

한국형 X11ARIMA 프로시저에 관한 연구

박유성¹⁾ 최현희²⁾

요약

X11ARIMA는 1965년 미국 센서스국에서 개발된 X11분석방법에 기초한 시계열 분석 방법으로 Dagum(1975)에 의해 개발되었다. 이 기법은 Dagum(1988)에 의하여 북미지역의 174개의 경제지수를 바탕으로 일부 기본모형이 수정·보완되어 오늘날에 이르고 있다. 최근에는 회귀모형과 ARIMA모형을 동시에 고려하여 특이치와 추세변환효과(outlier and Trend-change effects), 계절변동(seasonal effect), 그리고 달력효과(calendar effect) 등을 추정한 William 등(1995)과 Chen과 Findley(1995)의 X12ARIMA분석방법이 소개되었다. 그러나 위의 모든 기법들은 주로 북미지역의 경제지수를 기초로 하고 있다.

본 논문에서는 우리나라의 산업중분류에서 산출되는 102개(생산(27), 출하(27), 재고(27), 가동률(21))의 지수에 대한 우리나라의 표준 ARIMA모형을 제시하고, 우리나라에 적합한 이동평균항수를 제공하고자 한다. 그리고 우리나라의 설, 추석 등의 명절효과를 태양력으로 전환함과 동시에, 최근에 논의되고 있는 X12ARIMA에서 사용되는 회귀모형과 ARIMA모형을 동시에 고려하는 명절효과를 도출하고자 한다.

1. 서론

1.1. 계절조정의 필요성

경제 시계열은 기후, 생활습관, 명절 등으로 인해 매년 유사한 형태로 반복되는 비경기적 변동인 계절적 원인에 의해서 영향을 받는다. 시계열 속에 내재되어 있는 경기변동을 파악하는데 있어서, 계절변동의 영향을 제거하지 않을 경우 계절적인 증감 영향을 실제 경기변동으로 간주하여 경기변동을 과대 또는 과소평가하여 실제와는 전혀 다른 판단을 하는 경우가 발생하게 될 수 있다. 연간자료의 분석에서는 이러한 계절조정의 문제가 발생하지 않으므로 원계열을 그대로 사용할 수 있으나, 월별 또는 분기별 자료의 분석에서는 계절요인이 추세·순환요인과 결합하여 나타나므로 추세·순환요인을 명확하게 파악하기 위해서는 경기적 요인과 비경기적 요인을 분리하는 계절조정이 필요하다. 특히, 우리나라가 OECD의 회원으로 가입함으로써 모든 경제지수를 계절조정된 지수와 조정되지 않은 지수로 동시에 공표하여야 한다. 계절조정된 지수에 의해 그 나라의 경제현상을 파악하게 되므로 정교한 계절조정의 문제는 우리나라가 직면한 가장 시급한 문제중의 하나이다.

1) (136-701) 서울시 성북구 안암동, 고려대학교 통계학과, 교수

2) (150-010) 서울시 영등포구 여의도동, 한국아이비엠

1.2. 계절조정의 연혁

1920 ~ 1930년대 후반 미국의 NBER(National Bureau of Economic Research)을 중심으로 가변적인 계절조정방법에 대한 연구가 본격적으로 이루어졌다. 이 시기에 12개월 이동평균에 의한 계절조정방법을 개발하였고, 이 이동평균의 원리는 현재 각국에서 사용하는 계절조정방법의 기초가 되었다.

컴퓨터의 발달과 더불어 다양한 이동평균의 계산이 용이해짐에 따라 계절조정방법은 크게 발전하였다. 미국의 센서스국은 이동평균을 이용하여 결항처리, 요일변동추정 등을 연구하여 일련의 계절조정 프로그램을 개발하여 1965년 X11을 발표하였다. X11방법은 여러 상식과 경험을 근거로 작성된 일련의 복잡한 알고리즘으로, 계절변동 가변성을 인정하고 있다는 점과 완전 자동화된 알고리즘을 갖는다는 특성이 있다. 그러나 분석을 위해 사용하는 이동평균 때문에 몇 개의 초기 관찰치와 마지막 관찰치를 사용하지 못하는 약점을 가지고 있다. 1975년 캐나다 통계국의 Dagum은 이 결측치의 발생을 방지하기 위해 ARIMA절차에 의해 미리 1개년 또는 3개년의 과거치와 미래치를 예측, 보완하여 이를 X11ARIMA라고 하였다. 최근에는 회귀모형과 ARIMA모형을 동시에 고려하여 특이치와 추세변환효과, 계절효과, 그리고 달력효과 등을 추정한 X12ARIMA분석방법이 소개되고 있다.

이용사례를 소개하면, 일본, OECD 등은 X11, X11ARIMA 등을 사용하거나 계산과정의 일부분을 자국 실정에 적합하도록 변형한 방법을 사용하고 있고, 우리나라는 1973년 X11을 도입하여 사용하다가 79년부터는 X11ARIMA로 대체하여 현재까지 사용 중이다. 현재 X11ARIMA방법을 기초로 한 통계 프로그램은 X11ARIMA/80 과 X11ARIMA/88이 있다. X11ARIMA/80과 X11ARIMA/88의 주요 차이점은 계절조정시 필요한 양단의 계절연장을 위한 ARIMA모형과 이동평균에 관한 내용이다. X11ARIMA/88이 더욱 개선된 방법으로 본 논문은 X11ARIMA/88을 중심으로 전개되어질 것이다.

