

계절적 시계열 모형화를 위한 VSACF의 확장**

전태준*

Extension of the VSACF for Modelling Seasonal Time Series*

Tae-Joon Jeon*

Abstract

The purpose of this thesis is to develop the new technique for the analysis of seasonal time series by extending the vector sample auto-correlation function(VSACF), which was developed for ARMA modelling procedure.

After the problems of VSACF for modelling seasonal time series are investigated, the adjacent variance is defined and used for decomposing the seasonal factor from the seasonal time series. The seasonal indices are calculated and the VSACF is applied to the transformed series.

The automatic procedure for modelling seasonal time series is suggested and applied to the real data, the international airline passenger travel.

* Department of Industrial Engineering, Chonnam National University

** 이 연구는 1989년도 한국 과학 재단 연구비 지원에 의한 결과임. 과제번호 : KOSEF 893-0913-01-1

1. 서 론

경영 및 공학에서 나타나는 시계열 자료는 대부분 계절적 요소, 추세적 요소 그리고 랜덤 요소로 구성된다. 이러한 세가지 요소를 포괄하는 일반적인 형태는 SARMA(Seasonal Autoregressive Moving Average) 모형[3]이 있으나, 분석 과정에 전문가적인 판단이 요구되기 때문에 실용화 되지 못하고 있다.

계절적 요소를 제외하고 나머지 요소를 고려하는 경우에 ARMA(Autoregressive Moving Average) 모형이 된다. ARMA 모형을 전문적인 지식이 없이 분석할 수 있게 하는 기법으로, AIC(Akaike's Information Criterion)을 이용한 방법[2], DDS(Data Dependent System)[9], Pade Table을 이용한 방법, VSACF(Vector Sample Autocorrelation Function)을 이용한 방법[1] 등이 개발되었다. 이러한 기법들 중에서 본인이 개발한 VSACF 방법이 통계적 특성 및 계산에 있어서 효과적이지만, 계절적 요소가 존재하는 시계열에 적용하기 위해서는 시계열로부터 계절적 요소를 분리하는(Deseasonalize) 과정이 선행되어야 한다.

계절적 요소가 존재하는 시계열의 분석방법으로는 평활 모형을 사용하는 Winter's 방법[12]과 분할기법을 사용하는 Census II[7]를 들 수 있다. 그러나 Winter's 방법은 MSE를 최소화하는 평활상수를 구하는데 시행착오를 거듭해야 하는 문제가 있으며, Census II는 직관에 기초한 방법이기 때문에 이론적인 근거에 의한 통계적 특성을 평가하기 어렵다[7].

최근에는 계절적 시계열을 분석하는 방법으로

ARMA모형에 근거한 분할기법[4, 5], DDS를 이용한 방법[8]등이 제시되고 있으나, 전문적인 지식이 있는 사람이라도 여러번의 시행착오 과정이 요구되어 일반인이 사용하기에는 어려움이 있다.

본 논문은 계절적 요소의 분리 과정을 자동화하여 VSACF 방법을 사용할 수 있게 함으로서, 계절적 시계열을 모형화하는 새로운 방법을 개발함을 목적으로 한다. 이를 위해서 2장에서 VSACF의 성질과 계절적 시계열에 적용상의 문제점을 검토하고, 3장에서 계절적 요소를 분리하는데 사용할 분석 도구와 VSACF를 이용한 계절적 모형화 과정을 제시한 후, 4장에서 실제의 계절적 시계열에 적용한 예를 보았다.

