시점에서 검색하고자 하는 특이치(AO, TC 및 LS)에 대한 t-통계량을 계산한다. 특이치에 대한 t-통계량의 절대값이 임계치보다 크다면, 특이치로 검색이 되며 이에 따라 해당 특이치의 회귀변수가 모형에 추가된다. 이때 추가된 회귀변수에 의해 새로운 모형이 추정되며, 이후 특이치를 다시 탐색한다. 이러한 과정은 후속의 특이치가 발견되지 않을 때까지 반복된다. 이때, 후향 제외(backward deletion)과정이 모형에서 “유의적이지 않은(절대 t-통계량이 임계치보다 크지 않은)” 특이치를 제외하기 위해 적용된다. 후향 제외과정은 모형속에 남아 있는 모든 특이치가 유의적일 때까지 반복한다. 후향 제외과정이 실행되는 동안 잔차 분산의

추정은 일반적인 방법(또는 non-robust)이 사용되며 이는 특이치 검색과정에서 얻어지는 것보다 다소 다른 특이치에 대한 t-통계량이 계산될 수도 있다.

addall 방법은 각각의 특이치 검색 시에 t-통계량의 절대값이 임계치보다 큰 모든 특이치를 모형에 더한다는 것 외에 addone 방법과 같은 단계를 실행한다. 이 방법에 의해서 더해진 특이치는 새로운 모형이 추정되었을 때 유의적이지 않을 수 있다. 따라서 addall 방법은 검색측면에서 모형에 더해진 불필요한 특이치를 제외하기 위한 후향 제외과정에 매우 크게 영향을 받는다.

addone과 addall 특이치를 검색하는 과정상의 차이로 최종적으로 검색된 특이치가 다른 것일 수 있다. 두 방법의 차이는 다음과 같다. 첫째, addone 방법은 일반적으로 addall보다 계산시간이 더 걸린다. 둘째, addall 방법은 검색과정에서 모형이 허락하는 최대 회귀변수의 수를 넘는 너무 많은 특이치를 뽑을 수 있다. 이 경우 프로그램에서는 이러한 현상에 대한 에러 메시지를 인쇄하고 실행을 멈춘다. 개선방법은 보다 적은 특이치를 검색하도록 cutoff 값을 올리거나 addone 방법을 사용한다. 두 방법 모두 각 검색에서 모든 가능한 시점의 특이치에 대한 t-통계량은 print=tests를 지정함으로써 인쇄할 수 있다.

임계치의 선정은 경험과 판단을 요구한다. Chang, Tiao와 Chen(1988)은 낮은 차수의 비계절성 ARIMA모형으로부터 생성된 시점 길이가 200인 시계열에 대한 모의 실험결과, AO 특이치의 검색에 민감하게 작용하기 위해서는 임계치를 3.0, 중간은 3.5, 낮은 경우는 4.0을 추천하고 있다.

특이치 검색은 regression과 arima 명령에서 기술된 모형과 추정된 ARIMA 모수로부터 시작한다. estimate 명령이 없다면, outlier 명령은 특이치 검색전의 default 추정옵션으로 모형을 추정한다.

어떠한 관측치가 특이치라고 의심이 간다면, regression 명령에서 적절한 AO 또는 LS 회귀변수를 더함으로써 모형 속에 포함할 수 있다.

특이치 검색 결과는 지정한 RegARIMA 모형에 따라 달라질 수 있다. 즉, 어떠한 관측치가 대부분의 다른 관측치보다 모형에 덜 적합되기 때문에 특이치로 분류될 수 있다. 그러므로 부적절한 RegARIMA 모형은 부적절한 특이치 조정을 유도할 수 있다.

예 제

(예 1) addone 방법과 default 옵션에 의한 임계치(특이치를 검색하고자하는 구간내의 관측치 수에 따르는 임계치)를 이용하여 전체 계열에 대한 AO와 LS 두 특이치에 대한 탐색을 동시에 한다. 모형 속에 존재하는 수준변화의 수가 2개나 그 이상이라면, 2, …, 5개의 각 연의 연속적인 수준변화가 임시 수준변화를 상쇄하는지를 검정하기 위한 t-통계량을 계산한다. estimate 명령이 없더라도 outlier 명령으로 모형을 추정한다.

```
series { title = "Monthly sales"    start=1976.jan
         date = (138 128 … 297) }
arima { model = (0 1 1) (0 1 1)12 }
outlier {lsrun = 5 types=(ao ls)}
```

(예 2) addall 방법과 임계치 t=4.0을 이용하여 AO 특이치를 찾는다. span 옵션이 series 명령에 있으므로 1980년 1월부터 1992년 12월까지에서 모형을 추정하고 특이치를 검색한다. regression 명령에서 기술된 2개의 수준변화는 lsrun이 ‘0’의 default값을 취하므로 임시 수준변화의 상쇄에 대한 검정을 하지 않는다.

```
series { title = "Monthly sales"    start = 1976.jan
         date = (138 128 … 297)
         span = (1980.jan, 1992.dec) }
regression { variables =(ls1981.jun  ls1990.nov)}
arima { model = (0 1 1) (0 1 1)12 }
estimate {  }
outlier {types = ao   method = addall   critical = 4.0}
```

(예 3) (예 2)와 같은 기간에서 모형을 추정하나 1987년과 1988년에 LS 특이치를 찾는다. default인 addone 방법을 사용하며, 임계치는 t=3.0이다. 연속적인 LS의 각 쌍은 임시 LS을 상쇄할 수 있는지를 검정한다.

```
series { title = "Monthly sales"     start =1976.jan
         date =(138 128 ··· 297)
         span =(1980.jan, 1992.dec) }
arima { model =(0 1 1) (0 1 1)12 }
estimate {  }
outlier { types =ls    critical =3.0
          lsrun =2     span =(1987.jan, 1988.dec)}
```

(예 4) (예 2)와 (예 3)과 같은 기간을 이용하여 추정하나 AO와 TC 그리고 LS의 특이치를 찾는다. default인 addone 방법을 이용하며 AO 특이치에 대한 임계치는 3.0, LS 특이치에 대한 임계치는 4.5, TC 특이치에 대한 임계치는 4.0이다.

```
series { title = "Monthly sales"    start = 1976.jan
         date = (138 128 ··· 297)
         span = (1980.jan, 1992.dec) }
arima { model = (0 1 1) (0 1 1)12 }
estimate {  }
outlier { critical = (3.0, 4.5, 4.0) }
```

PICKMDL

내 용

REGARIMA 모형의 ARIMA 부분은 X-11-ARIMA/88에서 사용된 방법과 유사하게 자동 모형 선정 절차에 의해 구할 수 있다. 사용자가 시계열에 적합할 ARIMA 모형을 지정할 수 있으며, 모형 선정 기준치를 조정할 수 있다.

일반형식

```
pickmdl { mode  =  both
           method = best
          file = "my.mdl"
          fcstlim = 25.0
          bcstlim = 25.0
          qlim = 15.0
          overdiff = 0.99
          identity = all
          outofsample = yes
          print = (none autochoice)
          savelog = auomodel
        }
```

명 령

bcstlim backcast가 **mode=both**로 지정되었을 때, 표본 내(within -sample) backcast 오차 검정의 채택 기준치를 설정한다. backcast 값의 절대평균퍼센트오차는 이 기준치에 의해서 검정된다. 예를 들어, **bcstlim=25**는 이 기준치를 25%로 설정한다. 이 옵션의 값은 0보다 크고 100보다는 작아야 한다. default는 **20** 이다.

fcstlim 표본 내(within-sample) forecast 오차 검정의 채택 기준치를 설정한다. 임의의 모형이 채택되었다면, 최근 3년 내 외삽값의 절대평균퍼센트오차는 이 값보다 작아야 한다. 예를 들어, **fcstlim=20**은 이 기준치를 20%로 설정한다. 이 옵션의 값은 0보다 크고 100보다는 작아야 한다. default는 **fcstlim=15** 이다.

file **pickmdl** 자동 모형 선정 절차에서 사용되는 ARIMA 모형을 포함하고 있는 파일이름과 경로를 지정한다. 모형은 **arima** 명령의 **model** 옵션에서 사용된 기호를 사용한다. 이 옵션은 반드시 사용하여야 하며, default는 없다.

identify 특이치(outlier 명령), 요일효과 변수(regression의 aictest 옵션)의 자동 식별을 어떻게 할 것인지를 결정한다. **identify=all**이면, 자동 모형 선정 화일에 있는 각각의 모형에 요일효과 변수와 특이치가 적용된다. **identify=first** 이면, 요일효과 변수와 특이치에 자동 식별은 자동 모형 파일의 첫 번째 모형에 적용된다. 지정된 첫 번째 모형에 대한 결정이 이루어 지면 나머지 모형들에 적용된다. 선정된 모형이 첫 번째 모형이 아니라면, 선정된 모형에 대해서 자동 식별 과정이 재실시된다. default는 **identify=first**

method **pickmdl** 자동 모형 선정 절차에서 모형 선정 기준을 충족하는 첫 번째 모형(**method=first**) 또는 선정 기준을 충족하는 모든 모형 중에서 가장 작은 표본 내 예측오차를 갖는 모형(**method=best**)을 사용할 것인지 지정한다. default는 **method=first.**

outofsample 자동 모형 선정과 평가를 위한 pickmdl의 forecast 오차 종류를 결정한다. **outofsample=yes**를 사용하면, 표본 외 forecast 오차를 사용한다. 즉, 1년의 forecast와 모형 추정을 위하여 사용된 시계열 자료에서 forecast 기간의 자료를 제거함으로써 얻어진다. 만일 **outofsample=no**를 사용하면, 표본 내 forecast 오차를 사용한다. 즉, 전계열을 이용하여 추정한 모형의 모수 추정치가 최근 3년의 자료를 forecast하는데 사용된다. default는 **outofsample=no.**

mode X-11-ARIMA에서 사용된 기준치를 이용하여, 사용자가 기술한 모형 후보군 내에서 최적의 모형을 찾는다. 만일 **mode=fcst**이라면, 선정된 적합 모형은 1년의 forecast를 하는데 이용된다. **mode=both**이라면, 1년의 forecast와 backcast를 하는데 이용된다. default는 **mode=fcst.**

print save 옵션은 사용할 수 없다. output에 사용할 수 있는 표는 다음과 같다. 모든 표는 default로 인쇄가 된다.

<PICKMDL: 이용가능한 출력표>

이 름	단축	내 용
pickmdlchoice	pmc	pickmdl 자동 모형 선정절차의 모형 선택
header	hdr	pickmdl 결과에 대한 header
usermodels	umd	pickmdl 자동 모형 선정에서 사용된 모형 결과

overdiff 과도차분을 검정하기 위한 기준치를 MA 모수 추정치의 합을 이용하여 정한다. 계절 MA 혹은 비계절 MA 모수 추정치의 합을 계산한다. 비계절 MA 모수 추정치의 합이 설정한 기준치보다 크다면, 과도차분 때문에 모형은 기각된다. 계절 MA 모수 추정치의 합이 설정한 기준치보다 크다면, 자동 모형 선정 절차는 **regression** 명령에 고정계절효과변수를 사용할 것을 제안하고 경고 메시지를 인쇄해 준다. 그러나 선정된 모형은 기각하지 않는다. 이 옵션에 사용되는 값은 0.9보다 크고 1보다는 작아야 한다. default는 0.9이다.

qlim Ljung-Box Q-통계량의 p-값에 대한 채택 기준치를 설정한다. 모형 선정을 위한 Q-통계량의 p-값은 pickmdl의 qlim에 의해 지정되는 기준치보다 크다. 예, qlim=10은 기준치에 대해서 10%이다. 이 옵션에 사용되는 값은 0보다 크고 100보다는 작다. default는 5%이다.

참 고

pickmdl 명령은 automdl, arima 명령이나 estimate 명령에 file 옵션을 사용하는 경우에는 사용할 수 없다.

pickmdl 명령에 의한 default 자동 모형 선정 기준은 다음과 같다.

① 최근 3년 동안의 외삽 값(extrapolated value 또는 forecast value)에 대한 절대평균퍼센트오차가 15%보다 작아야 한다.

② 모형의 잔차상관이 존재하는지를 검정하기 위하여, 적합된 모형에 대한 Ljung- Box Q-통계량(Portmanteau-통계량)을 구하며, 이때 Q-통계량의 p-값이 5%보다 커야 한다.

③ 과도차분의 징후가 없어야 한다. 비계절 MA 모수 값(적어도 1차 비계절 차분을 갖는 모형)의 합이 0.9보다 크다면 과도 차분의 징후가 있다고 할 수 있다. 모형화일에 있는 모형들을 받아들일 수 없다면 모형 선정을 하지 않는다. 이들 기준은 fcstlim, qlim 혹은 overdiff 옵션을 이용해서 변경할 수 있다.

X-11-ARIMA에서 적용하고 있는 자동 모형 선정을 위한 ARIMA 모형은 다음과 같다. 이들 모형은 regression 명령에서 고정 계절효과 변수를 가정하면 사용할 수 없다.

$$(0, 1, 1)(0, 1, 1)_s$$
$$(0, 1, 2)(0, 1, 1)_s$$
$$(2, 1, 0)(0, 1, 1)_s$$
$$(0, 2, 2)(0, 1, 1)_s$$
$$(2, 1, 2)(0, 1, 1)_s$$

file 옵션에 의해서 만들어진 파일의 각각 모형에 대한 구별은 “X"로 하며, default 모형은 “*”로 나타낸다. 그러나 마지막 모형은 ”X" 혹은 “*" 을 사용하지 않는다. default 모형은 자동 모형 선정 과정에서 ARIMA 모형이 선정되지 않았으나 사용자가 사전조정을 위하여 regression 명령에 회귀변수를 지정한 경우, default ARIMA 모형을 RegARIMA 모형에 적용하기 위하여 사용한다. 다음은 자동 모형 선정을 위한 모형 화일의 형태이다.

(0, 1, 1)(0, 1, 1) *
(0, 1, 2)(0, 1, 1) X
(2, 1, 0)(0, 1, 1) X
(0, 2, 2)(0, 1, 1) X
(2, 1, 2)(0, 1, 1)

예 제

(예 1) 자동 ARIMA 모형 선정 절차에 의하여 모형을 선정하고, 1년 후의 시계열을 예측한다. 모형에는 요일 및 고정 계절성 효과를 포함하며, ARIMA 모형군에 default 모형을 설정한다.

```
series {title = “Monthly Sales” start =1980.1
        data =( …… …… …… ) }
regression {variables = (td seasonal)}
pickmdl { mode = fcst  file="my.mdl“ }
estimate {        }
x11 {   }
```

이때 my.mdl 형태는 다음과 같다.

(1 1 0) X
(2 1 0) X
(0 1 1) *
(0 1 2) X
(2 1 2)

(예 2) (예 1)과 동일하나 모형 선정 기준의 예측오차 기준을 20%, Q-통계량의 χ^2 채택 기준을 10%, 과도차분 기준을 0.99로 바꾸며, 이들 기준에 만족하는 첫 번째 모형을 선정한다.

```
pickmdl { mode = fcst  file="my.mdl“
          method=first
          fcstlim=20    qlim=10    overdiff=0.99  }
estimate {        }
x11 {   }
```

(예 3)　(예 1)과 같이 같으며, 모형 식별과 선정에서 표본 외 예측 오차기준을 이용하지 않는다.

```
pickmdl { mode = fcst   file="my.mdl“
          outofsample="no"  }
```

REGRESSION

내 용

RegARIMA 모형에 제거하고자 하는 효과에 대한 회귀변수를 지정하기 위한 명령이다. 프로그램에 내장된 사전에 정의된 회귀변수는 variable 옵션에 의해 지정된다. 사전 정의된 회귀변수들은 상수항, 고정된 계절성, 요일효과(trading day effect)와 휴일효과(holiday effect), 가법 특이치, 수준변화, 일시적 변화(temporary change)와 ramp 등이 있다. 구조변화(change of regime)에 대한 회귀변수는 계절성과 요일효과 회귀변수 등으로 정의할 수 있다. 사용자가 정의한 회귀변수는 user 옵션으로 모형에 포함할 수 있으며, 회귀변수의 data 값은 data나 file 옵션으로 불러올 수 있다. 즉, regression 명령은 사전에 정의된 회귀변수와 사용자가 정의한 회귀변수를 포함할 수 있다.

일반형식

```
regression{ variables = (const
              seasonal or sincos[1, 2, 3]
          td or tdnolpyear or tdstock[31] or td1coef or td1nolpyear
              lom or loq
              lpyear
              easter[8]    labor[8]     thank[1]
              ao1967.apr   ls1972.sep   rp1965.nov-1968.may)
        print = (none)
        save = (rmx)
        user = (temp precip)
        usertype = holiday
        start = 1955.jan
        data = (25 0.1 … ) or file = "weather.dat"
                              format = "(2f5.1)"
        aictest = (easter
          td or tdnolpyear or tdstock or  td1coef or td1nolpyear)
        aicdiff = -2.0                  }
```

명　　령

aicdiff　aictest 옵션에 있는 회귀변수를 갖는 모형의 AIC(혹은 AICC)와 회귀변수가 없는 모형의 AIC의 차이를 지정한다. default는 aicdiff=0.0이다.

aictest　회귀변수가 RegARIMA 모형에 포함된 경우에 AIC-검정을 실시할 지 지정한다. 이 명령은 td, tdstock, td1coef, tdnolpyear, td1nolpyear, easter와 user를 지정하였을 때 사용할 수 있다. 예를 들어, 요일효과 모형 선정을 한다면, AIC 값(시계열 길이의 수정이 필요한 경우는 AICC)은 요일효과 변수를 포함한 모형과 포함하지 않은 모형에서 구한다. default로서, 보다 작은 AIC 값을 갖는 RegARIMA 모형이 시계열의 예측과 특이치 식별 등에 사용된다. 하나 이상의 회귀변수를 갖는 경우, 다음 순서로 AIC-검정이 순차적으로 실시된다: (a) 요일효과 변수 (b) 부활절효과 (c) 사용자 정의 변수. 같은 형태의 회귀변수가 여러 개 사용되면(예: 여러 개의 요일효과 변수), aictest 절차는 그룹으로 적용된다. 즉, 이 형태의 모든 변수가 최종모형에 적용되던가 또는 적용되지 않는다. 이 옵션을 지정하지 않으면, AIC를 기초로 한 자동 모형 선정은 실시되지 않는다.

file　사용자가 정의한 회귀변수의 자료 값을 포함하고 있는 파일 이름으로 따옴표 안에 명명한다. 만일 파일이 현재의 폴더가 아니면 경로를 지정해야 한다. data 옵션을 사용하면, 자료 값의 시점 영역은 분석하는 시계열의 미래 및 과거 예측치의 시점 영역을 포함해야 한다. file 옵션이 사용되면, data 옵션을 사용할 수 없다.

data　사용자가 정의한 회귀변수의 값을 지정한다. 자료의 시점 영역은 시계열의 시점 영역이나 series 명령의 span에 의해 지정된 시점 영역을 포함해야 한다. 또한 forecast 명령에서 요구되는 예측의 시점 영역도 포함해야 한다. 자료 값은 free format으로 user 옵션에서 사용자가 지정한 회귀변수 순서대로 읽는다. data 옵션과 file 옵션을 같이 사용할 수 없다.

format 회귀변수의 자료 값을 읽는 file 옵션에서 지정된 파일의 입력 형식이다. format은 처음과 끝은 괄호로 묶어 따옴표 속에 입력한다. format 명령을 사용하지 않으면 자료는 free format으로 읽혀진다. format 명령을 사용하면 data 명령을 사용할 수 없으며, file 옵션만 사용할 수 있다. 입력형식은 series 명령의 format 옵션의 형식과 동일하다.

start 사용자가 정의한 회귀변수의 자료 값들에 대한 시작 월(분기)이다. default는 시계열의 시작 월(분기)이다. 시계열의 시작 월(분기)을 사용하거나, series 명령의 span 옵션에 의해서 지정된 기간의 시작 월(분기)이 사용된다.

user 모형 속에 포함하고자 하는 사용자가 정의한 회귀변수의 이름을 지정한다. 주어진 이름은 프로그램 결과에서 추정된 계수를 식별하기 위하여 사용된다. data나 file 옵션에 의해 사용자가 정의한 변수에 대한 자료값을 제공해야 한다. 사용자가 정의할 수 있는 회귀변수는 최대 52개이다.

usertype 사용자가 정의한 회귀변수에 의해 추정된 회귀효과를 지정한다. 이 옵션은 X-12-ARIMA에서 제공하는 회귀효과변수 대신 사용하는 경우 유용하다.

