

RegARIMA 모형을 이용한 음력 명절효과의 검정에 관한 연구

문 권 순¹⁾

요 약

본 논문은 시계열에 내재된 설·추석 등 음력 명절효과의 존재를 검정하기 위해 RegARIMA 모형의 잔차에 대한 t-검정 통계량을 제시하였으며 Box-plot에 의한 그래픽 진단을 시도하였다. 제시된 t-검정 결과를 X-12-ARIMA의 AICC-사전검정 및 RegARIMA 모형에 의해 추정된 명절효과 회귀계수의 t-값과 비교하였다. 사용된 명절효과 변수는 Bell과 Hillmer(1983)의 명절효과 변수이다.

주요용어 : 명절효과, RegARIMA, Box-Plot, X-12-ARIMA, AICC-통계량

1. 서론

계절조정(Seasonal Adjustment)은 시계열을 추세·순환요인, 계절요인과 불규칙요인으로 분해하는 것으로, 이들 요인을 분해하는 단계와 사전조정단계로 나눌 수 있다. 요인을 분해하는 방법은 이동평균법(Dagum, 1988)과 신호추출법(Burman, 1980) 등이 있으며 이동평균법은 X-11-ARIMA와 X-12-ARIMA, 신호추출법은 TRAMO/SEATS에 적용하고 있다.

사전조정단계는 윤년 및 요일효과, 명절효과 등의 캘린더 효과와 특이치를 계절조정하기 전에 조정해 주는 단계이다. 윤년 및 요일효과, 명절효과 등은 시계열의 변동이 경제적 원인보다 캘린더의 구성에 따라 달리 나타나게 되므로 계절조정 시 감안해 주어야 한다.

계절 및 요일 등의 주기성을 갖는 시계열은 Cleveland와 Devlin(1980)의 스펙트럼 분석에 의해 확인 가능하다. 명절효과는 명절효과 변수가 포함된 RegARIMA 모형의 최우도통계량과 명절효과 변수가 포함되지 않은 RegARIMA 모형의 최우도통계량을 비교함으로써 가능하다(U.S. Census Bureau, 2004과 Caporello, Maravall과 Sanchez, 2003). 그러나 우도함수에 의한 사전검정은 유의성 검정이 아니므로 유의수준을 알 수 없으며, 설과 추석을 구분하여 분석할 수 없는 단점이 있다. 따라서 설과 추석의 명절효과 검정을 위하여 유의성 검정 통계량인 t-통계량에 의하여 우리나라 41종의 산업생산지수 및 도소매판매액지수에 적용하였다.

분석된 41종 시계열중 2종의 시계열은 AICC-사전검정 결과와 같이 명절효과가 없음을 나타냈으나, 7종은 AICC-사전검정과 제안된 t-검정 간 다른 결과를 보였다. 추정된 명절효과의 회귀계수에 대한 t-값과 제안된 t-검정의 결과에서는 4종의 시계열에서 상이한 결과를 얻었다.

2 RegARIMA 모형과 사전조정 요인

X-12-ARIMA 및 TRAMO/SEATS에서 RegARIMA 모형은 윤년효과, 요일효과, 명절효과

1) 통계청 서기관, 대전광역시 서구 둔산동 920 (302-701), ksmoon@nso.go.kr

등의 캘린더 효과, 특이치 등 사전조정 요인을 추정할 뿐 아니라 시계열의 forecasting과 backcasting 등을 위한 모형으로 Bell과 Hillmer(1983)에 의해 제안되었다.

x_t 를 윤년, 요일, 명절 등의 캘린더 효과와 특이치 등을 나타내는 회귀변수라면 원계열 y_t 에 대해서 다음과 같은 회귀모형을 설정할 수 있다.

$$y_t = \sum_{i=1}^r \beta_i x_{it} + z_t \quad (1)$$

이때 회귀잔차 z_t 가 Box와 Jenkins의 승법계절 ARIMA (p d q)(P D Q)s 모형을 따른다고 가정하면, 회귀모형과 승법계절 ARIMA 모형이 결합된 전이함수모형(Transfer function)인 다음의 RegARIMA 모형을 설정할 수 있다.

$$\begin{aligned} \phi(B)\theta(B^s)(1-B)^d(1-B^s)^{D_s} z_t &= \alpha(B)\theta(B^s)a_t \\ \phi(B)\theta(B^s)(1-B)^d(1-B^s)^{D_s} (y_t - \sum_{i=1}^r \beta_i x_{it}) &= \alpha(B)\theta(B^s)a_t \end{aligned} \quad (2)$$

여기서, 잔차 a_t 는 평균 0 분산 σ^2 인 정규분포를 따르는 백색잡음(white noise)과정이다.