2. VSACF의 성질 및 계절적 시계열 적용상의 문제

2.1 VSACF의 성질

시계열 $\{Z_t\}$ 가 ARMA(p, q) 모형을 따른다고 하면 다음과 같이 표현된다.

$$\phi_p(B)Z_t = \theta_q(B)a_t, \quad t=1, \dots, n \quad (1)$$

여기서 $\phi_p(B) = \sum_{i=1}^p \phi_i B^i$

$$\theta_q(B) = \sum_{i=1}^q \theta_i B^i$$

$$a_t \text{는 } E[a_t]=0, E[a_t^2]=\sigma^2 \text{인 백색 잡음과정}$$

$$\phi_0 = \theta_0 = 0, \phi_p \neq 0, \theta_q \neq 0$$

식 (1)의 ARMA(p, q) 모형은 일반적인 것으로서 두 개의 특성방정식 $\phi(B)=0, \theta(B)=0$ 의 근이 단위원이나 밖에 존재하며 두 특성 방정식의 근들

간에는 공통근이 존재하지 않는다고 가정한다.

Jeon and Park[5]은 크기 h인 VSACF를 다음과 같이 정의하였다.

$$R_h^{(k, j)} = \{r_1^{(k, j)}, r_{j+1}^{(k, j)}, \dots, r_{j+h-1}^{(k, j)}\} \quad (2)$$

여기서 $r_1^{(k, j)} = r_1\{W_{k, t}^{(j)}\}$

$$w_{k,t}^{(j)} = \phi_j^{(j)}(B)Z_t$$

$$\phi_k^{(j)}(B) = \sum_{i=1}^k \phi_{ik}^{(j)}B^i$$

$\phi_{ik}^{(j)}$ 는 추차 회귀 추정치[10]

만약 시계열 $\{Z_t\}$ 가 ARMA(p, q) 모형을 따른다고 하면 VSACF 즉 $R_h^{(k, j)}$ 는 다음과 같은 성질을 만족한다[1].

$$R_h^{(k, j)} = 0 \quad j - q > k - p \geq 0 \quad (3)$$

$$\neq 0 \quad \text{otherwise}$$

(3)의 결과로 부터 $R_h^{(k, j)}$ 가 0 벡터이면서 k+j의 최소가 되는 k와 j를 찾으면 시계열 $\{Z_t\}$ 가 ARMA(k, j-1)모형을 따른다고 할 수 있다. 이러한 원리에 의해서 Jeon[1]은 시계열 모형의 차수 결정(Order determination)과정을 개발함으로써, 시계열 모형화 과정의 자동화를 가능하게 하였다.

VSACF의 요소 $r_1^{(k, j)}$ 는 실제 데이터에서는 0의 값을 가지지 않기 때문에, $\rho_1^{(k, j)} = 0, 1 \geq j$ 라는 가정으로 부터 $r_1^{(k, j)}$ 에 대한 분산값을 이용하여 분석한다.

$$\text{Var}[r_1^{(k, j)}] = \sum_{i=-j+1}^{j-1} \{r_1^{(k, j)}\}^2 / (n-k) \quad i \geq j \quad (4)$$

2.2 VSACF의 계절적 시계열 적용상의 문제점

대부분의 시계열은 ARMA(p, q)모형에서의 차수 p와 q가 3을 넘지 않기 때문에[3], VSACF의 적용

이 효과적이다. 그런, 계절적 시계열에서는 p와 q의 값이 상당히 큰 값을 가지게 된다. 예를 들어 간단한 계절적 시계열 모형인 SARMA(0, 1, 1)*(0, 1, 1)₁₂의 경우 ARMA(13, 13)에 대응되므로 p=13, q=13이 된다.

p와 q의 값이 크게 되면 VSACF 방법의 적용에 있어서 다음과 같은 문제가 발생한다.

첫째, VSACF 분석을 위해서는 최소 p+q+1 수준의 회귀식으로 부터 OLS(Ordinary Least Square) 추정치의 계산이 필요하다.

$$b = (X'X)^{-1}X'Y \quad (5)$$

식 (5)에서 보는 바와 같이 (p+q+1)*(p+q+1) 정방행렬 (X'X)에 대한 역행렬을 구하는 과정에 있어서 많은 계산량과 기억장소가 요구된다.

둘째, VSACF의 요소 $r_1^{(k, j)}$ 의 분산식인 식 (4)에 있어서 분산값이 간격(lag) i의 증가에 따라 커지므로, p와 q의 값이 커지면 분석의 효과를 기대할 수 없다.