사용자가 지정할 수 있는 회귀효과 형태는 다음과 같다;

상수(**constant**), 계절(**seasonal**), 요일효과(**td**), 재고 요일효과(**tdstock**), 월길이(**lom**), 분기길이(**loq**), 윤년(**lpyear**), 명절(**holiday, easter, labor** 등), 특이치 (**ao, ls, rp, tc**), 사용자 정의(**user**)

사용자가 정의한 모든 회귀변수에 대해 하나의 효과를 나타낸다면, 하나만 정의한다(usertype=td). 사용자가 정의한 모든 회귀변수에 대해 각각의 효과를 나타낸다면, 모든 회귀변수에 각각의 효과를 정의한다(usertype=(td td td holiday user). 이 옵션을 사용하지 않으면, 사용자 정의의 변수 형태는 user 옵션의 형태이다.

variables 모형에 포함할 사전에 정의된 회귀변수로 이들 변수에 대한 자료 값은 calender 함수와 같은 프로그램에 의해서 계산된다.

X-12-ARIMA에서 제공되는 사전정의 회귀변수

변 수	내 용
const	차분된 자료의 전체 평균이 '0'이 아닌 추세상수 회귀변수
seasonal	(s-1) 계절변수(s=계절주기)에 의한 고정된 계절효과. 변수는 월(분기)간 수준에서의 차분이 가능하지만, 전체수준에서는 영향력이 없다. 이 변수는 sincos 함수가 있거나, 계절차분을 한 모형에서는 사용할 수 없다.
sincos[ ]	계절빈도 $\omega_j=(2\pi_j/s)$ $(1\le j\le s/2)$에서 $\sin(w_j t)$와 $\cos(w_j t)$ 형태의 trigonometric 회귀변수이다. 빈도수는 지정되어야 한다. (예: 월별 시계열에서 sincos[1,2,3,4,5,6]은 모든 계절빈도수를 포함하지만, sincos[1,2,3]은 처음 세 개만을 포함한다.) 이 변수는 계절성이 있거나, 계절차분을 한 모형에서는 사용할 수 없다.
td	모형에 tdnolpyear를 포함하며 윤년효과는 transform 명령에 의해 변환을 한다. td는 tdnolpyear, td1coef, td1nolpyear, lom, loq, lpyear 혹은 tdstock[]과 함께 사용할 수 없다. td를 지정하면, transform 명령에서 adjust=lpyear 혹은 adjust=lom(loq)를 사용할 수 없다.
tdnolpyear	월(분기)별 계열에서 6개의 요일변수를 나타낸다. (월요일 수-일요일 수), … ,(토요일 수-일요일 수)이다. td, td1coef, td1nolpyear 혹은 tdstock과 같이 사용할 수 없다.
td1coef	모형에 td1coef 변수를 포함하며 윤년효과는 transform 명령에 의해 변환을 한다. td1coef는 td, tdnolpyear, td1nolpyear, lom, loq, lpyear 혹은 tdstock[]과 함께 사용할 수 없다. td1coef를 지정하면, transform 명령에서 adjust=lpyear 혹은 adjust=lom (loq)를 사용할 수 없다.
td1nolpyear	월(분기)별 계열에서 1개의 요일변수를 나타낸다. (주중요일 수)-2*(토・일요일 수)/5. td1nolpyear은 td, tdnolpyear, td1coef 혹은 tdstock과 같이 사용할 수 없다.
lpyear	윤년변수로, 2월(첫 분기) 윤년인 경우는 0.75, 2월(첫 분기) 비윤년인 경우는 -0.25, 이외는 0이다. lpyear는 td, td1coef 혹은 tdstock을 사용할 수 없다.
lom	회귀변수의 월길이이다. 분기계열은 loq를 사용한다. $s\neq12$(또는 4)이면 lom은 오류가 된다. lom은 td, td1coef 또는 tdstock과 같이 사용할 수 없다.

변 수	내 용
loq	회귀변수의 분기길이이다. 월계열은 lom을 사용한다. loq는 td, td1coef 또는 tdstock과 같이 사용할 수 없다.
tdstock[ω]	각 월의 ω^{th}날에 제공되는 stock 계열에 대한 주--일효과를 추정한다. ω는 1～31사이의 값을 입력한다. ω보다 임의의 월길이가 작으면, tdstock 변수는 월말을 이용한다. tdstock[31]은 월말 stock 계열에 사용한다. tdstock은 월별계열에서만 사용할 수 있으며, td, tdnolpyear, td1coef, td1nolpyear, lom 혹은 loq를 같이 사용할 수 없다.
easter[ω]	부활절 휴일회귀변수로 월(분기) 계열에서 사용가능하다. ω는 1～25 사이의 값을 입력한다.
labor[ω]	노동절 휴일회귀변수로 월별 계열에서만 사용가능하다. ω는 1～25 사이의 값을 입력한다.
thank[ω]	추수감사절 휴일회귀변수로 월별 계열에서만 사용가능하다. ω값은 -17～+8사이의 값을 입력한다. ω<0이면 추수감사절전 날수를 의미하고, ω>0이면 추수감사절 후의 날수를 의미한다.
ao*date*	가법특이치 회귀변수이다. 월별계열은 aoyear.month(예: ao1985.jul 또는 ao1985.7), 분기별 계열은 aoyear.quarter(예: ao1985.1 또는 ao1985.3), 연간계열은 aoyear(예: ao1985), 날짜가 없는 계열은 aoobservation number(예: 50번째 AO 관측치는 ao50)로 입력한다. 한 개 이상의 AO를 지정할 수 있다. 지정된 모든 특이치 날자는 계열 내에서 잇어야 한다. (AO는 계열 내에서 지정되지만, series 명령의 span에 의해 지정된 기간은 무시된다.)
ls*date*	수준변화 회귀변수이다. 예를 들면, 1990년 10월 초의 수준변화는 ls1990.oct 로 나타낸다. 1 이상의 수준변동을 지정해야 한다. 수준변동은 계열내에서 지정되지만, 계열의 시작일이나 series 명령의 span에 의해 지정된 기간을 지정할 수 없다.
rp*date-date*	시계열의 처음과 끝에서 발생하는 임시 ramp 효과 회귀변수이다. 예) rp1988.apr-1990.oct. 한 개의 ramp 효과를 지정할 수 있으며, 모든 ramp는 계열 내에 있어야 한다(만일 series 명령의 span 옵션에서 지정한 기간 밖에 있으면 무시된다.) ramp는 다른 ramp, AO, 수준변화(level-shif)와 겹칠 수 있다.

print & save　선택 가능한 출력표는 다음과 같다.

<Regression: 이용 가능한 출력표>

이　름	단축	저장	내　용
regressionmatrix	rmx	+	회귀변수와 관련된 값
aictest	ats	·	요일효과, 부활절과 사용자 정의의 회귀변수에 대한 AIC-검정 결과
outlier	otl	+	RegARIMA 특이치 요인
aoutlier	ao	+	RegARIMA 가법특이치 요인
levelshift	ls	+	RegARIMA 수준변화 및 ramp 요인
temporarychange	tc	+	RegARIMA 일시적 변화 특이치 요인
tradingday	td	+	RegARIMA 요일효과 요인
holiday	hol	+	RegARIMA 명절효과 요인
regseasonal	a10	+	RegARIMA 사용자 정의 계절요인
userdef	usr	+	사용자 정의 회귀변수에 대한 요인

기타명령

b　variables와 user 옵션에서 나타나는 순서대로 회귀모수의 초기 값을 지정한다. b 옵션은 RegARIMA 모형에 있는 모든 회귀 계수의 초기 값을 지정해야 하며 variables와 user 옵션 후에 있어야 한다. 모수에 부여하는 초기 값은 옵션에서 지정하거나 또는 다음 예처럼 결측값을 정확히 나타내주어야 한다. 결측값에 대한 default 값은 0.1이다.
두 개의 회귀계수를 갖는 경우, b=(0.7,)은 b=(0.7, 0.1)과 같으나 b=(0.7)과는 다르다. 세 개의 회귀계수를 갖는 경우, b=(0.8, ,-0.4)는 b=(0.8, 0.1, -0.4)과 같다. 특정한 값을 고정된 모수로 정하고자 하는 경우, 다음과 같이 "f"를 사용한다. b=(0.7f, 0.1).

noapply　x11 명령에 의해 계절조정단계를 실행하기 전에, 원계열에서 제거하지 않을 추정된 회귀효과요인을 정의한다. 사용할 수 있는 회귀효과변수는 다음과 같다;

td, holiday, ao, ls, tc, userseasonal, user

tcrate　tc 특이치 회귀변수의 감소 속도를 지정한다. 이 값은 0보다 크고 1보다는 작아야 한다. default 값은 tcrate= 0.7**(12/period), Period는 1년의 관측치 수이므로 월은 12개 계열, 분기는 4개 계열이다.

참　고

X-12-ARIMA는 예측을 실행하면, 전체 예측기간에 대해서 사전에 정의된 회귀변수를 선정하기 위한 자료 값이 만들어진다. 사용자는 52개까지 사전 정의된 회귀변수를 정의할 수 있으며, 회귀변수(개수 제한의 변경이 가능하다)를 포함하여 총 80개의 회귀변수를 사용할 수 있다. 또한 자동으로 검색된 특이치로부터 생성된 특이치 회귀변수까지 포함하여 사용자가 정의할 수 있는 회귀변수 계열의 최대길이는 예측기간을 제외하고 600이다.(이 제한 역시 변경이 가능하다)

variables 명령에서 const 옵션을 사용하면, 회귀변수의 결과는 차분된 시계열의 전체 평균을 따른다. ARIMA 모형이 차분을 포함하고 있지 않은 경우, 이것은 단순히 ‘0’이 아닌 전체 평균에 대한 상수 회귀항이 된다. 만일 ARIMA 모형이 차분한 시계열을 포함하고 있다면, 추세상수(trend constant)가 된다. 이때 생성된 실제 회귀변수는 차분 후에 1로 된 열을 생성한다.

월별(flow) 시계열에 요일효과가 존재한다면 variables 옵션에서 td 옵션을 사용할 것을 권장한다. 이 경우 윤년효과는 transform 명령에 의해서 좌우된다. 만일 시계열이 Box-Cox 또는 log 변환되었다면, 윤년효과(leap-year effect)는 사전조정에 의해서 제거된다. 즉, 시계열이 변환되기 전에 윤년요인인 lp_t에 의해서 조정된다.

$$lp_t = \begin{cases} \dfrac{28.25}{29} & t\ :\ \text{2월이 윤달(29일)인 경우} \\ \dfrac{28.25}{28} & t\ :\ \text{2월이 윤달이 아닌 경우} \\ 100 & \text{이외의 월인 경우} \end{cases}$$

시계열을 변환하지 않았다면, 윤년회귀변수 lpyear를 variables 옵션에

포함하며, 윤년회귀변수 LP_t는 다음과 같은 윤년회귀변수 값을 갖는다.

$$LP_t = \begin{cases} 29 - 28.25 & t\ :\ \text{2월 윤달(29일)인 경우} \\ 28 - 28.25 & t\ :\ \text{2월 윤달아닌 경우} \\ 0.00 & \text{이외의 월인 경우} \end{cases}$$

사용자가 변수 변환된 시계열에서 lom 회귀변수로 월 길이 효과를 조정하고자 한다면, 다음과 같이 lom과 tdnolpyear를 지정할 수 있다.

variables=(lom tdnolpyear)

사용자가 변수 변환되지 않은 시계열에서 월 길이 효과를 조정하고자 한다면, regression 명령에서 variables=(tdnolpyear …)와 transform 명령에서 adjust=lom를 지정한다. transform 명령에 adjust=lom를 지정하면, variables 옵션에 td나 lom를 사용하면 옵션간 충돌이 발생한다. 충돌을 피하기 위해서 transform 명령에서 adjust=lom을 제외하거나, variables에 있는 td 대신 tdnolpyear 를 사용하거나 lom를 제외하는 방법이 있다.

윤년효과를 regression 명령에 반영하기 위한 변수는 td, tdnolpyear, td1coef, td1nolpyear 등이 있다. 이중 td, td1coef는 윤년효과를 원계열에 대한 변수변환 시, 이들 효과를 반영한 변수이므로 lpyear나 lom을 사용할 수 없다. 이때 변수변환 결과는 X-12-ARIMA 결과의 A2표에 나타난다. tdnolpyear과 td1nolpyear은 요일효과만 포함하고 있는 변수이므로 regression 명령에 lpyear나 lom 변수를 사용할 수 있다. 이때 이들 변수에 대한 회귀계수가 추정된다. 그러나 transform 명령에서 adjust=lpyear 혹은 lom을 사용하면, 변수변환이 수행되며, 결과는 A2표에 나타난다.

분기별 flow 시계열에 대해서도 동일한 요일효과 옵션을 이용할 수 있다. 또한 lom을 loq로, period=12는 period=4로 대체하여 앞의 사항을 적용한다.

재고와 같은 stock계열에서, 요일효과는 월별 시계열에 대해서만 적용이 가능하다. tdstock[w]에서 w는 1～31의 범위를 가지며, 6개의 회귀변수를 생성한다. w값은 반드시 지정해야 하며, w는 stock을 결정하는 그 달의 날이나 월의 마지막 날을 나타낸다. 즉, tdstock[31]은 월말의 stock을 의미한다.

휴일효과(부활절, 노동절, 추수감사절 등)에 대한 회귀 변수는 flow계열에만 사용할 수 있다. 부활절 변수는 월별 또는 분기계열에 사용할 수 있으며, 노동절과 추수감사절변수는 월별계열에만 사용할 수 있다.

구조변화(change of regime)회귀변수는 계절성(seasonal), trigonometric seasonal(sincos), 요일(td, tdnolpyear 또는 tdstock), 윤년(lpyear), 월길이(lom) 또는 분기길이(loq) 등으로 지정할 수 있다. 구조변화 회귀변수는 두가지 기간 설정(전기간과 부분)방법이 있다.

다음 표와 같이 구조변화 회귀변수는 regression 명령의 variables 옵션에 회귀변수 이름에 하나 또는 두 개의 '/'를 사용하여 구조변화가 있는 시점을 나타낼 수 있다. 구조변화 기간은 구조변화가 발생한 이전 기간과 이후 기간으로 나눌 수 있다. 부분 구조변화 변수는 이 두 기간 중 하나에만 제약을 가하고 나머지 기간은 '0'으로 처리한다. 전기간 구조변화 변수는 구조변화 이전 기간에 대한 부분 구조변화 회귀와 기본적인 회귀를 이용한다. 예를 들어, 전기간 구조변화 표현 variables=(td/1990.jan/)은 variables=(td td/1990.jan//)과 같다. 이때 결합 요일효과 요인인 td, td/1990.jan/에 대한 회귀계수가 추정된다. 부분 구조변화 변수는 다음과 같다. variables=(td//1990.jan/)는 구조변화가 있는 1990년 1월 이후부터 td를 추정한다. 한편, variables=(td/1990.jan//)는 1990년 1월 이전의 td를 추정한다.

<구조변화에 대한 회귀변수와 형식>

형 태	형 식	예
전기간 구조변화	reg/date/	td/1990.jan/
부분 구조변화, 변화 이전기간 '0'	reg//date/	td//1990.jan/
부분 구조변화, 변화 이후기간 '0'	reg/date//	td/1990.jan//

부분 구조변화 변수에 대한 2가지 사용법이 있다. 첫째, 1990년 이후 시계열의 요일효과 추정을 위해서 variables=(td//1990.jan/)을 사용한다. 둘째, 계절 차분을 갖는 ARIMA 모형이 arima 명령에서 지정되었을 때 혹은 자동선정 명령인 automdl 명령에 의해 모형이 추정될 때, variables =(seasonal//1990.jan/)는 1990년 1월 이후 고정 계절요인 형태의 변화를 추정하고 통계적으로 유의한지 검정한다.

aictest는 variables 옵션에 있는 회귀변수들을 제외하거나 회귀변수를 추가하는데 이용할 수 있다. 양수의 aicdiff 값은 aictest 검정에서 검정하는 변수를 RegARIMA 모형에 포함하는 것을 어렵게 만든다. Δ_{AICC}를 aicdiff 옵션과 관련된 값, $AICC^{with}$(혹은$AICC^{without}$)를 aictest 옵션에서 지

정한 회귀변수를 RegARIMA 모형에 포함(비포함)했을 때의 AICC 값이라고 하자. 만일 variables 옵션에서 회귀변수들을 지정하지 않았더라도 aictest 옵션에서 회귀변수를 지정하여 검정하였다면, 다음과 같은 경우에 이들 회귀변수는 회귀모형에 포함된다.

$$AICC^{with} + \Delta_{AIC} < AICC^{without}$$

aictest=(tdstock) 사용하면, tdstock[w]의 default 값은 31이므로 stock 변수 tdstock[31]은 월말이 지정된다.

