

시계열 자료의 분할에 관한 사례 연구

문승호*, 이정형**

부산외국어대학교 데이터경영학과*

동아대학교 경영정보학과**

A Study on the Disaggregation Method of Time Series Data

Sungho Moon*, Jeong-Hyeong lee**

Dept. of Data Management, Busan University of Foreign Studies*

Dept. of MIS, Dong-A University**

요 약 마케팅 자료를 입수하는 경우, 시장조사에 시간이 많이 소요되는 등의 이유로, 월간으로 입수하지 못하고, 2개월 간격으로 합산되거나 분기별로 합산된 자료만 입수할 수 있는 경우가 있다. 이러한 자료를 활용하여 월간으로 시장을 평가 혹은 예측하거나 마케팅 전략을 수립하여야 하는 경우, 격월 혹은 분기별로 합산된 자료를 월간 자료로 변환하여야 한다. 본 논문에서는 두 달 간격으로 합산되어 집계되는 자료를 월별 자료로 변환하는 여러 가지 방법을 소개한다. 이런 변환 방법에는 2개월간의 자료를 단순히 2로 나누는 단순평균법, 2개월간의 자료의 증감률을 월별 자료의 증감률에 그대로 적용하여 월별 자료로 변환하는 방법, 전문가의 판단에 따른 가중치를 적용하는 방법, 단순 회귀모형 등의 모형을 정의하고 그 모형에 의해 월별 자료로 분해하는 방법 등이 있다. 본 논문에서는 유럽의 특정 국가의 가전제품 판매 사례를 활용하여, 두 달 간격으로 합산된 시장 자료를 월별 자료로 변환하는 모형을 활용한 방법을 소개하고자 한다. 나아가 이 모형을 활용하여 향후의 자료를 예측하는 방법도 소개한다.

주제어 : 월별 자료, 격월 자료, 분할, 요소분해, 예측

Abstract When we collect marketing data, we can only obtain the bimonthly or quarterly data but the monthly data be available. If we evaluate or predict monthly market condition or establish monthly marketing strategies, we need to disaggregate these bimonthly or quarterly data to the monthly data. In this paper, for bimonthly or quarterly data, we introduce some methods of disaggregation to monthly data. These disaggregation methods include the simple average method, the growth rate method, the weighting method by the judgment of experts, and variable decomposition method using 12 month moving cumulative sum. In this paper, we applied variable decomposition method to disaggregate for bimonthly data of sum of electronics sales in a European country. We, also, introduce how to use this method to predict the future data.

Key Words : monthly data, bimonthly data, disaggregation, variable decomposition, prediction

1. 서론

거시경제 지표와 경영 분야의 시계열 자료들은 조사

범위와 조사시간의 제약 등으로 인해 저해상도 자료를 고해상도 자료로 분할(disaggregation)하거나, 고해상도 자료를 저해상도 자료로 통합(agggregation)하는 경우가

* 이 논문은 2012학년도 부산외국어대학교 학술연구조성비에 의해 연구되었음.

Received 25 March 2013, Revised 25 April 2013

Accepted 20 June 2014

Corresponding Author: Jeong-Hyeong Lee(Dong-A University)

E-mail: jeonglee@dau.ac.kr

© The Society of Digital Policy & Management. All rights reserved. This is an open-access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/3.0>), which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

흔히 발생한다. 특히, 마케팅 분야의 시계열 자료들은 조사범위와 조사시간의 제약, 집계 지연 등으로 인하여 매월 자료를 제공하지 못하고 격월 또는 분기별 자료의 합산치를 제공하는 경우도 흔히 발생한다. 이러한 자료들은 과거 시장의 흐름을 이해하고 평가하기 위한 매우 귀중한 자료일 뿐 아니라, 수집된 자료의 분석과 예측을 통한 미래지향적인 마케팅전략 수립에 활용할 수 있는 많은 정보를 포함하고 있다(Chaiy, 1998). 이와 같은 실정에도 불구하고 많은 경우에 격월 자료의 전략수립에 그대로 활용하거나 단순 평균값으로 적용하고 있다. 그러나 이러한 절차는 월별로 얻어지는 주요한 자료와의 연계분석이 불가능할 뿐 아니라 홀수 월 또는 짝수 월과 같이 관측간격이 상이한 경우 자료 분석은 매우 복잡해지게 된다.