2. X11ARIMA 프로시저

2.1. ARIMA 모형의 선택

캐나다에서 개발된 X11ARIMA가 우리 여건에 적합한 방법인지의 여부에 대하여 의문이 제기되고 있으나 이에 대한 연구 및 검토는 다소 미흡한 실정이다. X11ARIMA/80의 경우 $L(011)(011)$ (원계열을 로그변환한 경우 $ARIMA(0, 1, 1) \times (0, 1, 1)_{12}$), $(011)(011)$, $(212)(011)$ 을, X11ARIMA/88의 경우 $L(011)(011)$, $L(012)(011)$, $L(210)(011)$, $L(022)(011)$, $(212)(011)$ 의 5개의 모형을 표준모형으로 제공하고 있으나, 이들 표준옵션으로 제공되는 모형들은 캐나다의 지수(174개)에 의해서 선정된 모형이므로 우리나라의 지수에 적합한 모형들과는 차이가 있을 수 있다.

본 절에서는 현재까지 문헌에 소개된 모형선정의 기준들을 제시한 후, 1980년 1월부터 1994년 12월까지 조사된 우리나라의 생산지수(27), 출하지수(27), 재고지수(27), 가동률지수(21)의 총 102개(자료출처:통계청)의 월별자료에 대한 최적 모형을 선정하고 X11ARIMA/88에서 제공하는 모형들과의 상이점을 설명하고자 한다.

2.1.1. 모형 선정의 기준

모형선정의 기준이 되는 통계량의 성질을 두 가지로 나누어 보면 적합잔차에 관한 것과 ARIMA모형의 차수에 관련된 것이 있다. 이 중에서 적합잔차와 관련한 통계량은 포트맨토우(Portmanteau)통계량, 절대오차평균(MAE), 절대오차퍼센트평균(MAPE), 평균제곱오차(MSE) 등이 알려져 있고, 모수의 차수에 관련한 통계량으로는 AIC, SBC, BIC, FPE, Parzen's CAT 등이 있다.

포트맨토우 통계량은 적합하지 않은 모형을 기각하는 기준으로는 우수한 통계량이나, 경쟁모형들 중 최적의 모형을 선택하는 기준으로서는 적당하지 않다. 그러므로 원계열 y_t 의 추정치를 \hat{y}_t 으로 표현할 경우 경쟁모형들 중 최적의 모형을 선택하는 기준으로 빈번히 사용되는 다음의 통계량들을 선택의 기준으로 추천한다.

(1) RMSE(Root Mean Square Error)

$$\text{RMSE} = \sqrt{\frac{1}{n-k} \sum_{t=1}^n (y_t - \hat{y}_t)^2}$$

(2) MAPE(Mean Absolute Percent Error)

$$\text{MAPE} = \frac{100}{n} \times \sum_{t=1}^n \left| \frac{y_t - \hat{y}_t}{y_t} \right| (\%)$$

(3) AIC

Akaike(1973)에 의해 제시된 통계량으로 어떤 시계열을 ARMA(p, q)모형에 적합시켰다고 가정하면 대수 우도함수는

$$\text{AIC}(p+q) = n \times \ln[\hat{\sigma}^2] + 2(p+q)$$

가 된다. 여기에서 $\hat{\sigma}^2$ 는 ARMA(p, q)의 오차항의 분산 추정치이다. AIC 기준에 의한 모형식별법은 AIC($p+q$)값을 최소로 하는 모형을 최적모형으로 판별하는 것이다. 이 AIC기준은 차수 (p, q)를 효과적으로 배당함으로써 과다추정을 막기 위하여 개발된 것이다.

2.1.2. 모형 선정기준에 따른 최적 모형

2.1.1절에서 제시한 기준들을 사용하여 다음의 절차에 따라 102개 지수의 최적모형을 선정하였다. 먼저 lag 36까지의 χ^2-p 값이 0.1보다 큰 모형들 중 비교적 큰 p 값을 가지는 모형을 간추린 결과 총 9개의 후보모형을 얻을 수 있었다. 이 9개의 후보모형에 대하여 RMSE, MAPE, AIC값을 비교한 후 가장 작은 RMSE, MAPE, AIC값을 갖는 모형을 대상지수에 적합한 모형이라 판단하였다. 지수 102개에 대한 모형 적합결과는 표 2.1과 같다. 표 2.1의 적합결과는 크게 ARIMA모형과 지수평활법모형 두 가지로 나타나고 있으며 ARIMA모형 중 특히 관심을 두어야 할 것은 X11ARIMA에서 표준옵션으로 제공되는 모형들이다.

표 2.1: 모형 적합결과

모형	생산지수		출하지수		재고지수		가동률지수		전체	
	RMSE AIC	MAPE	RMSE AIC	MAPE	RMSE AIC	MAPE	RMSE AIC	MAPE	RMSE AIC	MAPE
(011)(011) 지수평활법	17	14	17	16	23	23	14	10	71	63
L(110)(011)	2	3	3	6	1	0	1	5	7	14
L(011)(011) X11ARIMA	4	1	3	2	1	1	3	1	11	5
L(011)(210)	3	3	3	0	0	0	2	4	8	7
L(012)(011) X11ARIMA	0	3	1	1	0	0	0	1	1	5
L(110)(012)	1	1	0	0	0	2	0	0	1	3
L(110)(210)	0	0	0	1	0	0	1	0	1	1
L(210)(011)	0	1	0	1	0	0	0	0	0	2
L(110)(011) X11ARIMA	0	1	0	0	2	1	0	0	2	2
계	27	27	27	27	27	27	21	21	102	102

표 2.2: MAPE의 평균값(%)

통계	생산	출하	재고	가동률	전체
평균값	5.396	6.244	6.459	6.223	6.073
표준편차	2.421	3.095	4.614	3.314	3.436

시계열의 분석방법으로는 회귀분석방법, 분해방법, 지수평활법, 그리고 박스-젠킨스(Box-Jenkins)방법의 네 가지가 있다. 분해방법 중 X11기법이 대표적인 방법이며, X11기법은 특히 시계열의 계절인자가 느리게 변동할 때 사용되는 방법이다. X11에서 이동평균을 취함으로써 발생하는 양단의 결항치를 보정하기 위하여 전후 3개년치를 ARIMA방법에 의하여 예측하게 되어 있는데 ARIMA방법은 모든 인자가 빠르게 변동할 때 사용하는 기법으로 X11의 느리게 변동한다는 성질과는 상충하는 성질을 가지고 있다. 그리하여 본 논문에서는 X11ARIMA에서 제공하는 ARIMA방법 외에 통계적 추론에 바탕을 두지 않는 경험적 접근방법인 지수평활법을 고려하게 된 것이다.