요일효과 회귀변수가 aictest와 variables 명령에 지정되었을 때, 지정된 회귀변수 형태가 일치해야 한다. stock 요일변수에 대한 표본 월(분기)(sample day)과 구조변화 회귀변수에 지정된 월(분기)은 aictest 옵션에 포함하지 않는다. 예를 들어 variables=(tdstock[15] ao1995.jan)이라면, aictest에는 tdstock을 사용해야 한다.

사용자가 정의한 변수는 계절성이 제거된 형태로 이용되어야 한다. 만일 계절성이 제거되지 않은 상태에서 이용하면, 계절요인에는 추정된 모든 계절효과를 포함하지 못하게 된다. 사용자가 정의한 변수 형태를 usertype 옵션에서 지정하면, 사용자가 정의한 변수로부터 유도한 요인은 regression 명령에서 지정한 동일한 형태의 변수에서 유도한 회귀요인과 결합하게 된다. noapply 옵션에서 이들 회귀형태의 이름을 지정하지 않았다면, 이 결과의 요인은 시계열에 적용된다.

요일효과와 명절효과 변수는 noapply 옵션에 의해 regression이나 X11regression에 의해 추정된 효과를, 시계열에 적용하는데 사용하지 않는다고 지정하지 않는 한 regression과 X11regression에서 동시에 사용하지 않는다.

계절성이 없는 시계열에 대한 요일효과와 명절효과의 추정은 x11 명령의 type=trend를 사용하여 할 수 있다.

예 제

(예 1) 고정된 계절효과, (선형)추세상수와 ARIMA(0 1 1) 오차를 갖는 모형을 식별하고 추정하는 프로그램이다.

```
series {title = “Monthly sales”    start = 1976.jan
         data =(138 128  ...  297)}
regression { variables=(constant seasonal) }
arima {model = (0 1 1)}
estimate {          }
```

(예 2) 계절 빈도(seasonal frequencies)가 sine, cosine 함수로 결정되는 고정계절효과 회귀변수를 갖는 모형을 식별하고 추정하는 프로그램이다. (예 1)처럼 선형추세상수를 포함한다.
ARIMA (0 1 1)(0 1 1)$_{12}$모형은 회귀오차계열에 사용된다.
즉, 완전한 모형은 $(1-\Phi B^{12})(1-B)(y_t-\sum\beta_i x_{it})=(1-\Theta B)a_t$
이다. (x_{it} : sine, cosine 변수, t : 추세상수 변수)

```
series {title = “Monthly sales”    start = 1976.jan
         data =(138 128  ...  297)}
regression { variables=(const sincos[1,3,4,5]) }
arima { model = (0 1 1)(1 0 0)  }
estimate {          }
```

(예 3) 월별 시계열자료에 요일효과, 부활절, 노동절과 추수감사절효과를 적용한다. 지정한 기간은 휴일효과 기간이다. td를 지정하고 log 변환과 원계열을 윤년요인에 의해 나누고, tdnolpyear 회귀변수는 변수 변환된 계열에 적합시킨다. 회귀계수는 차분된 회귀변수에 대응되는 최대 차분된 계열의 회귀변수에 의해 identify 명령으로 추정된다.

```
series {title = “Monthly sales”    start = 1976.jan
         data =(138 128  ...  297)}
transform { function = log }
regression { variables = (td easter[8]  labor[10] thank[3]) }
identify { diff = (0 1)  sdiff = (0 1) }
```

(예 4) (예 3)과 같은 회귀효과를 갖고 있으며 윤년효과를 반영하도록 lom 회귀변수를 포함하는 경우의 모형이다. 회귀잔차계열은 (0 1 1)(0 1 1)12의 모형을 따른다.

```
series {title = "Monthly sales"   start = 1976.jan
        data =(138 128  ...  297)}
transform { function = log }
regression { variables  = (td6 lom easter[8]
                          labor[10] thank[3]) }
identify { diff = (0 1)  sdiff = (0 1) }
```

(예 5) 분기 시계열자료에 대해서 요일효과, 두 번의 AO와 LS가 있는 모형을 추정하는 프로그램이다. 이들 효과를 제외하고, 변환된 시계열은 ARIMA(0 1 1)(0 1 1)4모형을 따른다.

```
series {title = "Quarterly sales"  start = 1963.1
        period = 4
        data = (                    )}
transform { function =log }
regression {variables =(ao1967.1 ls1985.3 ls1987.2 ao1978.1 td)}
arima { model = (0 1 1)(0 1 1) }
estimate {  }
```

(예 6) 1985년 3분기부터 1987년 1분기까지 일시적인 수준변화가 있으며, 사용자가 정의한 회귀변수를 사용하는 프로그램이다. ARIMA 모형 식별을 위한 ACF와 PACF 계산을 하기 전에 발생하는 일시적인 수준변화(level shifts)는 identify 명령을 수행하여 제거한다. 수준변화 변수는 tls이며, 자료 값은 regression 명령에서 data 옵션에 의해 읽는다.

```
series {title = "Quarterly sales" start =1981.1
        data =(                    ) period=4 }
regression { user = tls
             data = (0 0 0 0 0 0 ……
                     1 1 1 1 1 1 1 1 1 1 1 0 0 ……0 ) }
 identify { diff = (0 1) sdiff = (0 1) }
```

(예 7) 상수, 고정된 계절효과와 2개의 사용자 정의 회귀변수를 포함하는 모형을 추정하는 프로그램이다. 사용자가 정의한 회귀변수 값은 weather.dat로 1960년 1월부터 입력되어 있다. 이 파일 속에는 모형에서 사용하지 않는 다른 변수들도 포함되어 있으며, 사용자가 정의한 회귀변수는 FORTRAN format를 이용하여 16 칼럼을 띄우고 읽는다. 사용자가 정의한 회귀변수의 시작 월이 모형화하고자 하는 시계열의 월(1970년 1월)보다 앞서므로 회귀변수의 시작 월을 지정해야 한다.

```
series {title = “Monthly Riverflow”    start = 1970.1
        data = (                    )  period = 12 }
regression { variables = (seasonal const)
             user = (temp precip)
             file = "weather.dat"
             format = "(t17,2f8.2)"
             start = 1960.1 }
arima { model = (3 0 0)(0 0 0) }
estimate {   }
```

(예 8) 하나의 AO와 요일효과를 가진 월말 stock 계열의 월별 소매업재고지수 모형을 추정하는 프로그램이다. 회귀효과가 제외된 변수 변환된 계열은 ARIMA(0 1 0)(0 1 1)$_{12}$모형을 따른다. stock 요일효과를 검정하기 위하여 AIC 검정을 실시한다.

```
series {title = “Retail Inventory - Family Apparel”
        start = 1967.1    period = 12
        data = (                    )  }
transform { function = log }
regression { variables = (tdstock[31] ao1980.jul )
             aictd = tdstock[31] )}
arima { model = (0 1 0)(0 1 1) }
estimate {   }
```

(예 9) 안정된 계절성과 요일효과를 갖는 월별 소매판매액지수에 대한 모형을 추정하는 프로그램이다. 두 회귀변수에 대해 1985년 12월에 구조변화가 있었다고 가정한다. 지수는 변수 변환이 되었으며, ARIMA(0 1 1) 모형을 따른다.

```
series {title = "Retail Inventory - Televisions"
        start = 1967.1     period = 12
        file = "tvsales.ori" }
transform { function = log }
regression { variables = (td/1985.dec/seasonal/1985.dec/)}
arima { model = (0 1 1) }
estimate {    }
```

SERIES

내　용

X-12-ARIMA의 series에서 요구되는 명령은 시계열 명, 시작 시점, 계절주기(월, 분기), 분석기간(span; 전체 시계열 의 일부분) 등이다. 시계열 자료는 series 명령의 data 옵션에 의해서 직접 읽어 들이거나, file 옵션에 의해서 파일 형태의 자료를 읽어 들일 수 있다. Single-mode(2.1절 참조)와 같이 data metafile의 자료를 이용하였다면, 시계열이 metafile에 기술되어 있으므로 시계열을 series 명령에서 기술할 필요가 없다.

일반형식

series{title = "Example Series"
  start = 1967.1
  period = 12
  span = (1968.1, 1989.12)
  name = "tstsrs"
  data = (480 ··· 1386) *or* **file** = "example.dat"
        format = "2r"
  decimals = 2
  precision = 1
  comptype = add
  compwt = 1.0
  print = (none+header)
  save = (spn)
}

명　령

start　시계열의 시작 년도와 계절주기(월, 분기)를 나타낸다.
default는 1.1이다.　*예) start=year.season period*

comptype 총합계열(또는 합성계열)이 어떻게 결합되는지 나타낸다. 총합계열은 가법(comptype=add), 감법(comptype=sub), 승법(comptype=mult), 제법(comptype=div)으로 각 구성계열을 결합하여 만들 수 있다. default는 결합하지 않는 것이다(comptype=none).

compwt 총합계열을 만들기 위한 입력 구성계열의 가중치를 기술한다. 이 명령은 comptype과 함께 사용한다. default는 1이다. *예) compwt=0.5*

decimals 계절조정 결과를 출력할 때 소수점 자릿수를 지정하며, 0과 5사이의 정수이다. default는 0이다. *예) decimals=5*

data 시계열 자료로, 열(row) 형태로 읽으며, 한 개 이상의 공백과 콤마가 있어야 한다. 관측치 수는 시계열 길이에 의해 결정된다. data와 file 옵션을 동시에 사용할 수 없으므로, data 옵션이 사용되면 file 옵션은 사용할 수 없다.

file 시계열 자료를 포함하고 있는 화일 이름으로, 따옴표 안에 표시한다. 화일이 현재의 디렉토리에 없으면 화일 경로를 포함한 완전한 화일 이름을 입력해야 한다.

format 지정된 화일로 부터 시계열 자료를 읽는데 사용한다. format은 file 옵션이 있을 때만 사용하며, 다음과 같이 여러 가지 형태로 입력할 수 있다.

(a) Free format, 한 줄에 있는 모든 숫자를 읽으며, 숫자는 하나 또는 그 이상의 space로 구분하며, “,” 또는 tab을 사용할 수 없다. *예) format="free"*

(b) FORTRAN format인 경우 양쪽에 괄호와 따옴표가 있어야 한다. *예) format="(6f12.0)"*

(c) X-11과 X-11-ARIMA에서 사용된 자료형태인 경우에는 2글자 코드를 사용한다. *예) format="1r"*

(d) “편집-날짜(edit-date)”형식. 년, 월(분기), 관측치의 값 순으로 format한다. 즉, 자료의 1991년 7월의 수치가 32531이면, 입력 형태는 1991 7 32531이다.
예) format=“datevalue"

(e) TRAMO와 SEATS에서 사용되는 입력형태를 이용할 수 있다. *예) format="tramo"*

(f) X-12-ARIMA에 의해서 미리 저장된 표를 읽어 들일 수 있다. *예) format="x12save"*

(g) 숫자들이 하나 혹은 그 이상의 공란으로 구별되고 소숫점은 콤마를 사용한다. *예) format="freecomma"*

(c)의 X-11-ARIMA 자료형태는 자료의 연도와 계열 값이6 또는 12자로 저장된다. free format에서는 한 줄에 있는 숫자를 모두 읽은 다음에 다음 줄을 읽는다. format 옵션이 없으면 자료는 free format으로 읽혀진다. Format 옵션은 data 옵션과 함께 사용할 수 없으며, file 옵션과 함께 사용한다.

modelspan REGARIMA 모형의 모든 계수를 결정하기 위하여 사용되는 자료 분석 기간을 지정한다. 사용자가 예측에 영향을 주는 시계열 초기 기간의 자료를 원하지 않거나, 사전조정을 위해 사용되는 회귀모형의 추정치에 영향을 주는 마지막 기간의 자료를 원하지 않을 때 사용한다. modelspan은 기간의 시작과 끝 2개 값을 가지며, 결측치가 있으면 분석하고자 하는 시계열의 시작 또는 마지막의 월(분기)이 지정된다. 월별 자료의 경우, REGARIMA 모형을 1968. 1월 부터 시계열의 마지막 기간까지 이용하고자 한다면, modelspan=(1968.1,) 으로 기술하며, 여기서 콤마(,)는 반드시 있어야 한다. 모형기간의 분석기간은 series 옵션에서 기술한 기간 내에 있어야 하며, 시작 월(분기)는 마지막 월(분기)보다 선행해야 한다. 한편, 마지막 기간의 자료는 "0.per"를 이용한다. 여기서 per는 월 또는 분기를 나타낸다. 즉 modelspan=(, 0.dec)는 REGARIMA 모형을 추정하기 위한 마지막 시점은 12월을 나타낸다.

name 시계열의 이름으로 64자 이내에서 따옴표 안에 표시한다. format에 의해 지정되었을 때, 이름의 첫 6자(또는 8자)는 올바른 시계열을 읽는지 확인하거나 혹은 여러 개의 시계열을 포함하고 있는 화일에서 특별한 계열을 찾을 때 이용한다.

period 시계열의 계절주기이다. 월(12), 분기(4)계열이 있으며, default는 12이다.

precision　입력 자료의 소수점 자릿수를 의미하며, format 옵션이 있을 때만 사용 가능하다. 값은 0과 5사이의 정수이며, default는 0이다.　*예) precision=5*

print & save　선택 가능한 출력표는 다음과 같다.

<Series: 이용 가능한 출력표>

이　름	단축	저장	내　용
header	hdr	·	X-12-ARIMA 실행에 사용된 옵션
span	a1	+	시계열자료 인쇄 또는 저장(span명령이 있는 경우 지정된 기간의 자료)
seriesplot	a1p	·	원계열의 plot
specfile	spc	·	실행에 사용된 입력화일 내용
savefile	sav	·	X-12-ARIMA 실행으로 만들어진 화일목록
specorig	sp0	+	원계열을 일차차분한 계열의 spectral plot
calendaradjorig	a18	+	RegARIMA 캘런더효과를 조정한 원계열
outlieradjorig	a19	+	RegARIMA 특이치를 조정한 원계열
adjoriginal	b1	+	원계열, 사전조정된 계열, 예측 확대된 계열

span　분석하고자 하는 시계열 기간의 처음 월(분기)과 마지막 월(분기) 두 개를 입력하며, 결측치가 있으면 분석하고자 하는 시계열의 처음 또는 마지막 월(분기)이 사용된다. 예를 들어 월별자료에서 span=(1968.1,)이면, 1968년 1월 부터 data 또는 file 옵션에서 기술한 시계열의 마지막 월이 지정된다. 여기서 콤마는 필수적이며, span에서 지정된 기간의 처음과 마지막 월(분기)은 입력 시계열 내의 기간이어야 하고, 시작 월(분기)은 마지막 월(분기)보다 앞서야 한다.

spectrumstart　원계열, 계절조정계열, 불규칙계열에 대한 spectrum 분석을 할 때, 분석 시작 월(분기)을 지정한다. spectrumstart= *year.seasonal period*로 표현한다. 자료의 마지막 8년의 자료로 조정된 결과에 요일효과 잔차 또는 계절효과가 있는지를 결정하기 위하여 사용할 수 있으며, 계절 또는 요일의 패턴을 계산할 때 이용한다. 월별자료의 스펙트럼 분석을 위한 시작점이 default의 경우, 최근 8년의 자료를 이용한다. 분기의 경우는 자료의 분석기간의 시작기간과 같다.
예) spectrumstart=1987.Jan

title 시계열 분석의 제목으로 79자 이내로 따옴표 안에 들어가야 한다. '-p' 옵션을 사용하지 않으면, 출력 시 각 쪽마다 title이 인쇄된다.

기타명령

diffspectrum diffspectrum=no이면, (변환된) 원계열 혹은 계절조정계열의 스펙트럼이 계산된다. default(diffspectrum=yes)는 이들 시계열 전월(분기)비 차이의 스펙트럼을 계산한다.

yr2000 yr2000=yes이면, X-11 형식으로 저장된 자료의 연도에 대한 2자리 절삭은 1945에서 이루어진다. 00-45년은 20xx년, 46-99년은 19xx년으로 인식한다. yr2000=no이면, 모든 연도를 20세기로 인식한다. default는 yr2000=yes이다.

참　고

X-12-ARIMA는 최대 600개 데이터를 읽을 수 있다. series 명령은 composite 명령과 함께 사용할 수 없다. composite 명령은 총합계열을 계절조정하는 것을 의미하므로 series 명령에서는 composite 옵션을 사용할 수 없다.

아래 표는 FORTRAN 형식 뿐만 아니라 format 옵션의 X-11 format code를 나타낸다. 아래의 형식들은 decimals 옵션을 사용하여 수정할 수 있다.

X-11의 Default Formats

code	월 별 자 료	분 기 자 료
1r	(12f6.0,I2,A6)	(4(12x,f6.0),I2,A6)
2r	(6f12.0,/,6f12.0,I2,A6)	(4f12.0,24x,I2,A6)
1l	(A6,I2,12f6.0)	(A6,I2,4(12x,f6.0))
2l	(A6,I2,6f12.0,/,6f12.0)	(A6,I2,4f12.0)
cs	(A8,I2,10X,12E16.10,18X)	(A8,I2,10X,12E16.10,18X)

주) CS는 캐나다 DB인 CANSIM의 format이다.

X-11 format code가 지정되었거나 혹은 format="datevalue", format ="tramo", format="x12save"이 사용하였다면, 시작연도는 데이터 화일에서 자동으로 읽히므로 series 명령에서 start 옵션은 필요 없다. tramo 입력형식은 (예 7)를 참고할 수 있다.

X-11 format code를 사용하지 않으면 decimals 옵션은 무시된다.

(12f6.0,I2,a6)의 자료형태는 다음과 같다.

자 료 값 column 수 1 6 12 ……………… 72											연도 74	계 열 명 80
1654	1712	1784	1797	……	1880	1869	1891	1906	1915		68	ICMETI
1938	1881	1916	1912	……	1973	2085	2103	2133	2185		69	ICMETI
2202	2235	2218	2242	……	2248	2251	2203	2160	2137		70	ICMETI

(6f12.0,/,6f12.0,i2,a6)의 자료형태는 다음과 같다.