관측간격이 격월인 경우 2개월 간 합산된 자료 또는 주기적 결측치가 존재하는 시계열 자료를 이용한 예측은 향후 월간 예측이 필요한 경우에 능동적으로 대처할 수 없다. 따라서 다양한 예측모형에 의해 예측되는 향후 예측치에 대해 월간으로 계획을 수립할 수 있도록 지원되어야 할 것이다. 이와 같이 분기별 또는 격월로 관측된 자료를 월별 자료로의 변환에 대한 요구가 증대되고 있는 실정이다.

시계열 자료의 분할은 전통적으로 Chow와 Lin(1971)과 Denton(1971)에 의해 제안된 참조 시계열(related time series)의 자기상관함수(autocorrelation function, ACF)의 형태에 의한 방법이 주로 이용되었다. 그러나 이용 가능한 참조 시계열이 존재하지 않을 경우 보간법(interpolation method)을 이용하여 시계열 자료를 분할한다. 따라서 본 논문에서는 2개월 간 합산된 자료들을 월간 자료로 변환하기 위한 마케팅 자료의 순환주기를 고려한 예측모형에 대해 살펴보고, 예측모형을 적용한 사례를 소개하고자 한다.

2. 격월 자료의 분할

격월로 입수된 1년간의 자료는 다음 Table 1과 같이 표현할 수 있다.

〈Table 1〉 Layout of bimonthly data

month	1	2	3	4	5	6	7
observed data		B_1		B_2		B_3	
converted data	y_1	y_2	y_3	y_4	y_5	y_6	y_7
month	8	9	10	11	12	1	
observed data		B_4		B_5		B_6	
converted data	y_8	y_9	y_{10}	y_{11}	y_{12}	y_{13}	

실 입수 자료는 B_1, B_2, \dots, B_6 이고 변환 목적 자료는 y_1, y_2, \dots, y_{13} 이라고 가정하자. 즉, i 번째 격월에 입수된 자료 B_i 는 식 (1)과 같이 쓸 수 있다.

$$B_i = y_{2i} + y_{2i+1} \quad (1)$$

시계열 자료의 분할을 위해 단순회귀모형, 비선형모형, 평활법, 시계열모형, 성장곡선모형 등의 다양한 형태의 모형을 정의하여 보간법에 의해 변환해 주는 방법을 고려해 볼 수 있으나, 모형을 이용한 시계열의 분할은 적용 모형에 따라 상이한 결과를 초래하거나 모형의 설정 단계에서 많은 계산시간과 노력을 필요로 하는 문제점을 가지고 있으므로 본 논문에서는 고려하지 않기로 하였다.

2.1 평균치 적용

〈Table 1〉에서 격월 자료를 단순 산술평균으로 정의하는 방법이다. 즉,

$$y_{2i} = y_{2i+1} = \frac{B_i}{2} \quad (2)$$

로 정의하여 합산된 월에서는 같은 값이 계산되어 입력된다. 이 정의는 계산이 매우 단순하지만 월별 증감이 고려되어 있지 않아 마케팅 자료와 같이 매출 마감시점(반기 말, 연말)의 매출물량이 증가하는 등의 자료 속성을 표현해 주지 못한다.

