경제적과정에 작용하는 계절변동영향의 존재성판단방법

김 종 철

위대한 령도자 **김정일**동지께서는 다음과 같이 교시하시였다.

《전자계산기를 비롯한 현대적기술수단을 받아들여 경영활동의 과학화를 다그치자면 경제조종학과 경제수학적방법이 이룩한 성과를 실정에 맞게 잘 리용하여야 합니다.》

(《**김정일**선집》 증보판 제15권 83페지)

현시기 경제강국건설을 위한 투쟁에서 나서는 중요한 문제의 하나는 경제수학적방법을 적극 받아들여 경영활동에 미치는 요인들의 영향을 세밀하게 타산하고 경제관리를 과학화, 합리화하는것이다.

경제생활에서 계절변동은 극히 보편적인 현상이다. 경제예측에서 계절변동은 대다수가 12개월 혹은 4개 분기를 한개 순환주기로 보며 매해 반복하여 출현하게 된다.

실례로 많은 상품들의 판매량은 기후변동의 영향을 받아서 명확한 계절적인 변동을 나타내고있다.

이러한 계절변동을 포함하고있는 시계렬로부터 경향변동을 분리해낼수 있거나 계절변동의 규칙성을 찾아낼수 있으면 량자를 결합하여 예측을 진행할수 있으며 예측의 정밀도에 대한 요구를 만족시킬수 있다.

계절변동예측방법의 기본원리는 우선 총체적인 시계렬의 경향변동을 묘사할수 있는 방정식을 찾아내고 다음 계절변동이 예측대상에 미치는 영향을 찾아내며 마지막으로 경향변동과 계절변동요소들을 결합시켜 시계렬의 총적인 발전법칙을 묘사한 예측모형을 작성하고 그에 기초하여 예측을 진행하는것이다.

여기에서 우선적으로 해결해야 할 문제는 시계렬에 계절변동요소들이 존재하는가 존재하지 않는가 하는것을 어떻게 판단하겠는가 하는것이다.

시계렬에 계절변동이 존재하는가를 판단하는 방법에는 우선 직관적인 판단방법이 있다.

직관적인 판단방법은 주어진 관측자료의 그라프를 작성하여 그의 변화법칙을 직접적으로 관찰하는 방법이다. 그라프를 그리면 그것이 계절변동의 영향을 받는가 받지 않는가 하는것을 판단할수 있으며 계절의 길이를 확정할수 있다.

이 방법의 우점은 직관적이라는데 있다. 부족점은 판단을 할 때에 주관성을 많이 가지게 되는것이다.

시계렬에 계절변동이 존재하는가를 판단하는 방법에는 또한 자기상관결수에 의한 판단방법이 있다.

시계렬을 $Y_t(t=1, 2, \cdots, n)$로 표시하면 우연변수 Y_t와 Y_{t+k}사이의 상관결수인 k계렬자기상관결수 r_k는 다음과 같이 규정된다.

$$r_k=\frac{\sum_{t=1}^{n-k}(Y_t-\overline{Y}_t)\cdot(Y_{t+k}-\overline{Y}_{t+k})}{\sqrt{\sum_{t=1}^{n-k}(Y_t-\overline{Y}_t)^2\cdot\sum_{t=1}^{n-k}(Y_{t+k}-\overline{Y}_{t+k})^2}} \tag{1}$$

여기서 $\overline{Y}_t=\frac{1}{n-k}\sum_{t=1}^{n-k}Y_t$, $\overline{Y}_{t+k}=\frac{1}{n-k}\sum_{t=1}^{n-k}Y_{t+k}$, n은 자료의 수이다.

r_k 는 변수 Y_t 와 Y_{t+k} 사이의 선형관계의 성질과 세기를 반영해준다.

그러면 자기상관결수를 리용한 계절변동영향의 판단방법에 대하여 보기로 하자.

만일 시계렬에 길이가 L인 계절변동이 존재하면 L계, 2L계 등의 자기상관결수는 정의 값을 취하고 동시에 매우 크다. 그것은 같은 계절의 자료는 동시에 크거나 혹은 작은것으로 하여 이 경우에 정상관관계를 가지며 상관성이 매우 강하기때문이다.

L/2계, L/2+L계 등의 자기상관결수는 일반적으로 부의 값을 취한다. 동시에 절대값은 매우 크다.(여기서 L은 짝수를 가정하며 만일 L이 홀수인 경우에는 (L+1)/2계, (L+1)/2+L계 등으로 고쳐야 한다.)