자 료 값 column 수 1 12 24 ……………… 72					연도 74	계열명 80
1654	1712	………………	1857	1865		
1855	1880	………………	1906	1915	68	ICMETI
1938	1881	………………	1979	1972		
1931	1973	………………	2133	2185	69	ICMETI
2202	2235	………………	2258	2294		
2256	2248	………………	2160	2137	70	ICMETI

하나의 입력 spc 화일과 일련의 자료 화일들을 data metafile로 읽어 들이는 경우(single-spc mode), X-11 format의 사용은 피해야 한다. data metafile 내의 개별 data 파일들은 같은 시계열 명을 사용하나, X-11 format은 자료가 있는 화일 내의 시계열 명을 사용하도록 되어 있어 같은 자료 화일일지라도 동일한 시계열 명으로 인식하지 않을 수 있다.

format되어 있는 자료 화일을 읽을 때, 시계열의 끝 관측치(시작 값과 마지막 값)가 0이라면 (trimzero=no)가 아닌 이상 X-12-ARIMA는 이 값을 사용하지 않는다. 그러나 format된 형식의 처음 또는 마지막 관측치가 실제로 0이라면 문제가 생긴다. 이러한 경우에는 trimzero=no를 사용한다.

예　제

(예 1)　1940~1993년 분기 시계열에서 시작(1940～1945년)과 끝(1991～1993년) 양쪽의 관측치를 제외하고 분석한다. 즉, 자료의 최초 6년을 분석기간에서 제외하고, '91～'93년은 예측정도를 검증하기 위한 표본외(out-of-sample)로 분석에서 보류한다.

```
series{data = (879 899 985 … )
       start = 1940.1            # ending in 1993.4
       period = 4                # Quarterly series
       span = (1946.1, 1990.4)   }
```

(예 2)　series 명령에 의해서 직접 시계열의 data를 읽는 경우

```
series{title = "A Simple Example"
       start = 1967.jan              #period defaults to 12
       data = (480 467 514 505 534 546 539 541 551 537 584 854
               …… …… …… …… …… …… …… …… …… ) }
```

(예 3)　X-11-ARIMA의 format형태(12f6.1,i2,a6)로 저장된 1971년 7월～1993년 2월 자료(c:\data\sales1.dat)를 X-12-ARIMA에서 읽어들인다. X-12-ARIMA의 형식은 ("1r")로 읽으며, 마지막 13열은 연도와 계열명이다

<sales1.dat의 입력된 형태>

```
                                    146.4 109.2 132.1 144.8 116.1 100.370sales1
142.9 158.8 196.2 244.0 251.6 245.5 244.2 213.8 188.9 197.2 181.2 161.371sales1
  …… …… …… …… …… …… …… …… …… …… …… ……
148.8 177.2   0     0     0     0     0     0     0     0     0     93sales1
```

```
series { title="Monthly data in an X-11 format"
         period = 12
         file = "c:\data\sales1.dat"
         precision =  1      format = "1r"  }
```

FORTRAN format에서는 앞의 6개월의 공백과 마지막 10개월의 0을 모두 '0'으로 취급하지만, X-12-ARIMA에서는 '0'은 무시하고 해당 월의 실제 자료만 읽는다. 해당 년도가 각 줄에 포함되어 있으므로 start 옵션은 사용하지 않아도 된다.

(예 4) X-12-ARIMA로 자료화일의 입력형태를 수정하는 예이다. 자료는 1980년 2월~1990년 11월까지 (6f4.0, 1x, i4)형태로 다음과 같이 저장되어 있다.

<입력된 자료형태>

```
column: 1                        24   26
             0 342-256 491    0  0001
      -234 922-111    2    0 199 0002
      ...... ...... ...... ...... ...... ...... ...
       581-987-423   10    0    0 0022
```

이 자료는 여러 개의 데이터가 구분없이 입력되어 있으며, 26~29번째 열에는 기록란(0001, 0002…)이 있어 free format으로 읽을 수 없다. REGARIMA 모형에서는 시작과 마지막 관측치가 0이면 무시하므로 ’80년 2월을 342로 읽는다. 따라서 기록 난을 지우고 개별 자료를 분리하여 수정한 후, free format으로 자료 화일을 읽는다.

<수정된 자료: exam1.dat로 저장>

```
         0    342  -256    491      0
 -234   922   -111    2      0    199
 ... ...... ...... ...... ...... ...... ...
  581  -987   -423   10      0      0
```

수정된 자료를 읽기 위한 명령은 다음과 같다.

```
series{title ="Data read correctly in free format"
       start = 1980.2   period = 12
       file = "exam1.dat" }  #file is in current directory
```

(예 5) 다음은 datevalue format방식으로 sales1.dat에 저장되어 있는 1970년 7월~1993년 2월의 자료이다.

<sales1.dat의 자료 형태>

```
1970   7   14624
1970   8   10952
1970   9   13251
       :
1993   1   14838
1993   2   17762
```

자료입력 명령은 다음과 같다.

```
series{title="Monthly data in an edit date format"
        period = 12
        file = "c:\data\sales1.dat"   # a DOS path and file
        format = "datevalue" }
```

(예 6) (예 5)의 자료를 구성계열로 하여 합성계열을 만드는 프로그램이다. 합성 기간은 1992년 12월까지이며, 소수 2자리로 출력한다.

```
series{title="Monthly data in an edit date format"
        period = 12
        file = "c:\data\sales1.edt"   # a DOS path and file
        format = "datevalue"
        comptype = add       decimals = 2
        modelspan = (,1992.dec)  }
```

(예 7) 다음 예는 시계열(명: TEST)이 tramo format방식으로 sales1.edt에 저장되어 있으며, 기간은 1970년 7월~1993년 2월이다.

<ramo format 형태>	<sales1.edt의 자료 형태>
시계열 명	TEST
n year mon peri	272 1970 7 12
	14624
여기서 n: 관측치 수	10952
year: 시작년도	:
mon: 시작 월	14838
peri: 주기	17762

자료입력 명령은 다음과 같다.

```
series{title="Monthly data in an edit date format"
        period = 12
        file = "c:\data\sales1.edt"   # a DOS path and file
        format = “tramo" }
```

SLIDINGSPANS

내 용

계절조정의 안정성을 분석하기 위한 이동기간(sliding span) 분석에 관한 명령이다. 시계열 자료의 부분중첩기간(overlapping subspan)들을 계절조정한 후 이들 결과에 대한 차이를 비교한다. 사용자는 이동기간 비교분석을 위한 시작점(start), 이동기간의 길이(length), 이동기간분석의 통계량에 대한 판정 기준값(cutsf, cuttd, cutchng), 각 이동기간에 대해서 REGARIMA 모형의 모수값을 계절조정을 하는 동안에 어떻게 구할 것인가(fixmdl), REGARIMA 모형의 특이치 식별을 자동으로 할 것인지(outlier)를 지정할 수 있다.

일반형식

```
slidingspans { start = 1975.jan
               length = 132
               cutchng = 3.0
               cutsf = 3.0
               cuttd = 2.0
               outlier = yes
               fixmdl = no
               fixreg = outlier
               print = (long - ssheader)
               save = (sfspan)
               }
```

명 령

cutchng 계절조정계열의 월(분기)간 혹은 연간 변화율에 대한 기준치이다. 한 기간 이상의 공통기간을 갖는 월(분기)자료에 대하여, 다른 기간의 기간 간 변화율의 최대 절대차이가 기준치보다 크면, 해당 월(분기)은 기간 변화에 대해서 신뢰할 수 없는 추정치로서 판정한다. cutchng 값은 0보다 커야 하며, default 값은 3.0이다. *예) cutchng=5.0*

cutseas 계절요인과 계절조정계열에 대한 기준치이다. 한 기간 이상의 공통기간을 갖는 월(분기)자료에 대하여, 다른 기간에 서 추정한 계절요인과 계절조정계열의 최대 절대 퍼센트 변화가 기준치 보다 크다면, 해당 월(분기)의 계절요인과 계절조정계열은 신뢰할 수 없다. cutseas 값은 0보다 커야 하며, default 값은 3.0이다. *예) cutseas=5.0*

cuttd 요일요인에 대한 기준치이다. 한 기간 이상의 공통기간을 갖는 월(분기)자료에 대하여, 다른 기간으로부터 추정된 요일요인의 최대 절대변화율이 기준치보다 크다면, 월의 요일요인은 신뢰할 수 없다. cuttd 값은 0보다 커야 하며, default 값은 2.0이다. *예) cuttd=1.0*

fixmdl 이동기간을 계절조정하기 전에 REGARIMA 모형에서 추정된 모수에 대한 초기 값을 어떻게 설정할 것인가를 지정한다. 이 옵션은 REGARIMA 모형이 적합되지 않으면 무시된다. fixmdl=yes라면, 각 기간에 대한 REGARIMA 모형의 모수 값은 원래의 REGARIMA 모형 추정으로부터 얻어진 모수 추정치가 설정된다. 이 모수는 고정된 값이 되며, 재추정되지 않는다. default는 fixmdl=yes이다.
fixmdl=no이면, REGARIMA 모형 추정이 완전한 계열에 대해서 이루어 질 때, 초기 값이 원래의 값으로 된다. 초기 값이 estimate 명령에서 고정되었다면, 같은 값으로 남는다. fixmdl=clear이면, 각 기간에 대한 초기 값은 모든 계수가 0.1인 default로 되며, 모형의 모든 모수는 재추정된다.

fixreg RegARIMA 모형이나 불규칙요인 회귀식의 회귀계수들을 고정시키기 위하여 지정한다. 이들 계수들은 series 혹은 composite 명령에서 지정한 모형 기간으로부터 얻어진 값에 대해서 고정한다. 요일효과(td), 휴일효과(holiday), 특이치(outlier), 사용자 정의(user) 회귀효과에 사용된다. RegARIMA 모형이나 불규칙요인 회귀식이 시계열에 적합이 되면 이 옵션은 무시된다. *예) fixreg=td*

length 이동기간 결과를 산출하기 위한 월 또는 분기의 각 기간 길이이다. 사용자에 의해서 선택된 길이는 3년보다 크고 17년보다 작은 기간이여야 한다. 기간의 길이를 사용자가 지정하지 않으면, 사용자가 선정한 계절 filter의 길이를 근거로 기간의 길이를 결정한다. 그러나 사용자가 계절 filter의 길이를 선정하지 않은 경우에는 프로그램이 선정한다. 월별자료의 경우 length=96 이다. *예) length=96*

numspans 비교를 위한 결과를 만들기 위한 이동기간의 수이다. 이동기간의 수는 2~4개가 선정되어야 한다. 사용자에 의해 이 옵션이 지정되지 않으면, 프로그램은 사용자에 의해 주어진 이동기간 길이에 의해 최대 기간 수(4개 까지)를 선정할 수 있다. *예) numspans=4*

outlier 각 기간에 대해서 REGARIMA 모형이 재추정될 때 자동 특이치 탐색을 실시하도록 한다. 이 옵션은 outlier 명령이 사용되지 않으면 의미가 없다. outlier=keep이면, 전체 기간에 대한 RegARIMA 모형의 원래 추정에서 자동 식별된 특이치가 적용되며, 이동기간의 RegARIMA 모형에서 자동 식별된 특이치가 적용되지 않는다. 전체 기간에서 식별된 특이치가 어떠한 이동기간 중에서 발견되지 않는다면, 특이치는 해당 이동기간에서는 제외된다. outlier=remove이면, 자동 특이치 식별이 전체 기간을 이용하여 실행될 때, RegARIMA의 회귀모형에 더해진 특이치 회귀변수는 이동기간 분석을 하는 동안 RegARIMA 모형에서 제외된다. 또한 이들 효과는 추정되지 않고 시계열에서 제외된다. 특이치 항이 regression 명령에 지정되면,

이들은 추정된 모형에 포함된다. outlier=yes이면, 시계열의 어떠한 기간에서 RegARIMA 모형이 추정될 때마다 자동 특이치 검색이 실시된다.

start 이동기간 비교분석을 위한 시작점이다. default는 두 번째 기간의 시작월이다. *예) start=1990.jan*

print & save 선택가능한 출력표는 다음과 같다.

<Slidingspan: 이용 가능한 출력표>

이 름	단축	저장	내 용
yypercent	pcy	·	연간 변화율의 불안전성을 나타내는 백분율
sfspan	sfs	+	모든 이동기간에 대한 계절요인
chngspans	chs	+	모든 이동기간에 대한 월(분기)간 변화율
saspans	sas	+	모든 이동기간에 대한 계절조정계열
ychspans	ycs	+	모든 이동기간에 대한 연간 변화율
tdspans	tds	+	모든 이동기간에 대한 요일효과요인
indyypercent	pyi	·	간접 계절조정에 의한 연간 변화율의 불안성을 나타내는 백분율 표
indyysummary	syi	·	간접 계절조정에 의한 연간 변화율을 나타내는표, 히스토그램, 사분위수
indsfspans	sis	+	모든 이동기간에 대한 간접 계절요인
indchspans	cis	+	모든 이동기간에 대한 간접 월(분기)간 변화율
indsaspans	ais	+	모든 이동기간에 대한 간접 계절조정계열
indychngspans	yis	+	모든 이동기간에 대한 간접 연간 변화율

<Slidingspan: default 출력표>

이 름	단축	저장	내 용
header	hdr	·	이동기간분석의 이름
ssftest	ssf	·	각 이동기간의 안정성과 이동계절성에 대한 F-검정
factormean	fmn	·	각 이동기간에 대한 범위분석
percent	pct	·	계절요인, 요일효과, 최종계절조정계열, 월(분기)과 연간 변화율의 불안정을 보여주는 백분율 표

이 름	단축	저장	내 용
summary	sum	·	불안전 계절요인·요일효과·최종계절조정계열·월(분기)과 연간 변화율의 백분율 히스토그램, 사분위수, 표
yysummary	smy	+	연간 변화율에 대한 백분율 사분위수, 히스토그램
indfactormeans	fmi	·	간접법 계절조정의 계절요인에 대한 범위분석
indpercent	pci	·	간접법 계절조정에 있어 계절요인과 월간(분기간) 변화율의 불안전성을 나타내는 백분율 표
indsummary	smi	·	간접법 계절조정에 있어 불안전 계절요인, 월간(분기간) 및 연간 변화율의 백분율 표, 히스토그램, 사분위 수

기타명령

additivesa 가법계절조정에서 이동기간분석을 계절조정계열의 최대 차이(additivesa=difference) 혹은 원계열에 대한 최종계절조정계열의 최대 비율(additivesa=percent)로 할 것인지를 지정한다. 계절조정계열에서 월간, 분기, 연간 변화율이 분석이 되면, 이 옵션의 차이나 비율은 결정된다. default는 additivesa=difference이다. 어떠한 기간의 계절조정계열에 0보다 작거나 같은 값을 갖는다면, 기간이동분석은 차이로 실시된다.

fixx11reg x11regression 명령을 사용하는 경우, 이동기간분석을 하는 동안 불규칙요인 회귀모형을 재추정할 것인지 지정한다. fixx11reg=yes이면, 불규칙요인 회귀모형의 회귀계수를 전체 계열에서 추정한 값을 분석하는 동안 고정해서 사용한다. fixx11reg=no이면, 불규칙요인 회귀모형의 회귀계수를 각 기간에서 재추정한다. default는 fixx11reg=yes이다.

x11outlier 이동기간분석을 하는 동안 x11regression 명령에서 지정한 불규칙요인 회귀모형 추정을 위하여 가법 특이치(AO)를 식별할 것인지를 지정한다. x11outlier=yes이면, 가법 특이치 식별이 매 기간 실시된다. 가법 특이치 자동 식별이 전체 시계열에서 실행되었을 때, 가법 특이치 회귀변수는 이

동기간분석 전에 불규칙요인 회귀모형에서 제거된다. x11outlier=no이면, 전체 시계열에 대한 가법 특이치 자동 식별은 매 기간 보류된다. 가법 특이치의 효과를 추정하는 계수는 불규칙요인 회귀모형의 모수를 재추정할 때마다 재추정한다. 이 옵션은 다음 경우에 무시되며, default는 x11outlier=yes이다. x11regression 명령을 사용하지 않는 경우, x11regression 명령에서 sigma 옵션을 사용한 경우, aictest 옵션에서 불규칙요인 회귀모형을 선정하지 않는 경우이다.

참 고

이동기간분석에서 계절조정의 형태와 요일조정의 실시여부에 따라 다른 통계량을 검토해야 한다. 승법, log-가법 계절조정모형인 경우, 계절요인, 계절조정계열의 월(분기)간 및 연간 변화율을 분석한다. 승법, log-가법 계절과 요일효과모형인 경우, 요일요인과 계절조정계열을 분석한다. 요일효과를 조정하지 않은 가법 계절조정모형인 경우, 계절조정계열과 계절조정계열의 월(분기)간 및 연간 변화율을 분석한다. 요일효과를 조정하면, 계절과 요일효과가 조정된 계열에 대해서 분석을 실시한다.

가법 계절조정의 경우, 조정된 계열이 너무 작거나 음의 값이 존재하면 이동기간 통계의 근간이 되는 변화율이 불안정하게 나타날 수 있다. 이러한 경우, print(또는 save) 옵션의 saspan을 이용하여 구한 조정된 계열에 대해 주관적 기간 분석을 통해 과도한 불안정성을 찾을 수 있다.

arima 명령에서 자동 ARIMA 모형 선정 옵션이 지정되었다면, 이 절차에 의해서 선정된 모형은 모든 이동기간에 대해서 사용된다. 이 절차에 의해서 모형이 선정되지 않았다면, 이동기간분석을 하는 동안 모형은 추정되지 않는다. slidingspan 명령의 많은 표는 개별 화일로서 저장할 수 없으며, 실행시 “-s”로 계절조정 진단요약 옵션을 지정하면 이동기간분석의 정보가 진단요약 화일인 .udg(또는 ·xdg)의 확장자 명으로 저장된다.

합성계절조정의 직·간접 조정에 대한 이동기간분석을 하고자 한다면, 이동기간 옵션은 구성계열에 대해서 각각 지정해야 한다. 계절 filter 길이가 각 구성계열마다 같지 않다면, 사용자는 분석하고자 하는 기간의 길이

가 같도록 하기 위해서 구성계열의 입력화일에서 length 옵션을 각기 사용해야 한다.

자동 계절 filter 선정 옵션을 사용하면, 원계절조정을 하기 위해 사용된 계절 filter를 각 기간에 대한 계절조정에 사용한다.

slidingspan은 history와 함께 실행시킬 수 없으며, 실행시키면 경고 메시지를 주며 실행을 멈춘다.