표 2.1의 결과를 요약하면 총지수 102개 중 RMSE, AIC 기준에서 69.6%, MAPE 기준에서 61.8%가 지수평활법에 의한 모형으로 적합되는 것으로 나타났다. 반면 X11ARIMA에서 표준모형으로 제공하고 있는 ARIMA모형들은 RMSE, AIC기준에서 13.7%, MAPE기준

에서 11.7%의 작은 비율로 나타났다. 그러므로 우리나라의 생산, 출하, 재고, 가동률지수는 X11의 이동평균을 취함으로써 발생하는 양단의 결함을 지수평활법을 사용하여 보정할 경우 더욱 개선된 계절조정결과를 얻을 수 있을 것으로 기대된다. 또한 X11ARIMA/88에서 제시한 표준모형을 MAPE기준에 의하여 (011)(011), L(110)(011), L(011)(011), L(011)(210), L(012)(011), L(110)(012)로 수정할 것을 제안한다. 지수들의 MAPE 평균값(%)들은 표 2.2와 같다. 전체적으로 MAPE는 약 6%를 나타내었다.

2.2. 계절조정을 위한 이동평균

X11분석방법이 다른 시계열적 접근에 대해 가지는 다른 특징은 이동평균의 분해기법을 적용하여 계열의 구성요인인 T_t (추세), S_t (계절), C_t (순환), I_t (불규칙)변동들을 이끌어 내는 것이다. X11절차를 그대로 소개하지는 않겠지만, 계절변동의 추정치를 구하는 과정에서 사용되는 몇 가지 이동평균의 방법을 소개하고, 앞장에서 사용한 102개의 지수에 대하여 우리나라와 같이 불규칙변동이 강한 계열에 적용되어야 하는 이동평균의 절차를 이끌어 내고자 한다.

2.2.1. X11ARIMA/88에서의 이동평균

X11ARIMA절차는 가중 5개항(3×3), 가중 7개항(3×5)의 이동평균을 기본적으로 적용한다. 가중 5개항 이동평균은 월별로 3개항 이동평균한 계열을 다시 3개항 이동평균하는 방법이다. 마찬가지로 가중 7개항 이동평균은 월별로 5개항 이동평균한 계열을 다시 3개항 이동평균하는 방법이다. 가중 5개항 이동평균은 일정한 방향으로 급격한 변화를 하는 시계열에 대해 매우 신축성있게 적용되지만, 이동기간이 짧아 계절·불규칙요인이 잘 평활화 되려면 불규칙요인이 작아야 한다. 반면에 평활너비가 넓은 3×5 , 3×9 이동평균은 3×3 이동평균보다 불규칙변동이 많은 경우, 즉, 3×3 이동평균을 그대로 사용하면 불규칙계열과 계절계열이 혼합되어 평활의 효과를 볼 수 없을 경우에 사용된다. 계절변동의 추정치를 구하기 위하여 X11ARIMA의 [절차2]에서는 기본적으로 4번의 이동평균, 즉, 3×3 , 3×3 , 3×3 , 3×5 의 차례로 이동평균을 산출한다. 이 중 첫 번째와 세 번째 이동평균은 동일절차의 반복이므로 이동평균의 형태가 항상 일치한다. 이 때 3×3 , 3×5 의 차례는 캐나다 지수에 의하여 경험적으로 얻어낸 이동평균의 절차이므로 불규칙 변동이 강한 우리나라의 지수에 이 절차를 그대로 적용하는 데에는 무리가 따르게 마련이다.

그러므로 본 논문에서는 표 2.3과 같은 방법으로 X11프로그램을 일부 수정하여 각 지수 내에서 잘 적합되는 이동평균을 구하고자 한다.

2.2.2. 이동평균의 선택

계절조정의 결과를 평가하기 위하여 X11에서는 D8(최종 수정 안된 계절·불규칙요인)표에 대하여 안정계절성에 대한 F검정, 크루스칼-왈리스 카이제곱 검정, 이동계절성에 대한 F검정결과를 제공한다. 안정계절성 F값과 크루스칼-왈리스 카이제곱 값은 값이 클수록 선택된 이동평균이 다른 이동평균보다 계절변동의 분해를 용이하게 한다는 것을 의미

표 2.3: 이동평균 적합단계

$3 \times 3, 3 \times 3, 3 \times 3, 3 \times 5(3335)$ (1) $3 \times 5, 3 \times 3, 3 \times 5, 3 \times 5(5355)$ 의 이동평균을 적용한다. $3 \times 9, 3 \times 3, 3 \times 9, 3 \times 5(9395)$ (2) 두번째 이동평균 3×3 을 $3 \times 5, 3 \times 9$ 의 차례로 바꾸면서 (1)의 차례를 반복한다. (3) 각 지수(생산, 출하, 재고, 가동률) 계열별로 총 9개의 이동평균모형에 의한 계절조정 산출 결과를 비교한다. (4) 9개의 이동평균모형 중 조정결과가 가장 좋은 모형에 대하여 네번째 이동평균 3×5 를 $3 \times 3, 3 \times 9$ 으로 수정한 후 마지막 3개의 모형 중 최적의 모형을 선택한다.
--

한다. 한편 이동계절성에 대한 검정 F값은 값이 클수록 사용된 방법이 다른 방법에 비해 이동평균 결과 계절변동과 불규칙변동이 혼합되어 있는 경우이다. 그러므로 이동계절성 F값은 값이 작을수록 적용한 이동평균이 계절변동과 불규칙변동을 잘 분해하고 있다는 것을 의미한다. 또 X11에서는 M1-M11 및 Q통계량이 제공되는데 모두 0~3의 값을 가지며 0에 가까울수록 시계열의 계절조정은 잘 되었다고 평가할 수 있고 통계량이 1을 초과할 때는 계절조정방법을 다시 고려해 보아야 할 경우가 많다.