셋째, VSACF의 크기 h는 AR과 MA의 차수 p와 q중에서 최대값이 적절하므로[1], p와 q의 값이 커지면 VSACF의 벡터의 수가 많아지기 때문에 계산량이 방대해진다.

위에서 제시한 주요 문제 때문에, VSACF 방법을 계절적 시계열 모형에 적용하기 위해서는 원래의 시계열로부터 계절적요소를 분리하는 과정이 선행되어야 한다.

3. VSACF를 이용한 계절적 시계열 모형화

3.1 계절적 시계열 모형

계절적 요소, 추세적 요소, 랜덤 요소가 공존하는 시계열 $\{Z_t\}$ 에 대한 모형으로 Census II 분할 기법의 승수(Multiplicative)모형을 사용하였다. 즉 $\{Z_t\}$

는 계절적 요소 I_t 와 비계절적 요소 D_t 의 곱으로 표현된다.

$$Z_t = I_t * D_t \tag{6}$$

비계절적 요소 D_t 에는 추세적 요소와 랜덤 요소가 포함되므로, D_t 는 ARMA 모형을 따른다.

$$\phi(B)D_t = \theta(B)a_t \tag{7}$$

따라서 원래의 시계열 Z_t 는 그림 1에서 나타나는 바와 같이 백색 잡음 a_t 로부터 2가지 필터를 거쳐서 생성된다.

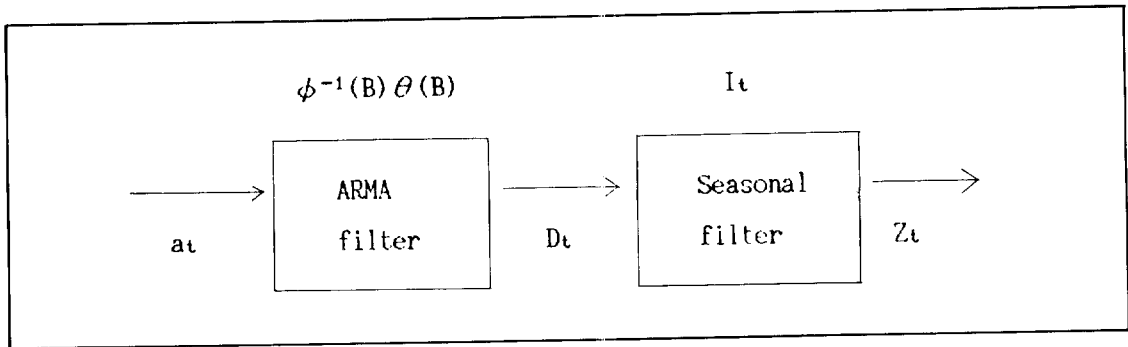


그림 1. 계절적 시계열 모형의 블록 다이어그램

3.2 인접분산

먼저 중심화된 이동평균을 정의한다.

정의 1. 시계열 Z_t 에 대한 k 기간의 중심화된 이동평균(Centralized Moving Average, CMA)을 $M_t(k)$ 로 나타내며 다음과 같이 계산된다.

$$M_t(k) = \sum_{i=t-m}^{t+m} Z_t/k \quad k \text{가 홀수일때} \tag{8}$$

$$= \{Z_{t-m} + 2 \sum_{i=t-m+1}^{t+m-1} (Z_i + Z_{i+m})\} / 2k$$

k 가 짝수일 때

여기서 $m = k/2$ 보다 작은 정수 중 가장 큰 수

$$t = m+1, \dots, n-m$$

만약 시계열 Z_t 가 계절주기(Seasonal Length) S 를 가지고 있다면 S 기간의 CMA, $M_t(S)$ 은 계절적 요소가 분리되어 비계절적 요소로만 구성된다. 그러나 계절주기 S 가 아닌 다른 기간의 CMA, $M_t(k)$ 는 계절적 요소가 분리되지 않기 때문에 계절적 요소가 그대로 존재하게 된다.