예 제

(예 1) 모든 월에 대해서 3x9의 계절요인을 갖는 승법 월별 계절조정의 이동기간분석에 대한 모든 옵션을 default로 설정한 프로그램이다.

```
series { file = "tourist.dat"   start = 1976.1 }
x11 { seasonalma = s3x9 }
slidingspans {    }
```

(예 2) 첫 2개 분기는 3x9 계절 filter, 나머지 2개 분기는 3x9 계절 filter, 7개항 Henderson 추세 filter를 사용하는 log-가법 계절조정의 이동기간분석에 대한 선택된 검정의 기준치를 0.5로 설정하는 예이다.

```
series { file = "qstocks.dat"
            start = 1967.1
            title = "Quarterly stock prices on NASDAC"
            freq = 4           }
x11 { seasonalma =(s3x9 s3x9 s3x5 s3x5)
        trendma = 7
        mode = logadd       }
slidingspans { cutsf = 5.0
                  cutchng = 5.0          }
```

(예 3) 요일효과 조정과 자동 특이치 식별·조정을 위하여 사용되는 회귀변수를 갖는 (0, 1, 2)(1, 1, 0)12의 계절 ARIMA 모형의 예이다. 서술된 회귀변수는 상수, 요일효과와 1982. 5월과 1982년 9월이 ramp이다. 특이치와 요일효과를 사전조정한 후, 계절조정한다. 이동기간분석 시에 자동 식별 절차에 의해서 발견된 특이치를 각 이동기간에 대해서 재추정된 REGARIMA 모형에 포함시킨다.

```
series { title = " Number of employed machinists"
         start = 1980.jan      file = "machine.emp"     }
regression { variables = (const td rp82.may-82.oct)  }
arima       { model = (0 1 2)(1 1 0) }
outlier     {  }
estimate  {  }
check      {  }
forecast  {  }
x11         { mode = add }
slidingspans{ outlier = keep}
```

(예 4) 추정하고자 하는 사전에 정의된 회귀효과는 상수, 요일과 고정된 계절성이며 ARIMA모형은 (3, 1, 0)으로 60개 시점을 예측한다. 추정된 요일에 대해서 사전조정을 한 후 시계열을 계절조정한다. 각 기간에 대한 REGARIMA 모형의 모수값을 재추정한다.

```
series { title = "Cheese Sales in Wisconsin
         file = "cheeze.fil"  start = 1975.1 }
transform { function = log }
regression { variables = (const seasonal tdnolpyear) }
arima  { model = (3 1 0) }
forecast { maxlead = 60 }
x11 { save = seasonal   appendfcst = yes }
slidingspans { fixmdl = no }
```

TRANSFORM

내 용

RegARIMA 모형을 추정하기 전에 계열을 변환 또는 조정하기 위한 명령이다. 변환은 Box-Cox의 역(power)변환 또는 log변환, 월길이 조정, 사용자가 사전에 정의한 조정요인에 의한 변환 등이 가능하다. 사용자가 정의한 사전조정요인의 자료는 data 또는 file에 의해 파일로 주어져야 한다.

일반형식

```
transform { function = log   or   power =0.0
            adjust = lom
            title = "prior adjustment factors"
            start = 1975.jan
            data = (1.25…1.90) or file ="prioradj.dat"
                                  format = "(6f12.3)"
            name = "Adjfac"
            mode = ratio
            aicdiff = 0.0
            type = temporary
            print = (none)
            save = (prioradj)
            savelog = atr
          }
```

명 령

aicdiff 자동 변수 변환을 수행할 때(function=auto), 변수 변환을 하지 않기 위해 필요한 AICC 차이를 지정한다. default 값은 aictest=-2.0이다.

adjust 월 길이(adjust=lom), 분기 길이 조정(adjust=loq)을 실시하며, 월 혹은 분기자료의 윤년효과 조정(adjust=lpyear)을 실시한다. default는 조정 안함(adjust = none)이다. td가 regression 명령의 variables에 의해 지정된 경우 adjust는 사용하지 않는다. 요일효과 변수 td, td1coef가 regression 혹은 x11regression 명령의 variables 옵션에서 지정된 경우나 가법 또는 pseudo-가법 계절조정 모형이 x11 명령의 mode 옵션에서 지정된 경우에는 adjust 옵션을 사용할 수 없다. 윤년효과 조정(adjust=lpyear)은 power 혹은 function 옵션에서 log 변환을 한 경우에만 할 수 있다.

data mode=diff일 때를 제외하고, 하나 또는 두 개의 사전조정 요인을 갖는 자료는 입력 시계열을 나눌 수 있도록 양의 값을 가져야 한다. default는 단위벡터(사전조정 안함)이다. data(또는 file) 옵션이 사용되면, 조정요인은 계열의 모든 관측치(또는 series의 span으로 지정된 기간)에 주어져야 한다. 조정요인은 향후와 과거 예측치가 주어져야 한다. 조정요인은 free format으로 읽는다. data 옵션이 사용되면 file 옵션을 사용할 수 없다. mode=diff일 때, data에 있는 값을 빼며, 이때 그 값은 양수일 필요는 없다. 영구 및 임시 사전조정요인이 type 옵션에서 주어질 때, data 옵션에 의해 2개의 자료를 입력할 수 있다.

file 사용자 정의의 사전조정요인을 포함한 파일 이름으로 따옴표 안에 기술한다. 파일이 현재의 폴더에 없는 경우에는 경로를 지정해 주어야 한다. file 옵션을 이용하면, data 옵션을 이용할 수 없다. 값의 제한은 data의 경우와 같다.

format 파일로 부터 사전조정요인을 읽기 위한 format을 나타낸다. 입력은 다음의 3가지 형태가 있다. series 명령의 format 옵션에서 사용된 형태와 같다.

name 사전조정요인 이름을 따옴표 안에 64자 이내로 기술한다.

start 사용자가 정의한 사전조정요인의 시작 월(분기)이며, default는 계열의 시작 월(분기)이다.

function log, square root, inverse, logistic을 이용하여 입력계열 Y_t를 변환한다. function=auto의 경우 regression과 arima 명령 혹은 (0 1 1)(0 1 1)에서 지정한 RegARIMA 모형을 이용하여 구한 AIC를 기초로 log 변환과 변환하지 않음을 결정한다. default는 변환하지 않음(function=none)이다. function과 power를 같이 사용할 수 없다.

<function 옵션에서 이용 가능한 변환>

value	변환	Y_t 범위	power 명령
none	Y_t	모든 값	power = 1
log	$\log(Y_t)$	$Y_t > 0$	power = 0
sqrt	$\frac{1}{4} + 2(\sqrt{Y_t} - 1)$	$Y_t \geq 0$	power = 0.5
inverse	$2 - \frac{1}{Y_t}$	$Y_t \neq 0$	power = -1
logistic	$\log(\frac{Y_t}{1-Y_t})$	$0 < Y_t < 1$	없음

mode 사용자 정의의 사전조정요인을 표현하는 형태를 정의한다. 원계열을 나누는 사전조정요인이 퍼센트(100 100 50 …)가 아닌 비율 (1.0 1.0 0.5 …)로 주어지면, mode=ratio를 사용한다. 사전조정을 원계열에서 빼는 경우, mode=diff를 사용한다. 계절조정의 mode가 x11 명령에서 승법 또는 log 가법일 때, mode=diff를 사용하면 사전조정요인은 log scale인 것으로 가정한다. 요인들을 지수화하여 원계열과 같은 수준으로 만든다. mode가 사용되지 않으면 사전조정요인은 퍼센트인 것으로 가정한다(mode=percent). type 옵션에 의해 영구(permanent)와 임시(temporary) 사전조정요인을 지정하였다면, 두 개의 값을 이 옵션에 의해 지정할 수 있다. mode는 ratio와 diff를 함께 사용할 수 없다.

precision 사전조정요일의 파일을 읽기 위한 소수점 자리수이다. 이 옵션은 format 옵션에서 미리 정의한 format을 사용할 때만 사용할 수 있다. 0~5까지의 정수를 사용할 수 있다. *예) precision=5* power Box-Cox 변환을 이용하여 입력계열 Y_t를 변환한다.

$$Y_t \quad \rightarrow \quad y_t = \begin{cases} Y_t & \lambda = 1 \\ \log(Y_t) & \lambda = 0 \\ (Y_t^{\lambda} - 1)/\lambda & \lambda \neq 0, 1 \end{cases}$$

power λ는 사전에 주어져야 하며(*예: power = .33*), default는 변환하지 않음(power=1)이다. function 옵션에서 log 변환(power=0), sqrt root 변환(power= 0.5), inverse 변환(power=-1)을 지정할 수 있다. power와 function 옵션을 같은 파일에서 사용할 수 없다.

print & save transform 명령에서 사용할 수 있는 결과표이다. prior, aictransform과 prioradjusted표는 default로 인쇄된다.

<transform: 이용할 수 있는 결과표>

이 름	단축	저장	내 용
aictransform	tac	·	변수 변환을 위한 AIC-검정 결과
priorfactors	a2	+	사전조정요인
permprior	a2p	+	영구 사전조정요인
tempprior	a2t	+	임시 사전조정요인
prioradjusted	a3	+	사전조정계열
permprioradjusted	a3p	+	영구 사전조정요인을 이용하여 조정된 사전조정계열
prioradjustedptd	a4d	+	사전요일 조정을 포함한 사전조정계열
permprioradjustedptd	a4p	+	영구 사전조정요인과 사전요일 조정을 이용하여 조정된 사전조정계열
transformed	trn	·	사전조정과 변수 변환된 계열

type 사용자가 정의한 사전조정요인이 영구적 요인(원계열 뿐만 아니라 최종 계절조정계열에서 제거된)인지 혹은 임시적 요인(계절요인을 생성할 목적으로 원계열에서 제거되었지만 최종 계절조정계열에서는 제거되지 않은)인지를 지정한다. 옵션에 하나만 주어졌다면(type=temporary), 사용자가 정의한 사전조정요인들 중 하나의 형태(임시적 요인)만 사용된다. 둘 다 주어졌다면(type=temporary permanent), 두 형식(임시적, 영구적 요인)이 다 사용된다. default는 type =permanent이다.

title 사용자 정의의 사전조정요인에 대한 이름으로 따옴표 안에 79자 이내로 명명해야 한다.

참 고

Box-Cox 또는 logistic 변환이 월길이 조정 또는 사전조정요인과 함께 지정되면, 시계열은 처음에 월 길이 또는 사전조정요인을 조정하고 Box-Cox 또는 logistic 변환을 한다. 월 길이와 사전조정요인이 동시에 지정된 경우에는 결합된 조정요인(월길이×사전조정)이 사용된다. 분기 길이 조정도 같은 방법으로 한다.

Adjust 옵션의 lom(또는 loq)이 지정되면, 정확한 조정요인은 series의 주기에 의해 결정된다. 월별 시계열(Y_t)의 경우, 각 관측치를 각 월(m_t)의 일수로 나눈 다음에 평균 월 길이(30.4375)를 곱해준다($30.4375*Y_t/m_t$). 분기 길이 조정도 같은 방법으로 수행하며, ($91.3125*Y_t/q_t$)에서 q_t는 t분기의 일수이다. 변환을 하고 월 길이가 조정된 자료의 예측치는 원래의 수준으로 다시 재조정해야 한다.

x11로 계절조정을 했을 때, 회귀계수에 의해 만들어진 요인들로 원계열이나 최종 계절조정계열을 조정하지 않는 한, 어떠한 값의 power나 function을 사용해도 무방하다. 그러나 회귀계수에 의해 만들어진 요인들로 원계열이나 최종 계절조정계열을 조정하는 경우에는 사용할 수 없다. 이 경우, 채택할 수 있는 변환은 (가법이나 log-가법 계절조정에서) log 변환과 (가법 계절조정에서는) power=1 변환만 할 수 있다. 계절조정을 하지 않으면 어떠한 power 변환도 가능하다.

회귀계수에서 유도된 요인의 형태와 계절조정과정에서 만들어진 요인의 형태가 같다. 이에 따라, 결합된 조정요인이 만들어 지며, 조정 진단을 할 수 있다. RegARIMA 모형에서 log-변환계열에 적용이 되면, 회귀요인은 가법(혹은 log-가법) 조정형태에 의해 생성된 계절요인과 같은 비율의 형태로 표시된다. 역으로 RegARIMA 모형이 원계열에 적합이 되면, 회귀요인은 원계열의 척도와 같다. 이것은 가법 조정형태에 의해 만들어진 계절요인의 척도와 같다.

function=auto와 시계열이 양의 값을 갖는다면, RegARIMA 모형을 변환한 계열과 변환하지 않은 계열에 적합하여 다음과 같을 때, log 변환을 실시한다.

$$AICC_{nlog} - AICC_{\log} < \Delta_{AICC} \quad \text{혹은} \quad AICC_{\log} + \Delta_{AICC} > AICC_{nlog}$$

여기서 $AICC_{\log}$는 RegARIMA 모형을 변수변환한 계열에 적합하여 구한 AICC 값이며, $AICC_{nlog}$는 변환하지 않은 계열에 적합하여 구한 AICC 값이다. Δ_{AICC}는 default로 -2로 갖으며, aicdiff 옵션에 의한 값이다.

가법 조정방법은 Census Bureau의 대부분 자료에 적용되었으나, 사용자가 가법 조정에 대한 통계적 지지가 강하지 않으면 승법 조정을 사용한다. 시계열이 0이나 음수 값을 갖는 경우, 변환을 하지 않는다.

RegARIMA 모형을 regression, arima 명령에서 지정하였다면, 검정을 위한 AICC 통계량은 이 모형에 의해 구해진다. 모형을 지정하지 않았거나 혹은 automdl, pickmdl 명령에 의해 자동으로 모형을 식별하도록 하였다면, airline 모형(ARIMA (0 1 1)(0 1 1) 모형)에 의해 AICC 통계량을 구한다.

사용자는 사전조정요인의 형태를 지정할 수 있다. 예를 들어, 1980년 1월부터 시작하는 임시 사전조정요인을 (6F12.5) format으로 temp.fil 파일에서 읽는다. 1975년 7월부터 시작하는 영구 사전조정요인은 perm.fil 파일에 (F15.3) format으로 저장되어 있다.

```
transform {
        type=(temporary  permanent)
        file=("temp.fil"  "perm.fil")
        format=("(6F12.5)"  "(F15.3)")
        start=(1980.jan  1975.jul)
        mode=(ratio  percent)
        }
```

data 옵션이 2개의 사전조정요인을 입력하는 데 사용되는 경우, 자료는 2개 행을 갖는 행렬을 취한다. tyep 옵션에 의해 자료의 형태를 지정한다. 다음 자료는 첫째 열은 일시적 사전조정요인 (1.055, 0.990, 1.025, …, 1.033), 두 번째 열은 영구적 사전조정요인 (1.000, 1.000, …, 1.000)으로 인식한다.

```
transform {
```

```
              type=(temporary  permanent)
              data=( 1.055      1.000
                     0.990      1.000
                     1.025      1.000
                      …          …
                     1.033      1.000 )
               start=(1980.jan  1975.jul)
               mode=(ratio)                 }
```

하나의 파일로 2개의 사전조정요인을 읽는 경우는 다음과 같다.

```
      transform {
              type=(temporary  permanent)
              file="both.fil"
               start=1980.jan
               mode=(ratio)                  }
```

X-12-ARIMA는 사용자가 정의한 영구와 임시 사전조정요인 가각에 대하여 최대 600개를 사용할 수 있다.

예 제

(예 1) 월길이 조정과 1967년 3월과 4월의 파업을 사용자 정의의 사전조정요인으로 하는 프로그램이다.

```
series {data = (879 899 462 670 985 973 … )
          start = 1967.jan    }
transform {data = (1  1  .5  .75  1  1  … )
              mode = ratio
              adjust = lom     }
```

(예 2) log 변환을 하고 경상금액을 불변금액으로 바꾸는 가격 디플레이터를 사용자가 정의한 사전조정요인으로 하는 프로그램이다. 시계열을 모형화하기 전에 사전조정을 하므로 변환시의 시작 월은 디플레이터 계열의 처음 월이다.

```
series{ title = "Total U.S. Retail Sales-Current Dollars"
        file = "retail.dat"
        start = 1980.jan}
transform { function = log
            title = "Consumer Price Index"
            start = 1970.jan
            file = "cpi.dat"
            format = "(12f6.3)"}
```

(예 3) (예 2)와 같으나, 사전에 정의된 형식을 사용자 정의 조정요인에서 읽는다.

```
series{ title = "Total U.S. Retail Sales-Current Dollars"
        file = "retail.dat"
        start = 1980.jan}
transform { function = log
            title = "Consumer Price Index"
            start = 1970.jan
            file = "cpi.dat"
            format = "21"   precision = 3
            name = "cpi" }
```

(예 4) 분기 시계열의 변동을 안정시키기 위해 $\sqrt[3]{}$ 변환을 하는 프로그램이다.

```
series { title = "Annual Rainfall"
         file = "rain.dat"
         period = 4
         start = 1901.1}
         transform { power = .3333 }
```

X11

내 용

X11 명령은 센서스국의 X-11과 X-11Q 프로그램보다 개선된 버전으로 계절조정을 하기 위한 명령이다. 이용자는 X11 명령에 의해서 시계열의 분해 유형(mode), 계절 및 추세 이동평균법(seasonalma와 trendma), 계절조정과정 중에 실행되는 극단 값의 조정방법(sigmalim) 등을 지정할 수 있다. 또한 1970년대 Bateman과 Mayer에 의해 개발된 방법론을 이용하여 부활절 조정을 지정하는 옵션(x11easter)을 사용할 수 있다. print와 save에 의해 지정된 output은 X-11 계절조정방법에 의한 최종 출력표와 진단결과를 보여준다. X-12-ARIMA의 부가적인 명령들에 의해 원계열과 조정상의 문제점을 진단하고 회귀모형에 의한 사전조정방법을 개선하며, 선행과 후행예측을 통해 시계열을 연장할 수 있다. 이 명령들을 이용함으로써 수정된 계열을 얻을 수 있으며, X-11은 더 나은 계절조정 인자를 도출할 수 있다. X-11 계절조정 진단에 대한 상세한 내용은 Shiskin, Young, and Musgrave (1967)와 Lothian and Morry (1978)을 참고할 수 있다.