이러한 특성을 리용하여 시계렬이 계절변동의 영향을 받는가 받지 않는가 하는것을 판단할수 있다.

실례로 어떤 상품판매액자료가 다음과 같이 주어졌을 때 계절변동의 영향이 존재하는가를 판단해보기로 하자.

$$\{y_t\}=\{11, 25, 31, 7, 12, 24, 30, 9, 13, 26, 32, 8, 10, 27, 31, 10\}$$

식(1)을 리용하여 앞의 8계까지의 자기상관결수를 계산한 값은 다음과 같다.

k	1	2	3	4	5	6	7	8
r_k	−0.113	−0.882 3	0.016 8	0.985 4	−0.135 4	−0.887 5	0.027 2	0.991

계산결과를 보면 r_4, r_8 은 정의 값을 취하고 r_2, r_6 은 부의 값을 취하며 동시에 절대값들은 모두 다 크다.

그리고 기타 계의 자기상관결수의 절대값은 모두다 매우 작다. 때문에 이 시계렬에는 계절변동이 존재하며 계절길이는 L=4이다.

경향변동을 가지고있고 또 계절변동의 영향을 받는 시계렬에 대해서는 그의 자기상관결수가 복잡하게 표현되며 두가지 요소가 서로 같이 포함되여있기때문에 때로는 판단이 매우 힘들다.

비교적 쉽게 처리할수 있는 방법은 먼저 주어진 시계렬에서 경향변동을 갈라내고 그 다음에 다시 계절분석을 진행하는것이다.

시계렬에 계절변동이 존재하는가를 판단하는 방법에는 또한 분산분석에 의한 방법이 있다.

주어진 시계렬의 변동은 여러가지 요인의 변동에 의하여 생긴다.

여러가지 요인들중에서도 어떤 몇개의 개별적요인을 생각하여 그 요인들의 변동때문에 생긴 시계렬의 변동이 어느 정도의 확실성을 가지는가 하는 문제를 해결하기 위한 통계적분석방법이 바로 분산분석이다.

분산분석에 의한 계절변동판단방법은 일정한 조건에서 주어진 유의수준에 대하여 이미 알고있는 수 L이 계절길이로 되는가 되지 않는가 하는것을 가려내는 일종의 검정방법이다. 그러므로 이 방법을 사용하기 전에 그라프나 경제적의미에 근거하여 가능한 계절길이 L을 찾아내고 그다음에 이 방법을 리용하여 검정을 진행한다.

분산분석판단방법의 기본원리는 다음과 같다.

우선 주어진 시계렬자료의 경향성을 제거해버려야 한다.

또한 자료를 L개 급으로 갈라놓고(매 급에는 같은 계절의 자료가 포함되여있다고 가정한다.) 매 급 자료의 평균값이 차이가 있는가 없는가를 검정한다. 만일 차이가 있을 때에는 L이 계절길이이라고 판정하며 차이가 없을 때에는 L이 계절길이가 아니라고 판정한다.

그러면 계절길이를 판정하는 구체적인 방법을 보기로 하자.

첫째로, 주어진 자료를 L개 급으로 나누어놓는다. 매 급에는 $n_1, n_2, \cdots, n_L$ 개의 자료가 들어있다. 즉 $n=\sum_{i=1}^{L} n_i$

둘째로, 분산분석의 요구에 따라 총변동 S_T 와 급내변동 S_E, 급간변동 S_A 를 계산한다.

$$S_T=\sum_{j=1}^{L}\sum_{i=1}^{n_j}(y_{ij}-\overline{y})^2=\sum_{j=1}^{L}SS_j-\frac{\left(\sum_{j=1}^{L}S_j\right)^2}{n} \quad (2)$$

$$S_E=\sum_{j=1}^{L}\sum_{i=1}^{n_j}(y_{ij}-\overline{y}_{\bullet j})^2=\sum_{j=1}^{L}SS_j-\sum_{j=1}^{L}\frac{S_j^2}{n_j} \quad (3)$$

$$S_A=S_T-S_E \quad (4)$$

여기에서 $\overline{y}=\frac{1}{n}\sum_{j=1}^{L}\sum_{i=1}^{n_j}y_{ij}$, $\overline{y}_{\bullet j}=\frac{1}{n_j}\sum_{i=1}^{n_j}y_{ij}$, $S_j=\sum_{i=1}^{n_j}y_{ij}$, $SS_j=\sum_{i=1}^{n_j}y_{ij}^2$ 이다.