2.2 증감률 적용

2개월간 합산값들의 증감률을 고려하여 월별 자료로 변환하는 방법이다. 격월 자료의 증감 패턴은 월별 증감 패턴에도 동일하게 작용한다는 가정을 둔다. 즉,

$$\frac{B_{i+1}}{B_i} = \frac{y_{2i+1}}{y_{2i}} \quad (3)$$

로 정의될 수 있다. 여기서 식 (1)과 (3)의 연립방정식을 풀어, 월간 자료 y_{2i} 에 대하여 정리하면 식 (4)와 같이 y_{2i} 와 y_{2i+1} 을 계산할 수 있다.

$$y_{2i+1} = \frac{B_i \times B_{i+1}}{B_i + B_{i+1}} \quad (4)$$

$$y_{2i} = B_i - y_{2i+1}$$

단, <Table 1>에서 최종 격월 자료 B_7 이 존재하지 않으므로 B_6 의 분할값, y_{12} 와 y_{13} 은 식 (4)로부터 계산이 불가능하다. 이때에는 1년간의 입수 자료의 패턴이 동일하다는 가정을 적용한다. 즉, 1년을 주기로 동일한 자료의 패턴이 반복된다고 가정하는 것이다. 따라서 y_{12} 와 y_{13} 은 $B_6/B_1 = y_{13}/y_{12}$ 의 가정을 두어 월간 변환을 정의할 수 있다. 이 방법은 1년(12개월) 주기로 자료의 패턴이 순환한다는 가정이 필요하며, 동시에 격월간의 증감이 월간에도 동일하게 적용된다는 강한 가정이 필요하다. 그러나 1년 주기의 순환이 파악이 되는 경우는 비교적 정확한 결과를 제공한다.

2.3 전문가 판단에 의한 가중치 적용

식 (2)에서 보인 바와 같이 격월 자료를 단순평균에 의하여 분할하는 경우 월간 자료가 동일시되는 단점을 보완하기 위해 월별 가중치를 적용할 수 있다. 이러한 면에서 식 (4)의 증감률에 의한 분할방법은 증감율의 가중치를 적용한 것과 동일하다. 가중치를 결정하기 위한 방법은 크게 두 가지로 나눌 수 있다. 우선, 전략 전문가의 경험을 바탕으로 한 정성적인 가중치 부여방법(expert judgement)과 다음으로 유사 자료에 의한 확률적 모형에 근거한 정량적 방법이 있다(Lee, 1987). 전자의 경우는 전문가의 면밀한 전략분석과 함께 경험을 바탕으로 하기 때문에 정확도는 개선이 될 수 있으나 관련 전문가에 대한 비용이 증가하며 과학적인 근거를 구성하기 어렵다. 후자의 경우는 수학적, 확률적 모형에 근거하기 때문에 전략 변화, 정책변화의 환경요인들에 대한 해석이 불가능하지만 시스템화를 위한 과학적인 배경을 구성할 수 있다.

가중치를 적용한 월별 자료 변환은 식 (1)에 다음과 같이 각각의 월간 자료에 가중치를 적용함으로써 구할 수 있다.

$$B_i = \omega_{2i} y_{2i} + \omega_{2i+1} y_{2i+1} \quad (5)$$

여기서 ω_{2i} 와 ω_{2i+1} 은 경험적, 수학적, 확률적으로 상정된 가중치이다.

이 방법에 대하여 정량적으로 접근하기 위해서는 가중치 부여에 대한 가정을 도입하거나 월간 자료에 대한 과거정보, 혹은 유사한 정보가 있어야 한다. 그러나 대부분의 시장조사 자료에서는 과거정보나 유사정보가 존재하지 않아 가정도입을 통해 가중치를 부여할 수밖에 없다. 이를 위해서는 그 기간 동안의 전략적 방침 등에 따라 임의로 가정을 세워야 한다.