어떤 시계열 $\{X_t\}$ 에 계절적 요소가 존재하지 않는냐를 검토하기 위해서 Census II에서는 인접점사(Adjacent Month Test)를 사용하였다. 인접 검사란 식 (9)에 나타낸 바와 같이 어떤 시점의 데이터와 그 시점 전후의 데이터의 평균과의 비율 A_t 를 가지고 계절적 요소의 잔존 여부를 검사하는 것을 말한다.

다.

$$A_t = 2X_t / (X_{t-1} + X_{t+1}) \quad t=2, 3, \dots, n-1 \quad (9)$$

만약 시계열 X_t 에 계절성이 없다면 A_t 는 1에 가까운 값이 되지만, X_t 에 계절성이 존재한다면 A_t 는 1의 값으로부터 편차가 커진다. 이러한 사실에 근거하여 인접 분산을 정의한다.

정의 2. $M_t(k)$ 에 대한 인접 분산(Adjacent Variance)을 $AV(k)$ 로 나타내며, 다음과 같이 계산된다.

$$AV(k) = \frac{\sum_{t=m+1}^{m-m-1} [2M_t(k) / \{M_{t-1}(k) + M_{t+1}(k)\} - 1]^2 / (n-2m-2)} \quad (10)$$

여기서 $M_t(k)$ 와 m 은 식 (8)에서 정의된 것과 같음

인접분산 $AV(k)$ 는 인접점사에서 사용한 A_t 의 개념을 확장한 것으로, A_t 와 A_{t+1} 의 이상치 1값과의 차를 제공한 것에 대한 평균이다. 따라서 $AV(k)$ 는 k 값이 계절주기 S 와 일치될 때 최소값을 가진다. $AV(k)$ 는 A_t 와 마찬가지로 통계적 성질을 규명하기는 어렵고, 직관적인 판단에서 만들어진 것임을 밝힌다.

3.3 계절적 시계열 모형화 과정

제2절에서 나타난 인접분산 $AV(k)$ 의 특성을 이용하여 원래의 시계열 Z_t 의 모형화 과정을 제시하고자 한다.

(1) 여러가지 k 값에 대해서 $M_t(k)$ 와 $AV(k)$ 를 계산한다. k 는 계절주기를 찾기 위한 값이므로 $n/6$ 을 넘지 않는 것이 바람직하다[2].

(2) $AV(k)$ 를 최소화 하는 k 값을 찾는다. 그 때의 k 값을 S 라 하면 S 가 계절주기가 되므로, $M_t(S)$ 는

이 비계절적 요소 D_t 로 구성된다.

(3) 원래의 시계열 Z_t 를 $M_t(S)$ 로 나누어 계절적 요소 I_t 를 계산하고, I_t 를 계절주기 S 별로 분류하여 계절 지표(Seasonal Indices)를 구한다.

(4) 비 계절적 요소 $M_t(s)$ 에 대해서 VSACF 기법을 적용하여 $ARMA(p, q)$ 모형에 해당하는 차수 p 와 q 를 결정한다.

이상의 모형화 과정의 특징은 다음과 같다.

첫째, 계절적 시계열 자료로부터 계절주기 S 를 자동적으로 계산할 수 있다. 그러나, Census II 방법이나, SARMA모형에 있어서는 계절주기 S 를 모형화 과정에서 계산할 수 없으며, 미리 알고 있다고 전제되어 있다. 따라서 본 논문에서 제시한 모형화 과정은 기존 방법에 비해서 절차가 훨씬 간단해진 다.

둘째, 모형화의 전 과정을 자동화 할 수 있다. 계절주기의 선정뿐만 아니라 $ARMA$ 모형의 차수 결정의 중요한 절차가 자동화 되기 때문에, 자료가 입력되면 사람의 판단이나 개입 없이 분석 가능하다.