윤년·요일효과 등과 함께 캘린더 효과 중 하나인 명절효과는 명절이 있는 월이 연도에 따라 변동하는 이동명절월(moving holiday)에 의해서 발생한다. 서양의 경우는 부활절이 (3월, 4월)에 있으며, 우리나라와 중국 등은 설이 (1월, 2월), 추석은 (9월, 10월)에 있다. 명절효과는 명절이 있는 월과 없는 월의 생산 및 소비 활동 등에 영향을 주어 전월비 뿐만 아니라 전년동월비 증감율에 영향을 주므로 시계열 분석 시 이동명절월의 존재에 유념하여야 한다.

<표 1>은 제조업생산지수(1985년 1월부터 2003년 12월)의 명절이 있는 월(HRS, HRC)과 명절이 없는 월(NHRS, NHRC)의 전월비 및 전년동월비의 평균이다. 여기서 명절효과(NO)는 명절효과가 조정되지 않은 계절조정 결과이며, 명절효과 조정은 Bell과 Hillmer의 명절효과 변수에 의해 사전조정된 계절조정 결과이다. 명절휴일로 전월비 뿐만 아니라 전년동월비에도 명절 월에는 비명절월보다 증가율이 낮음을 알 수 있다.

<표 1> 제조업생산지수의 이동명절월 증감률

	전월비(%)		전년동월비(%)	
	명절효과(NO)	명절효과 조정	원계열	명절효과 조정
HRS(설)	-1.9588	0.2257	6.1371	9.9085
NHRS	4.6547	1.0917	13.7948	9.7878
HRC(추석)	-1.5392	0.4523	7.5592	10.0941
NHRC	3.0642	1.4755	12.1614	9.1812

Bell과 Hillmer(1983)은 명절효과 요인의 추정을 위하여 부활절 일자와 부활절이 영향을 미치는 부활절효과 기간(ω)에 의해 결정되는 (식 3)과 같은 부활절효과 변수 $H(\omega, t)$ 를 (식 2)의 RegARIMA 모형에 회귀시킴으로써 부활절 요인을 추정하였다.

$$\begin{aligned} H(\omega, t) &= \frac{1}{\omega} \sum_{d=1}^{\omega} h(d, t) \\ &= \frac{1}{\omega} [t\text{월이 명절효과기간}(\omega)\text{내에 있는 일수}] \end{aligned} \quad (3)$$

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

3. 명절효과의 검정방법

Liu(1980)은 명절효과 등의 캘린더 효과를 개입변수로 가정한 개입모형(Intervention model)을 설정한 후, ARIMA 모형의 SACF(표본자기상관함수)을 이용하여 캘린더 효과의 유무를 분석하였다. Hillmer, Bell과 Tiao(1983)는 원계열에 캘린더 효과 변수를 회귀시킨 후, 회귀식의 잔차에 대한 SACF와 잔차분산의 크기를 이용하여 캘린더 효과의 유무를 분석하였다. Findley와 Soukup(2000)은 RegARIMA 모형의 AICC-통계량과 Out-of-Sample의 예측오차에 의해 명절효과에 대한 RegARIMA 모형을 선정할 수 있음을 보여주고 있다.

T 를 시계열의 길이, L_T 를 시계열의 우도함수(Likelihood function), n_p 를 추정된 모수의 수라고 한다면, 수정된 AIC-통계량인 AICC-통계량은 다음과 같다.