2.4 요소분해

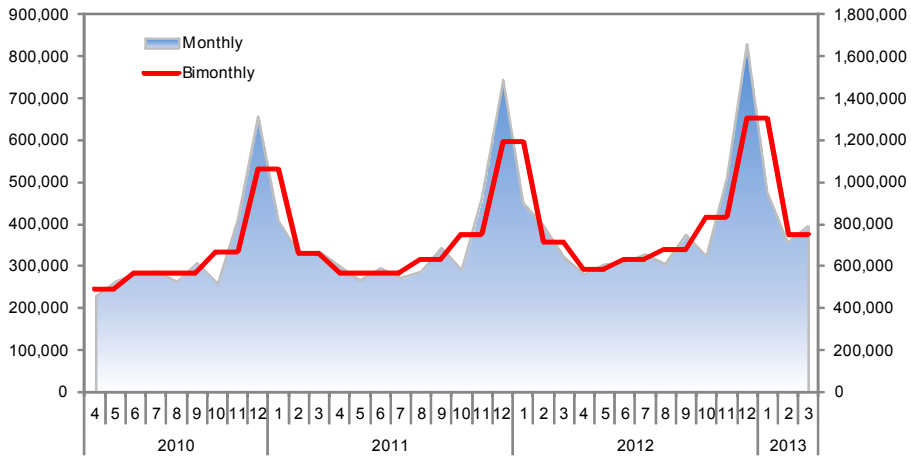
자료가 계절변동과 같이 고정적인 시점에 따라 순환되고 있고 어떤 경향성을 지니는 경우 요소분해법을 적용하여 모형화할 수 있다. t 시점의 매출 자료 y_t 는 식 (6)과 같이 쓸 수 있다(Cho and Sohn, 2011).

$$y_t = f(T_t, C_t, E_t), \quad (6)$$

여기서 T_t 는 t 시점의 추세변동의 구성요소, C_t 는 t 시점의 순환변동의 구성요소, E_t 는 t 시점의 오차 혹은 확률구성요소이다. 순환변동을 고려하기 위해 이동 누적합을 적용한다. 이동 누적합을 계산한 z_t 에 대하여 c 기간순환을 고려하는 경우 식 (7)과 같이 쓸 수 있다.

$$z_t = \sum_{i=0}^{c-1} y_{t-i} + \epsilon_t \quad (7)$$

여기서 c 는 순환주기를 나타내며, 계절변동이 존재하면 $c = 3$ 이 되고, 1년(12개월) 주기로 순환이 이루어진다면 $c = 12$ 가 된다. 따라서 1년간(12개월) 자료의 주기가 순환된다는 가정을 적용하면 t 시점에서 12개월간 이동 누적합 z_t 는 식 (8)과 같이 정의할 수 있다.



[Fig. 1] Bimonthly and monthly disaggregation data

$$z_t = \sum_{i=0}^{11} y_{t-i} + \epsilon_t \quad (8)$$

여기서 이동 누적합 z_t 에 대한 순환주기의 가정이 적합하다면 선형회귀모형으로도 충분히 설명될 수 있다. 즉, $z_t = \beta_0 + \beta_1 t + \epsilon_t$ 로 정의되며 결국, y_t 에 대하여 전개하면 식 (9)와 같다.

$$\begin{aligned} y_t &= z_t - (z_{t-1} - y_{t-12}) \\ &= (\beta_0 + \beta_1 t + \epsilon_t) - (z_{t-1} - y_{t-12}) \end{aligned} \quad (9)$$

3. 사례연구

유럽 특정 A국가에서 가전제품의 매출량에 대해 3년 동안 2개월간 합산치를 격월로 조사한 매출 자료에 대해 적용하고자 한다. 유럽 시장의 가전제품은 연말 특수를 가지는 것으로 알려져 있다. 따라서 월간 자료로 변환하기 위해 단순평균은 부적절하며 분석가의 주관에 개입하지 않도록 증감률에 의한 월변환 방법을 적용하였다. [Fig. 1]의 격월간 가전제품의 매출량에 나타난 바와 같이 1년을 주기로 자료가 순환하고, 시간에 따라 증가하고 있는 선형추세를 보이고 있다.