4. 실제 시계열에의 적용

본 논문에서 제시된 계절적 시계열 모형화 과정을 표 1.에 나타난 1949년부터 1956년 까지 8년 동안의 월별 국제 항공 승객수에 적용하였다.

먼저 중심화된 이동평균 $M_t(k)$, $k=1, 2, \dots, 16$ 까지 구하고 인접분산 $AV(k)$ 를 계산하였다. 계산된 결과는 표 2.에 나타내었다.

표 1. 국제 항공편 승객의 수

년/월	1	2	3	4	5	6	7	8	9	10	11	12
1949	112	118	132	129	121	135	148	148	136	119	104	118
1950	115	126	141	135	125	149	170	170	158	133	114	140
1951	145	150	178	163	172	178	199	199	184	162	146	166
1952	171	180	193	181	183	218	230	242	109	191	172	194
1953	196	196	236	235	229	243	264	272	237	211	180	201
1954	204	188	235	227	234	264	302	293	259	229	203	229
1955	242	233	267	269	270	315	364	347	312	274	237	278
1956	284	277	317	313	318	374	413	405	355	306	271	306

표 2. 인접 분산

k	av(k)*10e6
1	4293.714
2	854.794
3	653.890
4	256.816
5	221.083
6	88.015
7	116.383
8	68.015
9	76.393
10	35.141
11	32.428
12	1.649
13	24.308
14	18.432
15	26.786
16	15.949

표 2.의 인접분산 값이 $k=12$ 에서 최소값이 되므로, 계절주기 $S=12$ 임을 알 수 있다. 따라서 M_k (12)가 비 계절적 요소 D_k 가 되며, 이를 표 3.에 나타내었다.

계절적 요소 I_k 는 표 4.에 나타내었는데, 이는 표 1.의 값을 표 3.의 값으로 나눈 것이다. 표 4.의 맨 끝에 월별 계절 지표를 계산하였다.

표 3.에 나타난 비 계절적 요소에 대하여 VSACF 기법을 적용한 결과를 표 5. 나타내었다. 표 5.에서 지시 기호(Indicator Symbol)[10]는 VSACF의 각 원소의 크기를 기호로 나타낸 것으로, 크기가 2σ 범위 내에 있으면 (.), 범위 밖에 있으면 (-)와 (+)로 나타내었다.

표 3. 비계절적 요소

년/월	1	2	3	4	5	6	7	8	9	10	11	12
1949	0	0	0	0	0	0	127	127	128	129	129	130
1950	131	133	135	136	137	139	141	143	146	148	152	155
1951	157	160	162	164	167	169	171	174	175	177	178	180
1952	183	186	189	191	194	196	198	200	202	206	210	213
1953	216	218	221	223	224	225	225	225	225	225	224	226
1954	228	230	232	234	236	238	240	244	247	250	254	257
1955	262	267	271	275	278	282	286	289	293	297	301	305
1956	310	314	319	322	324	327	0	0	0	0	0	0

표 4. 계절적 요소

년/월	1	2	3	4	5	6	7	8	9	10	11	12
1949	0.00	0.00	0.00	0.00	0.00	0.00	1.17	1.16	1.06	0.93	0.81	0.91
1950	0.88	0.95	1.05	0.99	0.91	1.07	1.21	1.19	1.08	0.90	0.75	0.90
1951	0.92	0.94	1.10	0.99	1.03	1.05	1.16	1.15	1.05	0.92	0.82	0.92
1952	0.93	0.97	1.02	0.95	0.95	1.11	1.16	1.21	1.03	0.93	0.82	0.91
1953	0.91	0.90	1.07	1.05	1.02	1.08	1.17	1.21	1.05	0.94	0.80	0.89
1954	0.89	0.82	1.01	0.97	0.99	1.11	1.26	1.20	1.05	0.92	0.80	0.89
1955	0.92	0.87	0.98	0.98	0.97	1.12	1.27	1.27	1.06	0.92	0.79	0.91
1956	0.92	0.08	0.99	0.97	0.98	1.14	0.00	0.00	0.00	0.00	0.00	0.00
Season Indices	0.91	0.90	1.03	0.99	0.98	1.10	1.20	1.19	1.05	0.93	0.80	0.90