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

X-12-ARIMA는 명절효과를 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 최소인 RegARIMA 모형을 최적 모형으로 선정하고 있다. 그러나 AICC-통계량에 의한 명절효과 변수의 사전검정은 우리나라와 중국과 같이 2개 이상의 명절 중 어느 한 명절변수가 선호되지 않는 경우, 어느 변수가 선호되지 않았는지 알 수 없으며, AICC-사전검정은 유의성 검정이 아니므로 어느 정도의 유의수준을 갖는지 알 수 없다. 또한 AICC-통계량은 ARIMA 모형의 차분과 특이치 등이 같은 모형에서 비교하여야 하는 등 사용상 제약점이 있다(U.S. Census Bureau, 2004). 따라서 개별 명절효과 변수들에 대한 유의성 검정을 다음과 같이 생각할 수 있다.

명절효과 변수를 포함하지 않은 (식 2)의 RegARIMA 잔차 a_t 는 백색잡음과정으로 평균 0, 분산 σ^2 를 갖는 정규분포이며, a_t 에는 RegARIMA 모형에 의해 제거되지 않은 명절효과 성분들이 남아 있게 된다. 따라서 이동명절월 잔차 (a_t, a_{t+1})를 다음과 같이 명절월의 잔차(a_H)와 명절이 없는 월의 잔차(a_{NH})로 분류한 후, 두 집단에 대한 평균차이($\mu_H = \mu_{NH}$) t-검정을 실시한다.

$$a_H = \{a_t \mid t \in \text{명절이 있는 월}\} \quad a_{NH} = \{a_t \mid t \in \text{명절이 없는 월}\}$$

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

$$t = \frac{\bar{a}_H - \bar{a}_{NH}}{\sqrt{S^2 \left(\frac{1}{n_H} + \frac{1}{n_{NH}} \right)}} \quad (4)$$

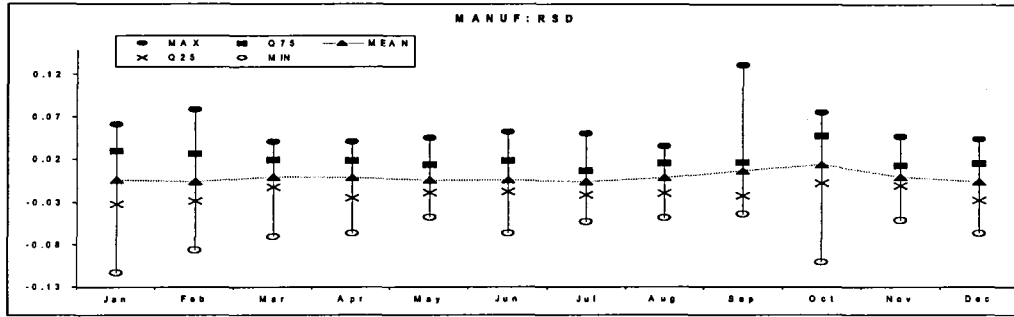
여기서, pooled 분산 $S^2 = \frac{(n_H - 1)S_H^2 + (n_{NH} - 1)S_{NH}^2}{(n_H + n_{NH} - 2)}$ 이다. 이때 두 집단의 분산이 동일하지 않은 경우는 다음의 Satterthwaite의 자유도(df)를 갖는 t-통계량을 이용한다.

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

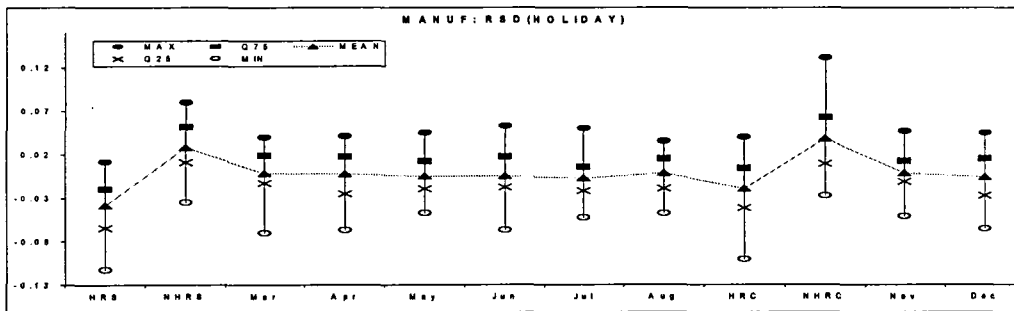
[그림 1]은 특이치와 요일효과 변수를 갖는 제조업생산지수(MANUF)의 월별 RegARIMA 잔

차에 대한 Box-plot이다. [그림 2]는 명절월(HRS, HRC)과 비명절월(NHRS, NHRC)을 분류한 그림으로 설(HRS)과 추석(HRC)의 평균인 “▲”는 다른 월보다 낮은 수준을 나타내고 있으나 비명절월인 NHRS와 NHRC는 다른 월보다 높게 나타나고 있어 명절효과가 있음을 보여주고 있다.