격월로 관측된 자료의 월별 자료로의 분할에 대한 효율성을 논의하기 위하여 전체 2010년 4월부터 2013년 3

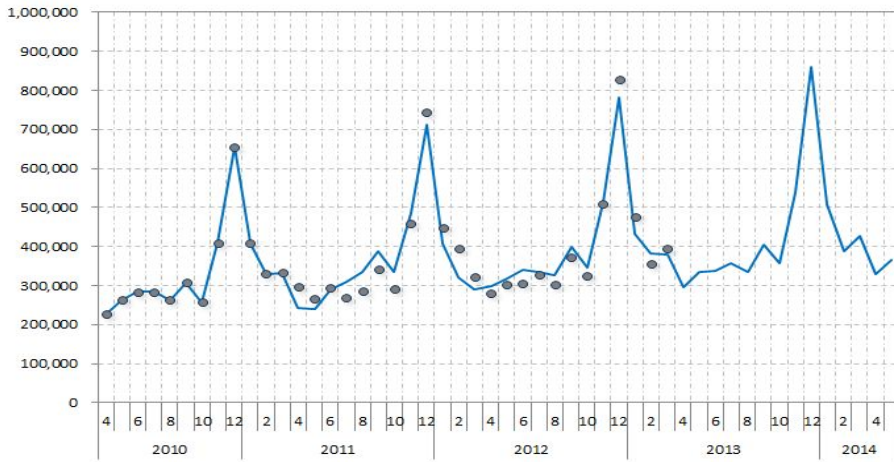
월까지 자료를 훈련자료(training data)와 검증자료(testing data)로 구분하였으며, 2010년 4월부터 2012년 11월까지를 훈련자료로, 2012년 12월부터 2013년 3월까지를 검증자료로 사용하였다.

식 (9)에서의 선형추세를 추정하기 위해 훈련자료에 대해 12개월간 이동누적합을 계산한 다음 이에 대한 z_t 을 예측하였다. 이동누적합 z_t 에 단순선형회귀모형을 적용한 결과 식 (10)과 같이 추정되었다.

$$\hat{z}_t = 3,622,367.9 + 31,714.6 \cdot t \quad (10)$$

적합된 누적합 예측치 \hat{z}_t 를 다시 y_t 에 대하여 전개하여 y_t 를 작수 월로 시작하는 2개월간의 합산치로 복원한 다음 Table 2에 정리하였다. 또한 [Fig. 1]에서 격월로 관측된 자료를 월별 자료로의 분할을 위해 식 (10)에서 추정된 선형추세를 식 (9)에 적용한 결과를 보이고 있다.

Table 2는 2011년 4월부터 2012년 11월까지 10시점(1년 6개월간)을 예측한 결과이다. y_t 에 대한 MAPE(mean absolute percent error)는 t 시점의 APE(absolute percent error) $|\hat{y}_t/y_t - 1| \times 100$ 의 평균으로 계산하였다. 여기서 y_t 는 실측치, \hat{y}_t 는 예측치이다. 2011년 4월부터 2012년 11월까지의 자료를 훈련자료로 예측모델을 수립하고 2012년 12월부터 2013년 3월까지를 검증자료로 타당성을 검증하였다. MAPE는 훈련자료의 경우 6.59%,



[Fig. 2] Final prediction results

검증자료는 6.58%로 우수한 예측성능을 나타냈다.