특히 Q통계량은 M1-M11까지의 가중평균으로 계절조정에 대한 종합평가를 위해 사용한다. 11개 M통계량의 가중치는 표 2.4와 같다.

표 2.4: M1-M11통계량의 가중치

통계량(Mi)	M1	M2	M3	M4	M5	M6	M7	M8	M9	M10	M11
가중치(wi)	0.13	0.13	0.10	0.05	0.11	0.10	0.16	0.07	0.07	0.04	0.04

이 때 $Q > 1.0$ 이면 계절조정지수를 받아들일 수 없다. Q통계량이 계절조정지수에 대해서 일반적인 평가를 해주지만 작은 변화에도 유의적인 결과를 줄 수 있으므로 주의해야 한다. 즉 $Q > 1.0$ 일 때 M통계량을 주의 깊게 관찰한 후 가중치가 큰 M1, M2, M3, M5, M6, 그리고 M7에 의해 원인을 교정해 주는 절차를 밟아야 한다. 가중치가 큰 M통계량은 주로 불규칙변동과 추세순환변동에 의해 발생하게 된다. 불규칙변동은 명절효과, 경제환경의 급격한 변화, 거래일효과 등에 의해 조절될 수 있고, 추세순환변동은 헨더슨 이동평균(Henderson Moving Average)을 조정함으로써 조절할 수 있다. 불규칙변동에 관한 사항은 2.3절에서 다시 논의될 것이며, 추세순환변동에 대한 연구는 추후과제로 남겨 놓는다.

표 2.5, 표 2.6, 표 2.7에 차례대로 안정계절성에 대한 F검정 결과, 크루스칼-왈리스 카이 검정 결과, 이동계절성에 대한 F검정결과를 사용한 이동평균방법별로 제시하였다. 표 2.8은 안정계절성에 대한 이동계절성의 상대적 크기(M7)를 이동평균방법별로 평균한 것이다. 표 2.5 ~ 표 2.8에서 ()안의 값은 표준편차를 의미하며, 이동평균 9993은 $3 \times 9, 3 \times 9, 3 \times 9$,

3 × 3의 이동평균을 의미한다.

표 2.5 ~ 표 2.8을 살펴보면, 원래의 이동평균인 3335는 우리나라의 경제지수에 잘 적합되지 않고 있음을 쉽게 알 수 있다. 자세히 설명하면, F값이 9993 또는 9999의 경우 계절변동이 안정적이고, 표 2.7의 이동계절성에 대한 F값은 9999가 가장 좋게 나타났으며 표 2.8에서 살펴본 M7값으로부터 9995 또는 9993의 이동평균을 취할 경우 의미있는 조정결과를 얻을 수 있을 것으로 보인다. 이러한 경향은 표 2.9의 Q통계량을 지수별로 분리한 결과에서도 확인할 수 있다. 즉 생산지수의 경우 9993, 출하지수 9999, 재고지수 9995, 그리고 가동률지수의 경우는 9999가 가장 작은 Q값을 나타냈다. 특히 가동률의 경우 이동평균 항수가 클 경우 적합이 잘 되는 것은 사실이지만 평균적으로 X11ARIMA 절차에서 얻은 계절조정지수를 받아들일 수 없는 것으로 나타났다.

표 2.5: 안정계절성에 대한 F값

F값평균	31.96 (37.49)	31.91 (37.42)	31.83 (37.35)	31.80 (37.27)	31.78 (37.12)	29.96 (35.07)	29.89 (35.05)	29.80 (34.96)	28.08 (30.90)	28.04 (32.80)	27.89 (32.81)
이동평균	9993	9595	9995	9999	9395	5955	5555	5355	3935	3535	3335

표 2.6: 안정계절성에 대한 크루스칼-왈리스 카이제곱값

카이제곱	95.75 (36.12)	95.55 (36.33)	95.54 (36.59)	95.54 (36.39)	95.26 (36.88)	93.40 (3.52)	93.25 (36.55)	92.92 (36.85)	90.84 (36.67)	90.79 (36.71)	90.54 (46.80)
이동평균	9999	9995	9595	9993	9395	5955	5555	5355	3535	3935	3335

표 2.7: 이동계절성에 대한 F값

F값평균	1.959 (1.24)	1.955 (1.22)	1.941 (1.22)	1.821 (1.07)	1.812 (1.08)	1.809 (1.07)	1.715 (0.92)	71.704 (0.88)	1.699 (0.88)	1.698 (0.89)	1.697 (0.90)
이동평균	3935	3335	3535	5355	5955	5555	9395	9993	9995	9595	9999

이는 표 2.10에서 볼 수 있듯이, 추세순환변동에 대한 헨더슨 이동평균의 조정이 필요하거나 명절효과와 같은 불규칙요인의 영향력이 매우 커서 안정적인 계절조정지수를 얻어내기 힘들다는 것을 의미한다.