일반형식

```
X11 { mode = pseudoadd
      seasonalma = s3x9
      trendma = 13
      sigmalim = ( 1.25  2.75 )
      title = "3x9 moving average, mad"
      appendfcst = yes
      force = totals
      final = user
      print = ( brief  +b2 )
      save = ( d10  d11 )
      savelog = ( m  7  q )
    }
```

명　　령

appendfcst　appendfcst=yes이면, 예측치는 x11과 관련된 표는 a16, b1, d10, d16, h1, regression과 관련된 표는 a6, a7, a8, a8.tc, a9, a10, x11regression과 관련된 표는 c16, c18에 저장되고, appendfcst=no이면 예측치는 저장되지 않는다. default는 appendfcst=no이다.

final　최종 계절조정계열에서 제거하고자 하는 사전조정요인들을 나타내며 이 요인들은 regression과 outlier 명령에 의해 구해진다. 가법 특이치(final=ao), 수준 변화와 램프 특이치(final=lc), 일시적 변화(final=tc), 그리고 사용자 정의의 회귀계수에 의해 추정된 사용자 요인(final=user)들을 제거할 수 있다. 이 옵션을 사용하지 않으면, 최종 계절조정계열은 위의 효과들을 포함한다.

force　계절조정계열은 다음 옵션을 이용할 수 있다.

(a) force=totals: 계절조정계열과 원계열의 연간 합이 같도록 조정한다.

(b) force=round: 반올림한 계절조정계열의 합이 반올림한 연간 합과 같게 되도록 매해 계절조정값을 조정한다.

(c) force=both: (a)와 (b)가 수행된다.

프로그램은 위의 (a) 또는 (b)의 과정을 default로 수행하지 않는다.

force=totals를 사용할 때, 연간 합간의 차이는 대략적으로 원계열의 월(분기)간 변동을 보전하도록 계절조정된 값에 분배된다. 자세한 내용은 Huot(1975)와 Cholette(1978)를 참고하라. 이 옵션은 계절패턴이 바뀌거나 요일효과조정이 수행되면 사용할 수 없다.

mode　원계열의 분해 유형을 지정하는 명령으로, 승법(mode=mult), 가법(mode=add), pseudo-가법(mode=pseudoadd), log-가법(mode=logadd) 등이 있다. transform 명령을 실행하지 않았다면, default는 mult이다. transform 명령을 실행하면, 분해유형은 계열에 맞게 선택된 변환과 매치된다. 그러나 변환명령의 옵션과 상충되면 log 변환 시에는 mode=mult, 변환을 하지 않았으면 mode=add가 적용된다.

seasonalma 계절요인을 추정하기 위한 계절이동평균(계절 “filter”로도 표기)을 지정한다. 계절이동평균은 n×m 이동평균으로 연속된 n개 연도와 m개 기간(월 또는 분기)의 단순평균이다(4.1절 참조). 다음의 계절 filter는 전체 계열이나 특정 월(분기)에 지정할 수 있다. 모든 월(분기)에 동일한 이동평균이 적용되면 하나의 계절 filter를 사용하지만, 특정 월(분기)에 각각 다른 계절 filter 사용하고자 하는 경우에는 각 월에 해당 계절 filter를 적어준다.

예) 분기계열 : seasonalma=(s3x3 s3x9 s3x9 s3x9)

계절이동평균을 지정하지 않으면, 자동으로 최종 계절 filter를 선택한다. 이 옵션은 seasonalma=msr로 수행되며, X-11-ARIMA/88의 이동계절성 비율(MSR: moving seasonality ratio)을 이용한다(4.1절, 4.2절 라 참조). 계절이동평균을 기술하지 않았을 때, 3x3 이동평균은 반복 계산시 초기 계절요인 추정시 사용되며 3x5 이동평균은 최종 계절요인을 추정하기 위하여 사용된다. 이러한 일련의 계절 filtering은 seasonalma=x11default로 수행된다.

< X-12-ARIMA **계절 Filter 옵션>**

이 름	설 명
s3x1	3×1 이동평균
s3x3	3×3 이동평균
s3x5	3×5 이동평균
s3x9	3×9 이동평균
s3x15	3×15 이동평균
stable	안정적인 계절 filter. 각 월(또는 분기)에 대한 하나의 계절요인은 (추세제거와 특이치를 조정한 후) 각 월이나 분기에 대한 단순평균에 의해서 생성된다.
x11default	3×3 이동평균은 반복계산시의 초기 계절요인, 3×5 이동평균은 최종 계절요인을 계산하기 위하여 사용된다.

print & save 표 A는 default에 의해 선택 가능한 출력표를 제공한다. 표 B는 print 옵션에 의해 출력되고 저장되는 기타 출력표를 제공한다. 표 C는 **print**에 기술된 선형플롯을 보여준다.

X11

<X11(A 표): default 출력표>

이　름	단축	저장	내　　용
x11easter	h1	+	x-11부활절조정요인
combholiday	chl	+	혼합된 휴일사전조정 요인, 표 A16
irrwt	c17	+	불규칙요인에 대한 최종 가중치
unmodsi	d8	+	수정하지 않은 최종 si-비율(차이)
ftestd8	d8f	·	안정과 이동계절성에 대한 F-검정, D8
replacsi	d9	+	극단 si-비율(차이)에 대한 최종 교체값, D 반복
movseasrat	d9a	·	각 기간에 대한 이동 계절성 비율
seasonal	d10	+	최종계절요인
seasonaldiff	fsd	+	최종계절요인 차이(단, pseudo-가법 계절 조정에 한함)
seasadj	d11	+	최종 계절조정계열
seasadjtot	saa	+	연간 합이 제약된 최종 계절조정계열 (force=roudn 또는 force=both)
saround	rnd	+	반올림된 최종 계절조정계열(force=round) 또는 반올림된 연간 합이 제약된 최종 계절조정계열 (force=both)
residualseasf	rsf	·	잔차 계절성에 대한 F-검정
trend	d12	+	최종 추세-순환
irregular	d13	+	최종 불규칙요인
adjustfac	d16	+	혼합된 계절요인과 요일효과요인
adjustdiff	fad	+	최종 조정차이(단, pseudo-가법 계절조정에 한함)
calendar	d18	+	혼합된 휴일과 요일효과요인
yrtotals	e4	·	원계열, 계절조정계열에 대한 연간 합의 비율
origchanges	e5	+	원계열의 증감률(차이)
sachanges	e6	+	계절조정계열의 증감률(차이)
revsachanges	e6a	+	연간 합으로 조정된 계절조정계열의 증감률(차이)
rndsachanges	e6r	+	반올림된 계절조정계열의 증감률(차이)
trendchanges	e7	+	최종 추세요인계열의 증감률(차이)
x11diag	f2	·	계절조정 진단의 요약
qstat	f3	·	품질 관리 통계량(quality control statistics)
tdaytype	tdy	·	월단위로 인쇄된 요일요인
specsa	sp1	+	차분된 계절조정계열(logadd나 mult라면, log 변환된 계절조정계열)의 스펙트럼 플롯
specirr	sp2	+	특이치가 완화된 불규칙계열의 스펙트럼 플롯

<X11(B 표): 기타 출력표>

이 름	단축	저장	내 용
ftestb1	b1f	·	안정된 계절성의 F-검정, 표 B1
trendb2	b2	+	초기의 추세-순환, B반복
modoriginal	e1	+	가중치가 없는 극단값을 수정한 원계열
modseasadj	e2	+	가중치가 없는 극단값을 수정한 계절조정계열
modirregular	e3	+	가중치가 없는 극단값을 수정한 불규칙 요인
robustsa	e11	+	로버스트한 최종 계절조정계열
totaladjustment	e18	+	총 조정 요인
mcdmovavg	f1	+	최종 계절조정계열의 MCD 이동평균
sib3	b3	+	초기의 수정되지 않은 SI-비율(차이)
replacsib4	b4	·	극한 SI-비율(차이)의 초기 대체값, B반복
seasonalb5	b5	+	초기의 계절요인, B반복
seasadjb6	b6	+	초기의 계절조정계열, B반복
trendb7	b7	+	초기의 추세-순환 요인, B반복
sib8	b8	+	수정되지 않은 SI-비율(차이)
replacsib9	b9	·	극한 SI-비율(차이)의 대체값, B반복
seasonalb10	b10	+	계절 요인, B반복
saesadjb11	b11	+	계절조정계열, B반복
irregularb	b13	+	불규칙 구성요인, B반복
irrwtb	b17	+	불규칙 구성요인의 초기 가중치
tdadjorigb	b19	+	초기 요일에 의해 조정된 원계열
extremeb	b20	+	극단값, B반복
adjoriginalc	c1	+	이상치, 요일 등의 사전요인을 조정한 원계열, C 반복
trendc2	c2	+	초기의 추세-순환 요인, C반복
modsic4	c4	+	수정된 SI-비율(차이), C반복
seasonalc5	c5	+	초기의 계절요인, C반복
seasadjc6	c6	+	초기의 계절조정계열, C반복
trendc7	c7	+	초기의 추세-순환 요인, C반복
replacsic9	c9	+	수정된 SI-비율(차이), C반복
seasonalc10	c10	+	초기의 계절요인, C반복
seasadjc11	c11	+	계절조정계열, C반복
irregularc	c13	+	불규칙 구성요인, C반복
tdadjorig	c19	+	최종 요일효과 조정한 원계열
extreme	c20	+	극단값, C반복
adjoriginald	d1	+	이상치, 요일 및 사전요인 조정한 원계열, D반복
trendd2	d2	+	초기의 추세-순환 요인, D반복
modsid4	d4	+	수정된 SI-비율(차이), D반복
seasonald5	d5	+	초기의 계절요인, D반복

<X11(B 표): 기타 출력표>

이 름	단축	저장	내 용
seasadjd6	d6	+	초기의 계절조정계열, D반복
trendd7	d7	+	초기의 추세-순환 요인, D반복
unmodsiox	d8b	+	이상치와 극단값의 라벨이 있는 수정되지 않은 최종 SI-비율
autosf	asf	·	계절요인을 자동으로 선택
trendadjls	tal	+	수준변화 특이치를 조정한 최종 추세-순환
irregularadjao	iao	+	점 특이치를 조정한 최종 불규칙 요인

< X11(C 표): print 명령에 기술된 플롯>

이 름	설 명
seasonalplot	월 또는 분기별 계절 요인 플롯
seasadjplot	최종 계절조정계열 플롯
trendplot	최종 추세-순환 플롯
irregularplot	최종 불규칙 요인 플롯
origwsaplot	원계열과 최종 계절조정계열 플롯
ratioplotori	원계열의 월별(또는 분기별)비율 플롯
ratioplotsa	계절조정계열의 월별(또는 분기별)비율 플롯

trendma 최종 추세-순환요인(trend-cycle)을 추정하기 위한 핸더슨 이동평균을 기술하는 명령으로 1보다 크고 101보다 작거나 같은 홀수로 표시하여야 한다. 예를 들어 **trendma=23**과 같이 입력한다. trendma를 지정하지 않으면, 자료의 통계적 특성을 고려한 추세이동평균을 선택한다. 월 자료인 경우에는 9항, 13항, 23항 핸더슨 이동평균을 선택하며, 분기 자료인 경우에는 5항, 7항 핸더슨 이동평균을 선택한다.

title 계절조정의 제목을 따옴표안에 입력한다. 소제목은 8개까지 쓸 수 있으며, 예를 들어 2개의 소제목을 가지고 있을 때는 다음과 같이 표기한다.

title= ("3×9, trading day adjustment"
"for sales of sporting goods")

각 제목은 줄을 달리 하여 사용한다. default는 없다.

savelog 로그 파일(4.4절 참조) 출력표에서 이용 가능한 진단내용을 보여준다.

<X11: 이용 가능한 로그 파일 진단 내용>

이 름	단축	내 용
m1	m1	M1 품질 관리 통계량
m2	m2	M2 품질 관리 통계량
m3	m3	M3 품질 관리 통계량
m4	m4	M4 품질 관리 통계량
m5	m5	M5 품질 관리 통계량
m6	m6	M6 품질 관리 통계량
m7	m7	M7 품질 관리 통계량
m8	m8	M8 품질 관리 통계량
m9	m9	M9 품질 관리 통계량
m10	m10	M10 품질 관리 통계량
m11	m11	M11 품질 관리 통계량
q	q	계절 조정 품질의 총지수
q2	q2	M2를 제외한 Q 통계량
movingseasratio	msr	이동 계절 비율
icratio	icr	$\bar{I}$/C비율
fstableb1	fb1	원계열의 안정한 계절성에 대한 F-검정
fstabled8	fd8	최종 SI-비율의 안정한 계절성에 대한 F-검정
movingseasf	msf	이동 계절성에 대한 F-검정
idseasonal	ids	식별할 수 있는 계절성 검정 결과

sigmalim 계절조정 과정을 반복적으로 수행할 때, 극단의 불규칙 값을 낮추기 위해 사용하는 상한 및 하한 sigma 한계를 기술한다. sigmalim은 상한과 하한의 두 개 실수 값을 가지며, 하한은 상한보다 작으나 0보다는 큰 실수 값이다. 예를 들면 sigmalim=(1.8 2.8)이다. default 값은 하한 1.5, 상한 2.5이며, 한계값을 지정하지 않았으면 default 값이 지정되고 반드시 콤마로 구분해야 한다. 즉, sigmalim=(, 3.0)인 경우 하한은 default인 1.5이고, 상한은 3.0이다.

type type=summary이면, 이미 계절조정을 하였거나 계절성이 없다고 가정한 원계열로 부터 잔차 계절요인 및 요일효과, 휴일요인과 함께 추세-순환요인, 불규칙요인 등과 관련된 진단 내용을 추정한다. 잔차 요인들은 제거되지 않는다. 최종 계절조정계열(D11)의 출력계열은 원계열(A1)과 같다.

type=trend이면, 계절조정 여부에 관계없이 최종 추세-순환 및 불규칙 요인을 추정한다. 입력계열은 이미 계절조정되었거나 계절성이 없는 것으로 가정한다. 이 가정 하에서 추세요인(D12)과 계절조정계열(D11)을 산출하기 위해 사전조정계열 뿐만 아니라 추정된 요일 및 휴일요인들을 제거한다. composite 명령을 갖는 메타파일로 총합계열을 간접 계절조정하려 할 때, 이 옵션은 계절조정 되지 않은 총합계열의 구성요소로 사용된다. Default로 type=sa가 지정되며 계열의 계절 분해요소를 계산한다. 위의 세 가지 유형의 type을 이용하여 최종 계절조정계열은 간접 계절조정을 만들기 위해 사용된다.

기타명령

forcestart force와 함께 사용되며, 계절조정계열의 총계를 합하는 연간 시작점이다. forcestart가 사용되지 않으면, 시작은 force=totals에 의해서 나타나는 연도로 간주한다. forcestart에서 시작하는 연간 기준시점을 월이나 분기(1/4분기; q1, 2/4분기; q2, …)로 지정할 수 있다. 예를 들면 지정된 회계연도의 시작월이 10월이고 다음해의 9월에 끝나는 경우, forcestart=october(또는 oct)이며, 회계연도가 3/4분기에 시작해서 다음해 2/4분기에 끝나는 경우는 forcestart=q3으로 입력한다.

trendic 핸더슨 이동평균의 최종 가중치를 생성하기 위한 불규칙 대 추세의 분산비율을 지정한다. 이 옵션이 지정되지 않았으면 trendic 값은 핸더슨의 추세 filter 길이에 의해 결정된다. default는 처음의 X-11과 X-11-ARIMA의 핸더슨 추세 filter의 최종 가중치에 가까운 수치이다.

forcetarget force와 같이 사용하며, 계절조정계열의 연간 합과 일치하여야 하는 목표 계열을 설정한다. Default는 forcetarget=original로 목표계열은 원계열이다. 기타 방법은 캘런더 조정 원계열(forcetarget=calendaradj), 영구 사전조정계열(forcetarget=permprior) 및 캘런더와 영구사전요인을 모두 조정한 원계열(forcetarget=both)을 목표 계열로 사용하는 방법이 있다.

itrendma 추세-순환의 초기 추정값을 만들기 위한 이동평균을 지정한다. itrendma=centered1yr이면 1년 중심화 이동평균이 사용된다. itendma=cholette2yr이면 2년 이동평균이 사용되며, 이는 캐나다 통계청의 Pierre Cholette에 의해 개발된 Leser 필터를 수정한 것이다. Default는 1년 중심화 이동평균이다. 2년 필터는 단기의 주기적 변동이 있거나 추세수준에 갑작스러운 변동이 있는 계열을 조정하는데 유용하다.

calendarsigma 극한값의 발견과 조정을 위해 사용되는 불규칙 관련 요인들(B4, B9, B17, C17)의 표준오차는 각 월(분기)이나 또는 2개월(분기)의 보충적 집합에 의해 각각 계산된다. calendarsigma=all이면, 각 월(분기)의 표준편차가 계산된다. calendarsigma=signif은 코크란 가설 검정에 의해 불규칙요인이 각 월(분기)마다 다른 경우, 각 월(분기)의 표준편차를 계산해 준다. calendarsigma=select 경우에는 모든 월(분기)를 두 개의 그룹으로 나누어 각 그룹의 표준오차를 계산한다. select에서 sigmavec은 월(분기)의 두 그룹 중 하나를 정의하는데 사용된다. calendarsigma가 지정되지 않으면 표준 편차는 5년간의 불규칙요인 값에 의해서 계산된다. 방법은 Dagum(1988)에 기술되어 있다. 극단값 조정은 <참고 4.1> 참조.

centerseasonal centerseasonal=yes이면, 사용자 정의의 계절회귀 효과와 결합하여 계절요인을 집중화한다. Default는 centerseasonal=no이다.

keepholiday 추정된 최종 계절조정계열에 휴일(명절)효과가 남아있는지 측정한다. Default는 keepholiday=no이며, 휴일조정요인은 최종 계절조정계열에서 제거된다. keepholiday=yes이면 휴일조정요인은 최종 계절조정계열에 남아 있는다. 계절요인 및 휴일요인 둘 다 조정된 계열을 만들기 위해 default가 사용된다.

spectrumaxis spectrumaxis=same이면, 차분된 원계열과 극단값을 조정한 차분된 계절조정계열의 스펙트럼이 같은 좌표축에 보여진다. Default는 spectrumaxis=diff이며 스펙트럼 그림이 따로 보여진다.

print1stpass	print1stpass=yes이면, X-11 Easter와 불규칙에 의한 회귀조정을 위해 불규칙 요인을 생성할 필요가 있는 계절조정 결과를 출력한다. print1stpass=no이면, 결과는 출력되지 않으며 단지 X-11 Easter와 불규칙 회귀 과정과 관련 있는 표만 출력된다. Default는 print1stpass=no이다. print1stpass=yes이면 print 명령은 표를 어떻게 출력할지에 대해 조정한다.
sfshort	시계열의 기간이 5년이라면, 계절요인을 구하기 위해서는 계절 filter를 조정해야 한다. sfshort=no라면 seasonalma 옵션이 있어도 계절요인을 구하기 위하여 안정계절 filter를 사용한다. sfshort=yes이면, X-12-ARIMA는 seasonalma 옵션에서 주어진 계절 filter를 사용한다.
sigmavec	calendarsigma=select 옵션 하에서 계산되는 불규칙요인의 표준편차를 구하기 위하여 월의 두 그룹 중 하나를 지정한다. 사용자는 한 그룹에 포함할 월 또는 분기(1/4분기: q1, 2/4분기: q2 등)를 입력해야 하며, 나머지 월은 다른 그룹에 속하게 된다. 예로 sigmavec=(jan feb dec)이면, 한 그룹은 1월, 2월, 12월의 표준편차를 구하며, 다른 그룹은 나머지 월의 표준편차를 구한다(결과표는 C17를 보면 된다). 이 옵션은 calendarsigma=select 옵션을 사용할 때만 사용할 수 있다.
x11easter	X-11의 Bateman과 Mayer의 비모수적 방법을 사용하여 부활절 효과를 추정할 것인지 여부를 지정한다. **x11easter=yes**이면 잠정 계절조정이 수행되고 이 조정의 3월과 4월의 불규칙성은 네 그룹으로 나누어, 절사평균을 구하고 부활절 효과를 추정하는데 사용된다. 자세한 내용은 Monsell(1989)를 참고한다. Default는 **x11easter=no**이다. x11easter 옵션은 휴일을 고려한 RegARIMA모형이 **regression** 명령에서 기술되었거나 부활절 조정이 **x11regression** 명령에서 지정되었다면 사용할 수 없다.
true7term	7개항 핸더슨 필터를 사용한 최종 가중치를 지정한다. true7term=yes이면, 중심화 핸더슨 필터가 적용되지 않은 계열의 마지막 관측값에 7개항 핸더슨 필터의 비대칭 최종 가중치가 적용된다. true7term=no이면, 5개항 핸더슨 필터의 중심화 비대칭 가중치가 적용된다. Default는 true7term=no.