[그림 1] 제조업생산지수의 월별 RegARIMA 잔차



[그림 2] 제조업생산지수의 월별 RegARIMA 잔차: (명절월, 비명절월)



4. 시계열 분석 결과

<표 2>는 제안된 t-검정 결과, Bell과 Hillmer의 명절효과 변수를 포함한 RegARIMA 모형의 AICC-통계량 및 추정된 명절효과 회귀계수의 t-값과 비교한 표이다. AICC-통계량, 제안된 t-검정에 의한 명절효과 변수의 사전검정 결과를 유의수준 5%에서 보면, NCRPNF와 CONORD 등 2종의 시계열은 명절효과가 없는 모형을 선호하는 것으로 나타났으며, 이때 추정된 (설, 추석) 회귀계수의 t-값이 각각 (0.18, 1.23)과 (-0.91, -1.15)로 작게 나타나 추정된 명절효과의 회귀계수도 의미가 없는 것으로 나타났다. BUILD, MVAF와 WRS는 제안된 t-검정과 추정된 회귀계수의 t-값에서 설 효과는 유의적이지 않은 것으로 나타났다. 그러나 UR, MGD, CPI와 RT는 제안된 t-검정과 추정된 회귀계수의 t-값과 다른 결과를 얻었다.

<표 4> 명절효과 변수의 사전 검정

변수명	AICC 사전검정	t-검정		회귀계수 t-값	
		설	추석	설	추석
NCRPNE	without	-0.20	1.38	0.18	1.23
CONORD	without	-1.32	-1.39	-0.91	-1.15
BUILD	with	-0.36	-2.33	-0.35	-3.16
MVAF	with	-0.19	-2.36	-0.52	-2.68
WRS	with	0.76	-3.84	1.56	-3.09
UR	with	-0.82	-0.30	-1.95	-0.97
MGD	with	1.41	-1.21	0.04	-2.71
CPI	with	1.60	0.89	1.60	1.75
RT	with	3.27	0.94	6.56	2.64

주) H: $\beta=0$ 에 대한 5%유의수준의 t-값은 약 1.96, 10% 유의수준은 1.64이다.

참고문헌

- Bell W.R. and S.C. Hillmer(1983), "Modeling Time Series with Calendar Variation", Journal of the American Statistical Association, 1983, Vol 78, 526-534.
- Burman, J.P.(1980), "Seasonal Adjustment by Signal Extraction", Journal of the Royal Statistical Society A, 143, 321-337.
- Cleveland, W.S. and S.J. Devlin(1980), "Calendar Effects in Monthly Time Series: Detection by Spectrum Analysis and Graphical Methods", Journal of the American Statistical Association, 75, 487-496.
- Caporello, G., Maravall A., and F. Sanchez(2003), 『Program TSW revised Reference Manual』, Banco de Espana.
- Dagum, B. E.(1988), 『The X11ARIMA/88 Seasonal Adjustment Method-Foundations and User's Manual-』, Time Series Research and Analysis Division, Statistics Canada.
- Findley, D.F. and R.J. Soukup(2000), "Modeling and Model Selection for Moving Holidays," American Statistical Association Proceeding, October 2000.
- Hillmer S.C., Bell W. R. and G. C. Tiao(1983), "Modeling Considerations in the Seasonal Adjustment of Economic Time Series", 『Applied Time Series Analysis of Economic Data』
- Liu Lon-Mu(1980), "Analysis of Time Series with Calendar Effects", Management Science, Vol. 26, 106-112.
- U.S. Census Bureau(2004), 『X-12-ARIMA/SEATS Reference Manual』, Time Series Staff, Statistical Research Division.