셋째로, F 검정량을 계산한다.

$$F=\frac{S_A/(L-1)}{S_E/(n-L)}$$

넷째로, 유의수준 α 를 결정한다. F 분포표에서 $F_\alpha(L-1, n-L)$ 를 구한다.

만일 $F>F_\alpha(L-1, n-L)$ 일 때 매 급들의 자료의 평균값은 차이가 존재한다고 즉 계절적영향이 존재한다고 판단하며 L은 계절길이로 된다.

만일 $F\le F_\alpha(L-1, n-L)$ 일 때 매 급의 자료의 평균값에 차이가 없다고 즉 시계렬은 계절영향을 받지 않으며 L은 계절길이가 아니라고 판단한다.

그러면 분산분석방법을 리용한 구체적인 실례를 통하여 시계렬(표 1)에 계절변동이 있으며 L=4가 계절길이로 되는가 안되는가를 판단해보자.

표 1

t	1	2	3	4	5	6
x_t	11	25	31	7	12	24
y_t	0.592 9	1.342 8	1.659 2	0.373 3	0.637 7	1.271 0
t	7	8	9	10	11	12
x_t	30	8	13	26	32	8
y_t	1.583 1	0.420 7	0.681 2	1.357 8	1.665 3	0.414 9

먼저 시계렬의 경향성을 제거한다.

x_t 의 분포도에 근거하여 시계렬 $\{x_t\}$에 선형경향이 있다는것을 알수 있다. 그렇기때문에 먼저 시계렬 $\{x_t\}$내의 경향을 제거하여 표 1의 3행(y_t)자료를 얻었다.

시계렬의 경향을 다음과 같은 방법으로 제거하였다.

최소두제곱법으로 선형경향방정식을 구한다.

$$T_t = 18.485 + 0.066\ 4t$$

t=1, 2, …, 12를 웃식에 대입하여 경향값 T_1, T_2, …, T_{12}를 구한 다음 대응하는 x_t 의 값으로 T_t를 나누어 y_t 값을 구한다.

또한 L=4는 계절길이이라는 가설을 제기하고 검정한다.

시계렬 y_t 를 계절길이의 가설값에 따라 4개 급으로 나눈다. 매개 급에는 3개의 자료가 포함되여있다. 즉 $n_i = 3(i = 1, 2, 3, 4)$

결과는 표 2와 같다.

표 2

급	1	2	3	4	$\sum$
y_{ij}	0.592 9	1.342 8	1.659 2	0.373 3	
	0.637 7	1.271 0	1.583 1	0.420 7	
	0.681 2	1.357 8	1.665 3	0.414 9	
$\sum y_{ij}$	1.911 8	3.971 6	4.907 6	1.208 9	11.999 9
$\sum y_{ij}^2$	1.222 2	5.262 2	8.032 4	0.488 5	15.005 3

총변동, 급내변동, 급간변동을 계산한다.

$$S_T = \sum_{j=1}^{L} SS_j - \frac{\left(\sum_{j=1}^{L} S_j\right)^2}{n} = 15.005\ 3 - \frac{11.999\ 9^2}{12} = 3.005\ 5$$

$$S_E = \sum_{j=1}^{L} SS_j - \sum_{j=1}^{L} \frac{S_j^2}{n_j} = 15.005\ 3 - \frac{1}{3}(1.911\ 8^2 + 3.971\ 6^2 + 4.907\ 6^2 + 1.208\ 9^2) = 0.013\ 8$$

$$S_A = S_T - S_E = 3.005\ 5 - 0.013\ 8 = 2.991\ 7$$

그러므로 F검정량을 계산하면 $F = \frac{S_A/(L-1)}{S_E/(n-L)} = \frac{2.991\ 7/3}{0.013\ 8/8} = \frac{0.997\ 2}{0.001\ 7} = 586.588\ 2$ 이다.

유의수준을 $\alpha = 0.05$ 로 정하고 F분포표에서 $F_{0.05}(3, 8) = 4.07$ 을 찾는다.

F=586.588 2>4.07이기때문에 매 급의 자료의 평균값에는 유의차가 존재한다. 따라서 시계렬은 계절의 영향을 받으며 L=4는 계절의 길이로 된다.