2.3. 명절효과

우리나라의 대부분의 경제시계열 자료들은 태음력의 명절효과에 의해 많은 영향을 받는다. 명절일자가 월초, 월중순 또는 월말에 오느냐에 따라 계절의 움직임에 주는 영향력의 정도가 달라지기도 하고, 해를 거듭함에 따라 명절효과가 상수로 작용하지 않기 때문에 시계열의 구성분해에서 불규칙요인에 나타난다. 따라서 명절일자의 움직임과 각 시계열의 특

표 2.8: 안정계절성에 대한 이동계절성의 상대적 크기

M7	0.7376 (0.473)	0.7324 (0.465)	0.7321 (0.466)	0.7018 (0.441)	0.6959 (0.432)	0.6956 (0.430)	0.6737 (0.425)	0.6716 (0.420)	0.6713 (0.423)	0.6695 (0.417)	0.6693 (0.416)
이동평균	3335	3935	3535	5355	5555	5955	9395	9595	9999	9995	9993

표 2.9: Q통계량의 평균과 표준편차

이동평균	3335	3535	3935	5355	5555	5955	9395	9595	9993	9995	9999
생산	0.792 (0.358)	0.788 (0.355)	0.794 (0.351)	0.778 (0.349)	0.780 (0.347)	0.778 (0.338)	0.761 (0.344)	0.753 (0.328)	0.750 (0.329)	0.753 (0.326)	0.751 (0.324)
출하	0.788 (0.350)	0.787 (0.347)	0.784 (0.346)	0.768 (0.338)	0.767 (0.337)	0.768 (0.340)	0.754 (0.343)	0.748 (0.332)	0.746 (0.337)	0.748 (0.334)	0.745 (0.331)
재고	0.908 (0.377)	0.894 (0.380)	0.897 (0.378)	0.863 (0.370)	0.843 (0.369)	0.841 (0.372)	0.818 (0.358)	0.809 (0.359)	0.804 (0.361)	0.801 (0.359)	0.804 (0.363)
가동률	1.066 (0.301)	1.062 (0.296)	1.067 (0.296)	1.055 (0.300)	1.046 (0.294)	1.043 (0.296)	1.036 (0.303)	1.039 (0.312)	1.028 (0.311)	1.031 (0.310)	1.026 (0.309)

성을 고려하여 각기 적합한 사전월조정요인(Prior Monthly Factors)을 산출하고 이를 원계열에 적용하여 명절효과가 제거된 안정적인 원계열을 만들어야 한다. 명절효과추정을 위한 방법으로는 단순평균법, OECD방법, Fisher 와 Pfefferman(1981)방법, Kendall 와 Oller방법, Pierre Cholette(1974)방법, X11ARIMA/88의 사전월요인 추정방법 등이 있다.

X11ARIMA/88의 사전월요인 추정방법은 부활절효과에 관한 것으로, 예를 들어 1(9)월 말에 명절이 있는 경우 그 영향은 2(10)월초까지 미칠 것으로 생각하여 명절월(1·9월)의 월말과 명절이동월(2·10월)의 월초에 대하여 명절의 파급효과 기간동안 명절일을 기준으로 명절효과가 점점 감소하는 가중치를 부여하여 명절의 효과를 추정하였다. 현재 통계청에서는 불규칙요인을 단순3등분 평균법으로 사전조정요인을 추정하고 있는 것으로 알려져 있다.

2.3.1. 명절효과에 대한 기본 개념

일반적으로 명절일에 가까울수록 명절효과의 지수에 대한 파급효과가 크고, 명절일에서 멀리 떨어져 있을수록 그 효과는 점점 작아지는 경향이 있다. 그리고 명절이 월초, 월중순, 또는 월말의 어느 시점에 위치하는가에 따라 각 월에 미치는 영향이 달라지기도 한다. 이 효과는 그림 2.1로 설명될 수 있다.

설의 경우를 예로 들면, 설이 2월중에 있을 경우 1월에 미치는 효과는 (1)의 어두운 부분과 같고, 설과 월말이 동일한 날에 위치할 경우 1월에 미치는 효과는 (2)의 어두운 부분과 동일하다. (3)은 설이 1월중에 있는 경우이다. 삼각형의 좌우 끝점은 명절효과가 사라지는 점, 즉 명절파급효과가 지속되는 마지막 날이다.

표 2.10: 가동률지수에 대한 M통계량 평균값

분류	M1	M2	M3	M4	M5	M6	M7	M8	M9	M10	M11
평균 (std)	1.926 (2.00)	0.973 (0.75)	0.434 (0.37)	0.414 (0.44)	0.637 (0.58)	0.324 (0.23)	0.695 (0.44)	1.007 (0.45)	0.547 (0.26)	1.094 (0.57)	1.041 (0.62)

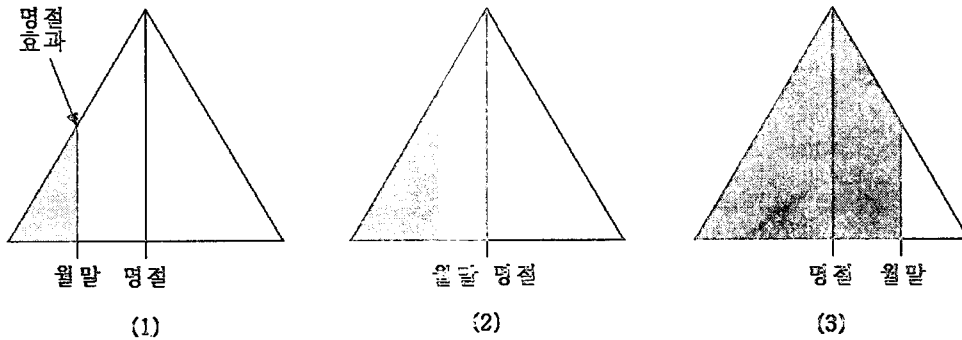


그림 2.1: 명절의 위치에 따르는 명절효과

위의 그림을 고려하여 명절과급효과기간의 길이에 따라 가중치를 달리 줄 수 있다. 통계청은 명절이전 7일과 명절이후 7일을 제안하였고, 한상태(1991)는 명절효과는 10일 전후가 가장 좋은 것으로 제안하였다. 본 논문에서는 명절 이전과 이후 각 10일을 영향을 주는 기간으로 선택하였다. 이 때 일별 가중치는 그림 2.2와 같이 표현될 수 있다.