<Table 2> Prediction results

period	actual	forecast	error rate	
			training	test
Apr-10 - May-10	488704			
Jun-10 - Jul-10	567367			
Aug-10 - Sep-10	569510			
Oct-10 - Nov-10	666582			
Dec-10 - Jan-11	1062867			
Feb-11 - Mar-11	662207			
Apr-11 - May-11	503213	495787	1.48%	
Jun-11 - Jul-11	563902	607078	7.66%	
Aug-11 - Sep-11	628838	725082	15.31%	
Oct-11 - Nov-11	750653	816564	8.78%	
Dec-11 - Jan-12	1190327	1112236	6.56%	
Feb-12 - Mar-12	716873	601502	16.09%	
Apr-12 - May-12	582822	599822	2.92%	
Jun-12 - Jul-12	632012	654703	3.59%	
Aug-12 - Sep-12	676569	699926	3.45%	
Oct-12 - Nov-12	831983	831544	0.05%	
Dec-12 - Jan-13	1304347	1179706		9.56%
Feb-13 - Mar-13	751239	724092		3.61%
Apr-13 - May-13		608963		
Jun-13 - Jul-13		693107		
Aug-13 - Sep-13		737664		
Oct-13 - Nov-13		893078		
Dec-13 - Jan-14		1365442		
Feb-14 - Mar-14		812334		
Apr-14 - May-14		670069		
MAPE			6.59%	6.58%

[Fig. 2]는 2013년 4월부터 2014년 5월까지 예측자료

에 대한 시계열도이다. [Fig. 2]으로부터 기존 자료의 선형추세와 월별 패턴이 제대로 반영되고 있음을 알 수 있다. 시계열 자료의 예측은 다양한 형태로 이루어지고 있으나, 식 (9)에 의한 예측은 매우 직관적이고 계산이 간단하다는 장점을 갖는 것으로 알려져 있다.

5. 결론 및 함의

본 연구에서는 격월로 합산하여 관측된 자료를 월별 자료로 분할하는 방법에 대해 논의하였다. 또한 실례를 통하여 모형적합 과정을 확인하였다. 대부분의 마케팅 자료는 1년을 주기로 반복적으로 순환하는 특징을 가지는 경우가 많으며, 특정 제품의 하위 모델은 계절적인 영향을 받는 경우가 많이 발생한다.

시계열 자료의 분할과 병합은 자료의 특성에 따라 다양한 방법이 존재한다. 본 연구에서는 매우 직관적이고 간단한 계산에 의해 격월로 관측된 시계열 자료를 월별 시계열 자료로 분할하였다. 관측간격이 격월인 경우 2개월 간 합산된 자료 또는 주기적 결측치가 존재하는 시계열 자료를 이용한 예측은 향후 월간 예측이 필요한 경우에 능동적으로 대처할 수 없었으나, 본 연구에서 제시한 12개월 이동 누적합을 이용한 요소분해법은 1년 주기를 갖는 격월 자료와, 분기별 자료의 월별 자료로의 분할에 능동적으로 적용할 수 있다.

ACKNOWLEDGMENTS

This work was supported by the research grant of Busan University of Foreign Studies in 2012.

REFERENCES

- [1] Chaib, S. (1998), *Marketing*, 126-128, Hakhyunsa, (in Korean).
- [2] Cho, S, Sohn, Y. S. (2011), *Time Series Analysis*, 3/e, Yulgok Books, (in Korean).
- [3] Chow, G. C., Lin, A. (1971). Best Linear Unbiased Interpolation, Distribution and Extrapolation of Time Series by Related Series. *Review of Economics and Statistics*, 53, 372-375.
- [4] Denton, F. T. (1971). Adjustment of monthly or quarterly series to annual totals: an approach based on quadratic minimization. *Journal of American Statistical Association*, 66, 99-102.
- [5] Lee, H. (1987), *Introduction to Management*, 394-398, Hyungseol Publishing, (in Korean).
- [6] Lee, J., Cho., S. (2001). Forecast of Foreign Tourist Using Time Series Model, *Journal of the Korean Data Analysis Society*, 3(1), 73-86. (in Korean).

문 승 호(Moon, Sungho)



- 1994년 9월 ~ 현재 : 부산외국어대학교 데이터경영학과 교수
- 관심분야 : 응용통계학
- E-Mail : shmoon@bufs.ac.kr

이 정 형(Lee, Jeong-Hyeong)



- 1998년 3월 ~ 현재 : 동아대학교 경영정보학과 교수
- 관심분야 : 성과평가, 통계학
- E-Mail : jeonglee@dau.ac.kr