표 5. VSACF 결과

(k, j)	VSACF $r_{\epsilon}^{k, j}$					지시 기호
	j	j+1	j+2	j+3	j+4	
(0,1)	0.95	0.90	0.86	0.83	0.79	++++
(1,1)	-0.26	0.01	-0.09	-0.11	0.03	-. . .
(0,2)	0.90	0.86	0.83	0.79	0.75	++++
(2,1)	0.29	-0.25	-0.20	-0.18	0.09	+-. . .
(1,2)	0.01	-0.09	0.11	0.03	0.16

VSACF의 모형화 과정의 분석 결과는 ARMA (2, 1)모형으로 제시되었으며 이때의 추차회귀 다항식(Iterated Regressive Polynomial)은 다음과 같다.

$$\begin{aligned} \phi_2^{(1)}(B) &= 1 - 1.98B + 0.09B^2 \\ &= (1-B)(1-0.98B) \end{aligned} \tag{11}$$

식 (11)에는 (1-B)가 포함되기 때문에 시계열 Z_t 에 추계적 요소가 존재함을 알 수 있다.

5. 결 론

본 논문에서는 계절적 시계열 자료를 분석하기 위

한 모형, 분석 방법, 모형화 과정을 제시하였다. 분석 방법으로는 인접 분산을 제시하였으며, 이를 이용하여 계절주기를 자료로부터 직접 계산할 수 있으므로 계절적 요인을 분리하는 절차가 간단해진다.

모형화 과정의 특징은 분리된 시계열로부터 VSACF 방법을 적용할 수 있으므로 전 과정을 자동화 할 수 있다는 점이다.

본 논문에서 제시된 모형화 과정을 국제 항공 승객 수에 관한 실제 시계열에 적용하였는데, 앞으로 이론적으로 발생시킨 시계열에 대한 시뮬레이션 연구와 기존 방법과의 비교 연구가 필요하다.

—參考文獻—

- [1] 전태준, 벡터 표본 상관 함수를 이용한 시계열 모형 선정, 한국 과학기술원 박사학위 논문, 1987.
- [2] Akaike, H., "A New look at statistical model identification", IEEE Trans, on Automatic Control, 19, pp716-723, 1974.
- [3] Box, G.E.P. and Jenkins, G.M., Time Series Forecasting and Control Holden Day, New York, 1976.
- [4] Cleveland, W.P. and Dempster, A.P., "Advances in model-based seasonal adjustment", ASA Proceedings in Business and Economics, pp 30-37, 1980.
- [5] Hillmer, S.C. and Tiao, G.C., "An ARIMA-model-based approach to seasonal adjustment", JASA, 77(377), pp63-70, 1982.
- [6] Jeon, T.J., and Park, S.J., "Automatic model indentification using vector sample autocorrelation function", Comm. in Stat., 15(4), pp 1147-1161, 1986.
- [7] Markridakis, S. and Wheelwright, S., Forecasting: Methods and Applications, John Wiley and Sons, 1978.
- [8] Pandit, S.M., "Data dependent system sproach to trend and seasonality", Time Series Analysis: Theory and Practice I, Anderson, O.D.(ed.), pp 515-526, 1982.
- [9] Pandit, S.M. and Wu, S.M., Time Series and Systems Analysis, John Wiley and Sons, 1983.
- [10] Tsay, R.S. and Tiao, G.C., "Consistent estimates of autoregressive parameters and extended sample autocorrelation function for stationary and nonstationary ARMA models", JASA, 79, pp 84-96, 1984.
- [11] Tucker, W.T., "On the Pade table and its relationship to the R and S arrays and ARMA modeling", Comm. in Statistics, All, pp 1335-1379, 1982.
- [12] Winters, P.R., "Forecasting sales by exponentially weighted moving average", Management Science, 6, pp 324-42, 1960.