1981년 설의 경우를 예로 들어보자. 양력날짜는 2월 5일, 1월말은 31일이므로 정확히 명절일의 10일 이전은 1월 27일, 10일 이후는 2월 14일이다. 이런 경우 1월에는 명절전 5일이, 2월에는 명절전 4일 명절후 9일이 속하므로 가중치는 각각

$$\text{가중치(1)} = 1/10 + 2/10 + 3/10 + 4/10 + 5/10$$

$$\text{가중치(2)} = 6/10 + 7/10 + 8/10 + 9/10 + 9/10 + 8/10 + \dots + 1/10$$

이 된다.

2.3.2. 명절효과의 적용

2.3.1의 그림 2.2에 의하여 산출한 1980년 1월부터 1994년 12월까지의 월별효과는 표 2.11과 같다. 설과 추석의 각각의 효과를 표준화하기 위해 최대효과를 1로 정하면 표 2.11의 값들을 9로 나누어 줌으로써 실제의 가중치를 얻을 수 있다. 1, 2, 9, 10월을 제외한 나머지 달에 대해서는 가중치는 0이다.

위의 월별효과를 불규칙요인을 설명해 줄 수 있는 변수들의 값이라고 생각하여 개입분

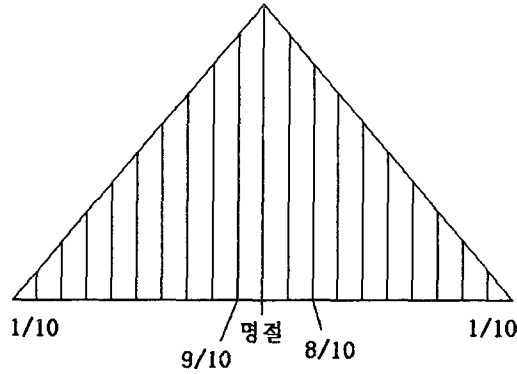


그림 2.2: 명절위치에 따르는 일별 가중치

표 2.11: 명절의 월별효과

년도	80	81	82	83	84	85	86	87	88	89	90	91	92	93	94
1월	0.0	1.5	8.4	0.0	3.6	0.0	0.1	6.2	0.0	1.0	7.5	0.0	2.1	8.9	0.0
2월	9.0	7.5	0.6	9.0	5.4	9.0	8.9	2.8	9.0	8.0	1.5	9.0	6.9	1.0	9.0
9월	8.7	9.0	4.5	9.0	9.0	5.4	9.0	6.0	8.0	9.0	2.8	8.9	9.0	4.5	9.0
10월	3.0	0.0	4.5	0.0	0.0	3.6	0.0	8.4	1.0	0.0	6.2	1.0	0.0	4.5	0.0

석 모형을 선정하면

$$\hat{N}_t = \beta_0 + \beta_1 A_{1t} I_{[1,2]} + \beta_2 A_{2t} I_{[9,10]} + \epsilon_t,$$

에에서 A_{1t} = 음력설에 해당하는 영향변수,

A_{2t} = 추석에 해당하는 영향변수,

$\epsilon_t \stackrel{iid}{\sim} N(0, \sigma^2)$, 그리고 $I_{[i]}$ 는 지시함수,

$$\hat{N}_t = Y_t(\text{original}) - \hat{Y}_t(\text{forecast})$$

와 같다. ARIMA모형 적합 후 얻어 낸 잔차(residual)를 종속변수로 하고 A_1, A_2 를 설명변수로 하여 회귀분석한 결과는 표 2.12와 같다. 참고로 분석에 사용된 생산, 출하, 재고, 가동률 지수는 각 지수영역에서 무작위로 추출된 것임을 밝힌다.

그림 2.2에서 제안한 모형에 의하여 설효과 또는 추석효과가 출하지수를 제외한 모든 지수에서 유의하므로 원계열에 대한 명절효과에 대한 조정이 필요함을 알 수 있다.

가동률의 예로 사용했던 ‘의료, 정밀, 광학기기 및 시계제조업’의 가동률지수의 원계열을 사전조정 한 후 명절효과에 의한 변화율은 표 2.13과 같다.

표 2.13에 의하면 80년 1월에는 명절에 의한 영향을 받지 않았으나, 동년 2월에는 설 명절에 의하여 가동률이 원래의 추이에 비하여 7.4% 만큼 낮아져 92.6%의 가동률을 달성하

표 2.12: 명절효과 회귀분석

영역		회귀계수	t값	p값	DW	R ²
생산	설	-64.569	-3.304	0.0120	2.054	0.0625
	추석	-9.780	-0.499	0.6184		
출하	설	-20.965	-0.846	0.3987	1.869	0.0045
	추석	1.803	-0.073	0.9423		
재고	설	0.002	0.000	0.9998	1.931	0.0321
	추석	17.887	2.317	0.0217		
가동률	설	-66.966	-1.743	0.0833	1.758	0.1074
	추석	-163.904	-4.252	0.0001		

였음을 알 수 있다.

표 2.13: 조정 후의 변화율(원계열/조정계열×100 %)

년도	80	81	82	83	84	85	86	87	88	89	90	91	92	93	94
1월	100	98.9	93.2	100	97.4	100	100	95.9	100	99.4	94.1	100	98.7	93.7	100
2월	92.6	93.8	99.5	93.6	96.4	93.6	94.0	98.5	95.3	95.0	99.0	93.1	94.7	100	93.6
9월	89.1	86.2	93.9	87.8	88.0	93.6	88.2	99.4	90.3	85.8	95.6	86.7	84.2	93.2	86.9
10월	99.6	100	93.3	100	100	95.4	100	88.5	98.5	100	88.7	99.9	100	93.7	100

조정후 계열을 X11ARIMA 절차에 적용시킨 결과는 표 2.14와 같다. 정상화된 원계열의 분산에 대한 불규칙요인의 상대적 기여도를 설명하는 M2, 연의 평균 연속기간에 의한 불규칙요인의 자기상관 M4, 불규칙요인 변화율 보다 큰 추세·순환요인의 변화율의 월간격 M5의 값이 낮아져 개입효과모형에 의한 명절효과제거에 의하여 불규칙요인이 많이 안정화되었음을 알 수 있다. 그리고 안정계절성에 대한 이동계절성의 상대적 크기인 M7이 낮아졌으므로 계절성을 이전 보다 더 잘 식별할 수 있음을 알 수 있다. M7의 분석결과 계절성을 인식할 수 없다면 계절조정을 하지 않는 것이 좋은 것으로 알려져 있다. 무엇보다도 전체적인 M통계량의 가중평균인 Q값이 조정결과 낮아졌음을 알 수 있다.