참　　고

X-12-ARIMA 계절조정에 있어서, 원계열(O)은 3가지의 기본적인 요인으로 나누어진다.

· 추세 · 순환요인(C): 추세요인은 대체로 10년 이상 동일한 방향으로 상승 또는 하강하는 경향을 나타내는 장기적인 변동이며, 순환요인은 확장과 수축 기간이 반복되는 주기적인 변동이다.

· 계절요인(S): 일정한 기간 동안에 반복되는 변동으로 기후, 온도의 변화와 이에 따른 생활습관 등에 따라 나타나는 변동이다.

· 불규칙요인(I): 추세 · 순환과 계절요인에 의한 변동 이외의 변동으로 천재지변, 전쟁, 파업 및 급격한 경제정책 변화 등과 같이 극히 단기적이고 우발적인 사회현상 또는 자연현상 등에 의해 발생하는 변동이다.

원계열(Y)은 계절변동 특성에 따라 추세 · 순환요인(C), 계절요인(S), 불규칙요인(I) 등의 구성요인들로 결합된 몇 개의 모형으로 표현할 수 있다. X-12-ARIMA는 아래 표와 같은 4가지 분해모형에 적당한 계절조정모드를 제공하며, default는 승법모드이다.

계절조정 모드와 모형

모　드	설　명	원계열(Y)의 모형	계절조정계열 (SA)의 모형
mult	승법	Y=C×S×I	SA=C×I
add	가법	Y=C+S+I	SA=C+I
pseudoadd	Pseudo-가법	Y=C×[S+I-1]	SA=C×I
logadd	Log-가법	Log(Y)=C+S+I	SA=exp(C+I)

대부분 계절성을 가진 시계열은 계열수준이 증가하고 감소함에 따라 계절변동도 비례적으로 증가 · 감소하는 형태로 나타난다. 이러한 계열을 승법 계절성을 가진 계열이라고 한다. 여기서는 승법 구성요인을 추정하기 위해서 Dagum(1988)의 비율 이동평균방법(ratio-to-moving average method)을 사용한다. Pseudo-가법 모형은 어떤 월(분기)에는 방학이나 기후 등에 의해 계절성이 극히 작고, 나머지 월은 승법 계절성을 나타날 때 사용한다. 계절성의 크기가 계열의 수준에 의해서 영향을 받지 않는 것으로 나타나

면, 가법 계절성을 가지고 있다고 보고 가법 모드를 사용한다.

승법 분해방법의 대안으로 주로 사용하는 log-가법 모드는 경제분석에서 유용하게 사용할 수 있다. log-가법 계절조정 모드는 먼저 추세요인을 log 변환한 원계열($\log(O)$)을 가법 분해하여 구한 후, 원계열과 같은 단위를 갖는 추세를 나타내기 위해 지수화(exponential)한다. 그 결과 추세의 추정에 편의를 갖게 되며, 이 편의는 2002년 Thomson과 Ozaki의 편의수정(bias-correction)으로 조정한다.

승법, Pseudo-가법, Log-가법 계절조정에서 계절요인과 불규칙요인은 1에 대한 비율로 추정되고, 그 결과자료는 100에 대한 백분율로 나타난다. 가법 계절조정에서, 계절요인과 불규칙요인은 원계열과 동일한 단위이며, 0을 중심으로 변한다.

RegARIMA 모형을 regression과 arima 명령에서 사용 시, regression 명령에서 정의된 요일효과, 휴일, 특이치와 기타 회귀 요인들은 RegARIMA 모형에서 구해지며, 계절조정에 앞서 원계열을 조정하기 위해 사용된다. 이러한 의미에서 사전조정(Prior Adjustment)이라고 하며, 이들 요인들을 사전조정요인이라고 한다. 위의 요인들은 총합 계절조정 요인을 도출하고 계절조정결과를 진단하기 위하여 계절조정 과정에서 생성되는 요인들과 같은 유형이어야 한다.

regARIMA 모형을 로그변환 계열에 적합시킨다면, 회귀요인은 비율의 형태로 표시되며, 이들 요인들은 승법 또는 Log-가법조정모형에서 생성된 요인들과 같은 형태이다. 반대로 regARIMA 모형을 원계열에 적합한다면, 회귀요인들은 원계열과 같은 척도로 측정되며, 이들 요인들은 가법조정모형에서 생성된 요인들의 척도와 동일하다. 그러므로 transform 명령의 function과 power 옵션에 의해 지정된 변환과 x11 명령의 mode 옵션에 의해 지정된 계절조정이 양립할 수 있도록 주의해야 한다. 특히, mode의 default는 '승법계절조정'이므로, transform 명령의 function과 power옵션의 default인 '변환하지 않음'과 충돌할 수 있음을 조심해야 한다. 현재, Pseudo-가법 계절조정에서 원계열을 사정조정하기 위한 회귀요인은 사용할 수 없다.

시계열의 계절 진폭이 크지 않은 경우, 승법 계절성을 가진 시계열에서 승법 계절조정모형과 Pseudo-가법 계절조정모형에 의한 계절조정결과는 매우 비슷하게 나타난다. 그러나, 만약 최소 계절요인이 0.7보다 작거나 같으면 두 방법은 뚜렷한 차이를 보인다.

승법 계절조정을 했을 때, 큰 계절요인 값을 갖는 월(분기)보다 작은 계절요인 값을 가진 월에 극한값(100.0미만, 특히 0)이 더 많다면 Pseudo-가법 계절조정방법을 사용하는 것이 적절하다. Pseudo-가법 계절조정에 대한 자세한 내용은 Baxter(1994)자료를 참고할 수 있다.

극한값을 가진 불규칙 구성요인의 하향 가중치: 불규칙 요인(승법 계절조정모형은 1.0, 가법 계절조정모형은 0.0)의 평균을 μ_I, 표준편차를 σ_{X11}라 할 때, $I_t-\mu_I$의 절대값이 σ 한계값의 하한에 σ_{X11}를 곱한 수치보다 작으면 불규칙 값 I_t는 완전가중치(full weight, 1)를 준다. σ 한계값의 상한에 σ_{X11}를 곱한 수치보다 크면 불규칙요인 값 I_t의 가중치는 '0'이 주어지며, I_t는 계절요소 μ_t로 대체된다. 이외의 I_t는 부분적으로 하향 가중이 된다. 이를 수식으로 표현하면 다음과 같다.

$$\omega_t = 0.0, \quad |I_t-\mu_I| > 2.5\sigma_{X11} \text{ 일 때}$$
$$\omega_t = 1.0, \quad |I_t-\mu_I| < 1.5\sigma_{X11} \text{ 일 때}$$
$$\omega_t = 2.5-|I_t-\mu_I|/\sigma_{X11}, \quad 1.5\sigma_{X11} \leqq |I_t-\mu_I| \leqq 2.5\sigma_{X11} \text{ 일 때}$$

계절 filter 자동 선택: X-11-ARIMA/88(Dagum, 1988)의 과정과 같다. 두 번의 계절조정 반복 추정에서 3×3 이동평균은 초기의 계절요인, 3×5 이동평균은 최종 계절요인을 구한다. 3번째와 마지막 계절조정에서는 3×3 이동평균으로 초기 계절요인을 구하고, 최종 계절요인을 추정하기 위한 계절 filter는 이동계절성 비율 $\overline{I}/\overline{S}$(global moving seasonality ratio: GMSR 또는 MSR)로 결정한다. X-12-ARIMA는 GMSR의 크기에 의해 3×3, 3×5, 3×9 이동평균 중 하나를 선택한다. GMSR에 의한 이동평균법의 결정은 제 4.1장이나 Lothian(1978)을 참조한다.

예측 시계열 확장: RegARIMA 모형은 가장 최근(최초) 관찰치의 계절조정 결과를 향상시키기 위해 미래(과거)시점에 대한 시계열을 예측하여 확장한다. X-12-ARIMA에서는 RegARIMA 모형에 forecast 명령이 있으면 계절조정을 하기 전에 향후 1년(forecasting)과 최초 1년 전(backcasting)의 시계열을 예측한다. 예측 시계열 확장 없이 계절조정을 하려면 forecast에서 maxlead=0으로 지정한다.

계절조정계열의 잔차 계절성 및 요일효과: X11 명령에 의해 계절성 및 요일효과를 파악하기 위한 작업이 반복적으로 수행되며, 뚜렷한 계절 및

요일효과에 대한 정점을 발견 시에는 경고 메시지와 plot이 출력된다. 출력 시 제한 옵션(-n flag)을 사용하면, plot은 출력되지 않으나, -n flag 없이 프로그램을 다시 실행하라는 내용의 메시지가 출력된다.

수준변화 및 최종 핸더슨 추세: 계절조정을 하기 전에 수준변화 특이치가 추정되고 추정된 수준변화 특이치가 제거되었을 때, 수준변화는 최종 핸더슨 추세-순환에 되돌려지며, 이 요인은 관찰된 자료의 수준을 갖게 된다. print=trendadjls 옵션에 의해 수준변화를 조정한 시계열의 추세-순환 표를 산출할 수 있다.

부활절 조정: 휴일을 고려한 RegARIMA 모형을 regression 명령에서 지정하면, 부활절 조정 옵션은 x11 명령에서 사용할 수 없으나, x11regression 명령에서는 사용 가능하다.

연간 합: 계절조정계열의 연간 합을 원계열의 연간 합과 강제로 같게 하면 계절 조정의 질이 낮아질 수 있으며, 특히 계절 패턴이 시간의 흐름에 따라 변화하고 있을 때는 더욱 질이 낮아질 수 있다. 일년 간 요일효과요인의 총합이 가변적이고 0과 꽤 차이가 나기 때문에 요일효과요인을 조정하는 것은 부적절하다.

극단값 식별표시를 가진 SI 표: D 8.B는 사용자에게 X-11 계절 조정 프로그램에 의해 생성된 수정되지 않은 계절-불규칙 값(SI 값)을 극단값(C 17의 불규칙 가중치) 조정을 통해 수정할 것인지 또는 RegARIMA 특이치에 의해 영향을 받을 것인지 등의 정보를 제공한다. X-11의 극단값으로 식별된 SI 값은 "*"와 함께 출력된다. RegARIMA 특이치가 한개만 나타날 때의 SI 값은 "#", 적어도 한 개 이상의 RegARIMA 특이치와 관련된 극단값으로 식별된 SI 값은 "&"와 함께 출력된다. 한 개 이상의 RegARIMA 특이치를 가지고 있는 SI값은 "@"와 함께 보여 진다. 램프 특이치의 처음과 끝 시점 사이의 모든 관찰값들은 그것들이 특이치인 것처럼 표시된다.

승법 계절조정모형에서 수준변화(LS)에 의해 영향을 받을 수 있는 수준변화 특이치 전 후의 SI값은 "-"와 함께 표시된다. 이렇게 표시된 관측값의 갯수는 회귀계수 추정치로 측정된 수준변화 특이치의 크기와 SI-비율을 구하는 추세에 사용된 핸더슨 필터의 길이에 따라 다르다.

<수준변화에 영향받았다고 가정할 때, D8의 수준변화 갯수>

Percent Change	Length of Henderson Filter				
in Level ($\triangle L$)	23	13	9	7	5
$\triangle L \leq 11$	0	0	0	0	0
$11 < \triangle L \leq 12$	1	1	0	0	0
$12 < \triangle L \leq 13$	1	1	1	0	0
$13 < \triangle L \leq 1$	2	1	1	0	0
$15 < \triangle L \leq 18$	2	1	1	1	0
$18 < \triangle L \leq 19$	2	2	1	1	0
$19 < \triangle L \leq 20$	3	2	1	1	0
$20 < \triangle L \leq 26$	3	2	1	1	1
$26 < \triangle L \leq 29$	3	2	2	1	1
$29 < \triangle L \leq 36$	4	2	2	1	1
$36 < \triangle L \leq 55$	4	3	2	1	1
$55 < \triangle L$	5	3	2	1	1

비계절성을 지닌 계열의 처리: 계절성을 가지고 있지 않은 계열은 **type=trend**를 이용하여 추세-순환요인과 불규칙요인으로 분해할 수 있다. 핸더슨 추세와 극단값 탐지와 관련된 X-11 계절조정 분해과정을 단순화함으로서 분해요인들을 얻을 수 있다.

예 제

(예 1) 계절 filter 길이를 선택하기 위하여 이동계절성 비율을 이용하는 등의 모든 옵션을 default로 한 승법계절조정의 예이다. 1976년 1월부터 시작되는 월별자료는 free format으로 현재 디렉토리내에 klaatu.dat로 저장되어 있다.

```
series{file ="klaatu.dat"   Start = 1976. 1 }
X11    {     }
```

(예 2) 모든 월의 계절요인을 추출하기 위한 3×9 이동평균, 추세·순환요인을 구하기 위한 23-개항 핸더슨 이동평균을 하는 승법계절조정모형 예이다. AIC-통계량을 사용하여 불규칙요인의 회귀에서 요일효과와 회귀계수에 대한 유의성을 검정한다.

```
series{file ="klaatu.dat"    Start = 1976. 1 }
x11    {SeasonalMa = s3x9    TrendMa = 23      }
```

(예 3) 처음 2개의 분기 계절요인을 추출하기 위하여 3×3 이동평균, 나머지 분기의 계절요인을 추출하기 위하여 3×5 이동평균, 7-개항 핸더슨 추세 이동평균으로 추세·순환요인을 추정하는 분기 계절조정의 예이다.

```
series{file ="qhstarts.dat"
        start = 1967.1        period = 4      }
x11 {seasonalma = (s3x3   s3x3 s3x5 s3x5)   trendma = 7    }
```

(예 4) 사전조정 회귀변수를 갖는 월별 시계열로 계절 ARIMA 모형으로 예측 시계열은 연장하지 않는다. 지정된 회귀변수는 상수, 요일효과, 1972년 5월과 1976년 9월의 수준변화이다. ARIMA모형은 $(0,1,2)(1,1,0)_{12}$이며, REGARIMA 모형에 의해서 추정된 특이치, 수준변화와 요일효과에 대한 사전조정을 한 후 가법 계절조정을 한다. 불규칙요인에 포함되어 있는 특이치를 σ한계 범위 (2.0, 3.5)에서 찾고, 계절조정결과 출력시 alltable 옵션을 사용한다.

```
series{title ="exports   of leather goods"   start=1969.jul
        data =(815 866 926 … 942)    }
regression { variables =(const td ls1972.may ls1976.oct) }
arima { model = (0 1 2) (1 1 0)        }
estimate   {       }
forecast   { maxlead = 0   }
x11 { mode = add   print = alltable   sigmalim = (2.0   3.5)}
```

(예 5) 자동 ARIMA 모형 선택에 의해서 12개월이 예측되어 연장된 월별자료를 3×9 이동평균으로 승법계절 조정 예이다. 계절조정계열에 대한 회계기간의 연간 합을 원계열의 합과 같도록 만든다. 이때 회계기간의 시작은 10월이다.

```
series{title ="exports of truck parts"    start=1967.1
        file = "x21109.ori"  }
automdl {    }
x11 {seasonalma = s3x9    force=totals   forcestart = oct}
```

(예 6) 사전 회귀효과는 요일효과변수와 상수이다. 사용자 정의의 회귀변수는 1988년과 1990년의 판매율 증가효과로 special.dat에 있다. ARIMA 모형은 $(3,1,0)(0,1,1)_{12}$이다. 계절주기는 12이나 default이므로 지정하지 않는다. RegARIMA에 의해 추정되는 요일효과와 사용자 정의 회귀효과를 사전조정하고 계열의 미래와 과거에 대해 각각 12개월을 예측한 후 승법계절조정모형을 수행한다. 계절조정 제목은 줄을 달리하여 지정한다.

```
series{title ="unit auto sales"    file = "autosal.dat"
        start = 1985.1    }
transform { function = log }
regression { variables =(const td)       user =(sale88 sale90)
               file ="special.dat"    format ="(2f12.2)"    }
arima { model = (3 1 0)(0 1 1)12     }
forecast  { maxlead = 12   maxback = 12  }
x11{ title =("Unit Auto Sales"
               "Adjusted for special sales in 1988, 1990" )      }
```

(예 7) X-11의 format으로 사전에 정의한 자료를 읽는다. 시작일은 데이터 file의 초기 날짜를 이용하며 따로 지정할 필요가 없다. 특이치가 있는 log 변환 자료에 대해 RegARIMA 모형을 지정한다. 이 모형으로 5년간을 예측한다. 모든 월에 대해 3×9 계절이동평균을 이용하여 승법 계절조정을 한다.

```
series{ title ="Northeast one family housing  starts"
        file ="cne1hs.ori"  name ="cne1hs"  format ="2r"}
transform { function = log }
regression {variables =(ao1976.feb  ao1978.feb  ls1980.feb
                        ls1982.nov ao1984.feb)   }
arima { model = (0 1 2)(0 1 1) }
forecast  { maxlead = 60  }
x11   { seasonalma = (s3x9)
        title = ("adjustment of  1  family  housing  starts " ) }
```

(예 8) 사전 정의된 회귀효과는 상수이다. 사용자가 정의한 회귀변수는 1980년, 1985년, 1991년 파업이며 strike.dat에 있다. ARIMA 모형은 (0, 1, 1)(0, 1, 1)12이며, default로 1년을 예측한다. 사용자가 정의한 회귀효과로 사전조정을 한 후, 계절조정을 한다. RegARIMA모형을 추정하기 전에 X-11 요일효과와 Bateman-Mayes과정으로 추정한 휴일효과로 계열을 조정한다.

```
series { title = "Automobile  sales"
         file = "carsales.dat"   start = 1975.1   }
transform { function = log }
regression { variables = (const)
             user = (strike80  strike85  strike90)
             file = "strike.dat "
             format = "(3f12.0)"   }
arima { model = (0 1 1)(0 1 1)12    }
X11 { title = ("Car Sales in US - Adjust for strikes in 80,85,90")
      save = seasonal   appendfcst = yes
      x11easter=yes     }
X11regression {variables= td }
```

X11REGRESSION

내 용

결측치없는 시계열에 대해서 x11과 함께 사용하는 선택적 명령이다. 이 명령은 불규칙요인에 포함된 캘런더 효과를 프로그램에 의해 사전에 정의된 회귀변수 또는 사용자가 정의한 회귀변수를 이용하여 회귀모형으로 추정하기 위한 명령이다. 프로그램에서 사전에 정의된 변수는 캘런더 효과(요일효과와 명절효과)와 가법 특이치이며, variable 옵션에 의해 지정할 수 있다. 구조변화 (change of regime)는 요일효과 회귀변수와 함께 사용할 수 있다. 사용자가 정의한 우리나라의 명절(설・추석)과 같은 회귀변수는 user 옵션으로 모형에 포함할 수 있으며, 회귀변수의 data 값은 data나 file 옵션으로 불러올 수 있다. 즉, 불규칙요인에 의한 회귀모형은 사전에 정의된 회귀변수와 사용자가 정의한 회귀변수를 독립변수로 사용할 수 있다.