그러나 표 2.14에서 볼 수 있듯이 추세·순환변동에 밀접한 관계를 가지고 있는 M1, M3, M5의 값이 큰 것으로 보아 추세·순환변동요인을 산출하는 이동평균법에 대한 연구가 필요함을 알 수 있다. 또한 본 논문에서 사용하고 있는 중분류 경제지수는 소분류 경제지수를 기초로 하고 있다. 그러므로 Q값이 큰 중분류 경제지수는 소분류 경제지수의 계절조정을 본 논문에서 제시하는 방법에 의해 조정하여 그 효과를 검증할 필요가 있다.

본 논문에서 제시한 명절효과 조정방법은 기존의 X11ARIMA방법 또는 기존의 연구에

표 2.14: 사전 조정 결과 비교

		Q값	M1	M2	M3	M4	M5	M6	M7	M8	M9	M10	M11
기본모형 (3335)	조정전	1.38	3.000	2.088	1.160	0.161	1.392	0.449	0.968	1.343	0.710	1.485	1.358
	조정후	1.26	3.000	1.854	0.905	0.023	1.334	0.256	0.883	1.280	0.754	1.339	1.241
최적모형 (9999)	조정전	1.46	3.000	2.336	1.339	0.253	1.586	0.720	0.962	1.338	0.670	1.460	1.303
	조정후	1.24	3.000	1.883	0.983	0.115	1.308	0.548	0.774	1.096	0.631	1.234	1.110

서 제시된 방법과 두 가지 점에서 차이가 있다. 첫째, X11ARIMA 방법과는 순서상의 차이이다. X11ARIMA에서는 명절효과(서양의 부활절효과)를 X11ARIMA 절차중 D13에 나타난 최종불규칙요인에서 산출하게 된다. 이것은 이미 이전에 이동평균에 의하여 평활한 계열로부터 얻어내는 효과이기 때문에 그 효과의 신뢰성이 떨어질 수밖에 없다. 그러므로 시계열의 분해절차에 들어가기에 앞서 원계열에 대하여 명절효과를 수정해야 계절적인 증감 영향을 경기변동으로 평가하는 오류를 범하지 않을 수 있다는 것이다. 둘째, 기존의 문헌들에 소개되었던 방법들과의 차이점은 명절이전·이후 효과에 대하여 날짜별 파급효과를 달리 고려하였다는 것이다.

3. 결론

한국의 경제시계열 자료들로부터 경제지수들을 산출해내는 과정에서 X11ARIMA에서 제공하고 있는 표준모형을 그대로 사용함으로써 인해 발생하는 문제점들을 살펴보고, 이동평균법의 선택과 태음력의 영향을 받는 명절효과를 사전조정함으로써 올바른 계절조정지수들을 이끌어내고자 하였다.

생산, 출하, 재고, 가동률의 4가지 종류의 지수들을 SAS/ETS의 ARIMA 절차를 이용하여 적합한 결과 총 102개 자료 중 63(61.8%)개 자료셋이 MAPE기준에 의하여 지수평활법의 모형에 잘 적합되는 것으로 나타났으며, 기존의 X11ARIMA 모형에 적합되는 것은 12개(11.7%)에 불과했다. 우리나라의 표준모형으로 (011)(011), L(110)(011), L(011)(011), L(011)(210), L(012)(011), L(110)(012)를 제안한다. 이는 MAPE기준에 의하여 총 102개의 지수 중 95%가 적합되는 모형이 된다.

X11ARIMA/88 표준옵션에서 제공하는 이동평균의 항수가 우리나라 지수에 알맞게 적합되는가 하는 문제에서, 대부분의 계열이 불규칙변동에 큰 영향을 받고 있었기 때문에 작은 항수의 이동평균으로는 매끄러운 결과를 얻을 수가 없었다. 우리나라의 경우 생산지수는 9993, 출하지수 9999, 재고지수 9995, 그리고 가동률지수는 9999의 이동평균을 선택하는 것이 바람직하다.

현재까지 시도되어 온 명절효과의 산출방법들은 태음력의 영향을 받는 달력효과에 의하여 명절효과를 제대로 이끌어 내지 못했다. 본 논문에서는 명절효과를 산출함에 있어 음력달력을 양력달력으로 환원한 후 명절이전과 이후 각 10일씩을 고려하여 명절의 위치에

따라 월별로 상이한 가중치를 적용하였다. 그 결과 임의로 선정한 가동률의 3D33000(의료, 정밀, 광학기기 및 시계제조업)계열이 본 논문에서 선정한 모형에 의하여 잘 설명되는 것으로 나타났다.

4. 프로그램

본 논문에 소개된 새로운 접근방법들은 모두 화면상에서 사용자가 자동으로 처리할 수 있도록 X11ARIMA를 수정하였다.

(1) 표준모형

X11ARIMA/88은 L(011)(011), L(012)(011), L(210)(011), L(022)(011), (212)(011)을 표준모형으로 제공하고 있으나, 본 논문에 소개된 6개의 새로운 모형 (011)(011), L(110)(011), L(011)(011), L(011)(210), L(012)(011), L(110)(012)를 표준모형으로 수정하였다.

(2) 이동평균

입력단계에서 옵션으로 1.생산, 2.출하, 3.재고, 4.가동률을 입력하여 종류별 선택이 가능하게 하였다. 기본으로는 0을 주도록 하였는데, 이는 지수구분을 하지 않을 경우 5555의 모형 등을 적합에 이용하는 것이다.