일반형식

x11**regression{ variables** = (td *or* td1coef *or* tdstock[31]
easter[8] labor[8] thank[1]
ao1967.apr)
user = (temperature precip)
start = 1955.jan
data = (25 0.1 …) *or*
file = "weather.dat"
format = "(2f5.1)"
tdprior = (1.4 1.4 1.4 1.4 1.4 0.0 0.0 0.0)
aictest = (easter user
td *or* tdstock *or* td1coef)
aicdiff = -2.0
span = (1980.jan, 1995.dec)
sigma = 2.75 or **critical** = 3.5
outliermethod = addone
outlierspan = (1973.may, 1992.sep)

usertype = holiday
prior = yes
print = (brief+b15)　　　　　　**save** = (c16　c18)
savelog = aictest
}

명　　령

aicdiff　aictest 옵션에서 회귀변수를 갖는 회귀모형과 갖지 않는 회귀모형의 AIC(혹은 AICC) 차이 값을 지정한다. default는 aicdiff=0.0이다.

aictest　지정된 회귀변수를 불규칙요인 회귀모형에 포함할 것인지 검정하기 위한 AIC 비교 검정을 지정한다. 이 명령은 td, tdstock, td1coef, easter와 user에서만 사용할 수 있다. 예를 들어, 요일효과 모형을 선정한다면, 요일효과 변수를 포함한 모형과 포함하지 않은 모형에서 AIC(시계열 길이의 수정이 필요한 경우는 AICC) 값을 구한다. default로서, 보다 작은 AIC 값을 갖는 모형이 시계열의 예측과 특이치 식별 등에 사용된다. 하나 이상의 회귀변수를 갖는 경우, 다음 순서로 AIC-검정이 순차적으로 실시된다. (a) 요일효과 변수 (b) 부활절효과 변수 (c) 사용자 정의효과 변수. 같은 형태의 회귀변수가 여러 개 사용되면(예: 여러 개의 요일효과 변수), aictest 절차는 그룹으로 적용된다. 즉, 이러한 형태의 모든 변수가 최종모형에 적용 되던가 또는 적용되지 않는다. 이 옵션을 지정하지 않으면, AIC를 기초로 한 자동 모형 선정은 실시되지 않는다.

critical　가법 특이치(극단의 불규칙 값)를 검색하기 위한 t-통계량의 절대값에 대한 임계치를 지정한다. 이 옵션은 sigma 옵션을 사용하지 않거나 flow 요일효과를 추정하는 회귀변수를 사용하는 경우 사용한다. 지정된 값은 0보다 큰 값이여야 한다. default 임계치는 특이치 탐색을 위해 사용된 관측치 수에 의해 결정된다(OUTLIER 명령의 표 “특이치 식별을 위한 default 임계치” 참조).

data 사용자가 정의한 회귀변수 값을 지정한다. 자료의 시점 영역은 시계열의 시점이나 series 명령의 span에 의해 지정된 시점을 포함해야 한다. 또한 forecast 명령에서 사용되는 예측 시점 영역도 포함해야 한다. 자료 값은 free format으로 user 옵션에서 사용자가 지정한 회귀변수 순서대로 읽는다. data 옵션은 file 옵션을 같이 사용할 수 없다.

file 사용자가 정의한 회귀변수의 자료 값을 포함하고 있는 파일 이름을 지정한다. 파일이 현재의 폴더가 아니면 경로를 지정해야 한다. data 옵션을 사용하면, 자료의 시점은 분석하는 시계열의 미래 및 과거 예측치의 시점을 포함해야 한다. file 옵션을 사용하면, data 옵션을 사용할 수 없다.

format file 옵션에서 지정한 회귀변수에 대한 자료 값의 파일 입력 형식이다. format 명령을 사용하지 않으면 자료는 free format으로 읽혀진다. format 명령을 사용하면, data 명령을 사용할 수 없으며 file 옵션만 사용할 수 있다. 입력 형식은 series 명령의 format 옵션의 형식과 동일하다.

outliermethod 특이치 식별 방법으로 method=addone 혹은 addall을 지정한다. default는 method=addone이며, sigma 옵션과 함께 사용할 수 없다.

start 사용자가 정의한 회귀변수의 자료 값들에 대한 시작 월(분기)이다. default는 시계열의 시작 월(분기)이다. 시계열의 시작 월(분기)을 사용하거나, series 명령의 span 옵션에 의해서 지정된 기간의 시작 월(분기)이 사용된다.

outlierspan 특이치를 찾기 위한 불규칙요인의 시작과 끝 시점을 입력한다. 입력 기간의 시작과 끝 시점은 시계열의 기간 내에 있어야 하며, 시작 시점은 끝 시점보다 선행하여야 한다. 시작 시점이나 끝 시점이 결측 값이 될 수 있으며, 결측 값은, outlierspan=(1976.jan,), series 옵션에서 사용된 기간이 default로 사용된다. 이 옵션은 sigma 옵션과 함께 사용할 수 없다.

print & save 선택 가능한 출력표는 다음과 같다.

<x11regression: default 출력표>

이 름	단축	저장	내 용
priortd	a4	+	사전 요일효과의 가중치와 요인
extremeval	c14	·	불규칙 회귀모형에서 추출된 극단의 불규칙요인 값
x11reg	c15	+	최종 불규칙 회귀계수와 진단 통계량
tradingday	c16	+	최종 사전요인효과의 요인과 가중치
combtradingday	c18	+	결합된 최종 요일효과요인
holiday	xhl	+	최종 명절효과요인
calendar	xca	+	최종 캘런더효과요인(요일 및 명절효과)
combcalendar	xcc	+	결합된 일별 가중치로 구한 최종 캘런더 요인
outlierhdr	xoh	·	임계치와 특이치 span을 포함한 특이치 검색을 위해 사용된 옵션
xaictest	xat	·	요일 및 명절효과에 대한 AIC-test 결과

prior 불규칙요인에 대한 회귀모형으로 구한 캘런더 요인을 계절조정하기 전에 추정하고, 사전조정 과정(prior=yes) 혹은 계절조정 과정(prior=yes)에서 적용할 것인지 지정한다. prior 옵션은 regARIMA 모형을 사용하거나 x11 명령에서 x11easter= yes옵션을 사용하면, 사용할 수 없다. 따라서 불규칙요인 회귀는 항상 계절조정하기 전에 실행한다.

sigma 요일효과 회귀를 실행하기 전에 불규칙요인의 극단값을 제거하기 위한 sigma 한계를 지정한다. 평균(승법은 1.0, 가법은 0.0)에서 (sigma*표준편차)보다 큰 불규칙요인 값은 극단 값으로 판정하며 제외한다. 이 옵션은 요일효과 변수 이외의 변수가 모형에 포함되어 있거나, critical 옵션을 사용되면 사용할 수 없다. 지정되는 값은 0보다 커야 하며 default는 2.5이다. 예: sigma=3.0

tdprior 사용자가 계절조정하기 전에 요일효과를 조정하기 위하여 월요일부터 입력하는 요일 가중치를 지정한다. 이들 가중치 합은 프로그램에 의해 7.0이 되도록 조정된다. 이 옵션은 승법 또는 log-가법모형에서 사용된다. 입력되는 값은 0 또는 0보다 커야 한다. 예: tdprior=(0.7 0.7 0.5 1.05 1.5 1.5 1.05) span 회귀모형의 계수를 추정하기 위해 사용될 불규칙요인의 기간을 지정한다. span 옵션에는 시작 시점과 끝 시점이 있다. 결측 값이 있는 경우, 입력 시계열의 시작 시점 혹은 끝 시점이 default로 이용된다. span=(1968.1,)의 경우, 불규칙요인 회귀모형 추정은 1968년 1월부터 시작하며 끝 시점은 입력 시계열의 끝 시점을 이용한다. 끝 시점을 *0.per*로 지정함으로써 분석하고자 하는 시계열의 가장 최근의 시점을 이용할 수 있다. 여기서 *per*는 월(분기)이며, 12월 이외의 월을 끝 시점으로 한다면, span=(, 0.per), 입력된 시계열의 최근 시점을 이용한다.

user 모형에 포함하고자 하는 사용자가 정의한 회귀변수의 이름을 지정한다. 주어진 이름은 프로그램 결과에서 추정된 계수를 식별하기 위하여 사용된다. data나 file 옵션에 의해 사용자가 정의한 변수에 대한 자료 값을 제공해야 한다. 사용자가 정의할 수 있는 회귀변수는 최대 52개이다.

usertype 사용자가 정의한 회귀변수에 의해 추정된 회귀효과를 지정한다. 사용자가 지정할 수 있는 회귀효과 형태는 다음과 같다; 요일효과(**td**), 스톡 요일효과(**tdstock**), 명절(**holiday, easter, labor** 등), 가법 특이치 (**ao**), 사용자 정의(**user**).
사용자가 정의한 여러 개의 회귀변수에 대해 하나의 효과를 나타낸다면, 하나만 정의한다(usertype=td). 사용자가 정의한 여러 개의 회귀변수에 대해 각각 다른 효과를 나타낸다면, 회귀변수 각각에 대한 효과를 정의한다. 예: usertype=(td td td holiday user).

variables 모형에 포함할 사전 정의된 회귀변수로 이들 변수에 대한 자료 값은 calender 함수와 같은 프로그램에 의해서 계산된다.

기타명령

eastermeans 장기간의 월별 평균을 easter[w]와 같은 부활절 변수의 계절조정에 사용할지 지정한다. eastermeans=yes 또는 =no로 지정한다.

b variables와 user 옵션에서 나타나는 순서대로 회귀모수의 초기 값 또는 불규칙요인 회귀모수의 고정된 값을 지정한다. b 옵션은 불규칙요인에 의한 회귀모형에 있는 모든 회귀 계수의 초기 값을 지정해야 하며 variables와 user 옵션 후에 지정하여야 한다. 결측 값에 대한 default 값은 0.1이다. 두 개의 회귀계수를 갖는 경우, b=(0.7,)은 b=(0.7, 0.1)과 같으나, b=(0.7)과는 다르다. 세 개의 회귀계수를 갖는 경우, b=(0.8, ,-0.4)는 b=(0.8, 0.1, -0.4)과 같다. 특정한 값을 고정된 모수로 정하고자 하는 경우, 다음과 같이 "f"를 사용한다. b=(0.7f, 0.1).

centeruser 사용자가 정의한 회귀변수에서 평균 또는 계절 평균을 뺀다. centeruser=yes이면, 사용자 정의의 회귀변수에 대한 평균이 회귀변수에서 제외된다. centeruser=seasonal이면, 각 월의 계절평균이 사용자가 정의한 각 회귀변수에서 제외된다. 이 옵션이 사용되지 않으면, 사용자 정의의 회귀변수는 적절한 방법에 의해 중심화된 것으로 가정한다.

forcecal 캘런더 조정요인을 명절요인과 요일요인의 곱(가법 모형은 합)으로 제약할 것 인지 지정한다. 제약을 한다면 forcecal =yes, 아니면 forcecal=no이다. default는 forcecal =no이다.

noapply x11 명령에 의해 계절조정 단계를 실행하기 전에, 원계열이나 최종 계절조정계열에서 제거하지 않을 회귀효과요인을 정의한다.

umdata 사용자 정의의 회귀변수에 대한 계수를 추정하기 전에 불규칙 계열 I_t(혹은 $Log\ I_t$)에서 제거하고자 하는 평균 조정값의 입력 행렬이다. 이 옵션과 umfile은 사전에 정의된 회귀변수에 대한 평균 함수가 사전에 정의된 회귀변수를 포함한 모형과 다른 경우 사용한다. 평균 조정함수는 조정형태(mode)에 따른다.

기타 X11REGRESSION에 대한 자세한 내용은 U.S. Census Bureau (2004) 참조한다.

<x11regression: 이외의 출력표>

이 름	단축	저장	내 용
extremevalb	b14	+	B반복의 불규칙회귀에서 제거된 불규칙요인 값
x11regb	b15	·	사전 불규칙 회귀계수와 진단 통계량
tradingdayb	b16	+	사전 요일효과요인과 가중치
combtradingdayb	b18	+	결합된 요일 가중치로 구한 사전 요일효과요인과 가중치
holidayb	bxh	+	사전 명절효과요인
calendarb	bxc	+	사전 캘런더 요인
combcalendarb	bcc	+	결합된 일별 가중치로 구한 사전 캘런더 요인
outlieriter	xoi	· +	특이치의 검색과 제거 등을 포함한 특이치 검색의 반복절차 결과
outliertests	xot	·	특이치 검색을 위한 반복 절차 시의 t-통계량
finaloutliertests	xft	·	최종 특이치 검색 반복 절차 시의 t-통계량
xregressionmatrix	xrm	+	불규칙 회귀변수와 관련된 값
xregressioncmatrix	xrc	+	불규칙 회귀모수 추정치의 상관행렬(print 옵션 사용 시), 공분산행렬(save 옵션 사용 시)

<x11regression에 사전 정의된 회귀변수>

변수	변수정의
td	6개 요일효과를 추정한다. 윤년효과는 변수변환(승법모형)에 의해 조정된다. td는 td1coef나 tdstock[ ] 옵션과 함께 사용할 수 없다.
td1coef	주말 및 주중 요일효과를 추정한다. td1coef는 td나 tdstock[] 옵션과 함께 사용할 수 없다.
tdstock[ω]	스톡자료에 있어 ω일에 대한 6개 스톡 요일효과를 추정한다. ω는 1일부터 31일까지의 범위 내에서 사용된다. 지정된 ω일이 해당 월의 길이보다 작다면, tdstock 변수는 해당 월의 마지막 일을 사용한다.
easter[ω]	부활절효과를 추정한다. ω는 1일부터 25일 사이의 값이다. 이외에도 labor[ω], thank[ω] 등도 사용 가능하다.
odate	모형에 가법 특이치를 포함한다.

참　고

X11REGRESSION 명령은 사전 조정효과가 조정되지 않은 불규칙요인 I_t로부터 캘런더 효과 또는 다른 사전 조정효과의 요인을 추정하기 위하여 사용된다. 불규칙요인에 의한 사전 조정요인의 추정은 다음과 같은 회귀모형을 설정하여 최소자승법(OLS)으로 한다.

$$I_t - 1.0 = \beta' X_t + e_t$$

여기서 X_t는 관심대상의 효과를 나타내는 회귀벡터이다. e_t는 회귀잔차로 독립이며 일정한 분산을 갖는다.

불규치요인에 의한 회귀모형에 의해 회귀계수에 대한 t-통계량, χ^2-통계량, AIC-통계량이 OLS 추정치로부터 구해진다. 그러나 이들 통계량으로부터 검정된 회귀효과에 대한 통계적 유의성은 regression 명령에서 regARIMA 모형으로 검정한 결과보다 신뢰성이 낮다. 즉, x11regression의 통계량은 사전조정변수의 효과가 없더라도, 사전조정변수가 유의적인 효과가 있는 것으로 결정을 내릴 가능성이 regression에 의해 구해진 통계량보다 높다.

x11regression은 주어진 시계열이 regARIMA 모형에 적합이 잘 되지 않는 경우에 유용하게 사용할 수 있다. x11regression은 x11 명령의 B와 C 계산과정에서 추정되는 불규칙요인을 이용하여 사전조정요인을 추정한다. spec file에 arima, automdl, pickmdl, regression 명령을 포함하면, x11regression에 의해 효과가 먼저 추정이 되고, 이들 사전 조정효과는 추정치 혹은 forecast와 backcast 등을 위해 사용된 시계열에서 제거된다. 다른 명령에 의해 계산된 결과 값은 프로그램에 의하여 계절, 추세, 캘런더 효과, 불규칙 요인을 구하기 위한 x11 명령의 두 번째 계산과정에 의해서 분해된다.

x11regression에 대한 자세한 내용은 U.S. Census Bureau(2004)를 참조하기 바란다.