(3) 명절효과에 의한 가중치

사용자가 시작해와 마지막해를 입력한 후 설정기간 동안의 음력달력을 양력달력으로 환원하여 각 월에 주어야 할 가중치 테이블을 출력하도록 하였다.

수정된 프로그램이 10,000라인에 달하므로 본 논문에서는 게재하지 않는다.

5. 제언

본 논문에서 다루어진 절차들은 우리나라의 산업중분류 지수 102개를 기준으로 수정·보완되었다. M1~M11과 Q값이 수정에 의하여 향상되었으나 다소 큰 값을 나타내어, 소분류를 중심으로 모든 절차를 다시 수행하는 것도 의미있는 결과를 이끌어 낼 수 있는 한 방법으로 생각되어진다.

명절효과의 형태에 있어서 본 논문에서는 효과가 명절일을 중심으로 지속적으로 증가하다가 명절후에는 지속적으로 감소하여 사라지는 형태를 제안하였으나, 명절을 중심으로 동일한 효과를 가지는 모형과 비교하여 어느 형태가 우리나라 지수에 더 잘 적합하는가에 대하여는 검증하지 않았다. 그러므로 두 모형 더 나아가서는 전·후효과를 달리 고려하는 형태도 비교의 대상이 될 수 있을 것이다. 명절전·후 영향을 주는 기간으로 본 논문에서는 이전과 이후 각 10일을 고려하였으나, 3, 5, 7과 같이 다양한 다른 방법들과도 비교가 필요하다. 마지막으로 본 논문에서는 계절변동의 조정만을 고려하였으나, 추세변동의 수정에 대한 연구가 요구된다.

참고문헌

- [1] 박유성, 허명희 (1996). <개정판 시계열자료분석>. 자유아카데미.
- [2] 오광우 (1996). <경제통계입문>. 자유아카데미.
- [3] 최기헌, 이종협 (1994). <SAS/ETS를 이용한 시계열분석과 그 응용>. 자유아카데미.
- [4] 통계조사국 통계분석과 (1995). <계절조정방법에 관한 해설서(X11AR IMA/88을 중심으로)>.
- [5] 한상태 (1991). 한국 경제 시계열의 계절조정 - 태음력 명절 효과의 조정을 중심으로 -. <고려대학교 통계학과 석사학위 논문>.
- [6] 허명희 (1993). <탐색적 방법에 의한 통계자료분석>. 자유아카데미.
- [7] Akaike, H. (1973). Information theory and an extension of the maximum likelihood principle. *Proc. 2nd International Symposium Information Theory*. 267-281. Akademiai Kiado. Budapest.
- [8] Akaike, H. (1978). A Bayesian analysis of the minimum AIC procedure. *Annals of the Institute of Statistical Mathematics*. 30A. 9-64.
- [9] Chen, B. and Findley, d.F. (1995). The X-12-ARIMA Projection Collaborators at the U. S. Bureau of the Census. *1995 Annual Research Conference and Seasonal Adjustment Workshop Advance Program*. Bureau of the Census, USA.
- [10] Dagum, E. B. (1980). *The X-11-ARIMA Seasonal Adjustment Method*. Time Series Research and analysis Division, Statistics Canada.
- [11] Dagum, E. B. (1988). *The X11ARIMA/88 Seasonal Adjustment Method - Foundations and User's manual*. Time Series Research and analysis Division, Statistics Canada.
- [12] Lothian, J. and Morry, M. (1978). A set of Quality control Statistics for the X-11-ARIMA Seasonal Adjustment Method. Research paper. *Seasonal Adjustment and Time Series Staff*. Statistics Canada.
- [13] Pfeiffermann, P. and Fisher, J. (1981). Festival and Working days Prior Adjustments in Economic Time Series, *Time Series Analysis*. Anderson, O. D. and Perryman, M. R., ed., North-Holland Publishing Company.
- [14] SAS/ETS User's Guide. Version 6. Second Edition.
- [15] Schwartz, G. (1978). Estimating the dimension of a model. *Annals of Statistics*. 6. 461-464.

- [16] Taylor, J. M. (1986). The Retransformed Mean After a Fitted Power Transformation. *Journal of the American Statistical Association.* 81. 114-118.
- [17] William, B., Findley, d., Monsell, B., Otto, M. (1995). The X-12-ARIMA Projection Collaborators at the U. S. Bureau of the Census, 1995. *Annual Research Conference and Seasonal Adjustment Workshop Advance Program.* Bureau of the Census, USA.

[1997년 12월 접수, 1998년 6월 최종수정]

X11ARIMA Procedure

YouSung Park¹⁾ Hyunhee Choi²⁾

ABSTRACT

X11ARIMA is established on the basis of X11 which is one of smoothing approach in time series area and this procedure was introduced by Bureau of Census of United States and developed by Dagum(1975). This procedure had been updated and adjusted by Dagum(1988) with 174 economic index of North America and has been used until nowadays. Recently, X12ARIMA procedure has been studied by William Bell et.al.(1995) and Chen. & Findly(1995) whose approaches adapt adjusting outliers, Trend-change effects, seasonal effect, and Calender effect. However, both of these procedures were implemented for correct adjusting the economic index of North America.

This article starts with providing some appropriate and effective ARIMA model for 102 indexes produced by national statistical office in Korea; which consists of production(21), shipping(27), stock(27), and operating rate index(21). And a reasonable smoothing method will be proposed to reflect the specificity of Korean economy using several moving average model. In addition, Sulnal(lunar happy new year) and Chusuk effects will be extracted from the indexes above and both of effects reflect contribution of lunar calender effect. Finally, we will discuss an alternative way to estimate holiday effect which is similar to X12ARIMA procedure in concept of using both of ARIMA model and Regression model for the best fitness.

1) Department of Statistics, Korea University, Seoul, Korea

2) IBM Korea, Seoul, Korea