

신제품 수요예측을 위하여 누적자료를 활용한 회귀모형에 관한 연구[†]

박상규¹ · 오정현²

¹중앙대학교 수학과통계학부 · ²중앙대학교 대학원 통계학과

접수 2008년 12월 30일, 수정 2009년 1월 14일, 게재확정 2009년 1월 17일

요약

시계열자료에 계절효과가 존재할 때 성공적인 수요예측을 위해 Winters 방법과 같은 다양한 통계적 방법이 존재지만 신제품과 같이 과거 매출자료가 충분하지 않을 경우 통계적 방법 적용에 한계가 존재한다. 본 연구논문은 신제품과 같이 과거 매출자료가 충분하지 않아 계절효과 등을 추정하기 어려울 때 누적자료를 활용한 통계적 예측방법을 제안한다. 제안된 통계적 방법은 회귀모형이론에 기초하고 있으며 이 방법의 유효성을 최근 화장품 매출자료를 이용하여 검증하였다.

주요용어: 계절효과, 누적자료, 이동평균법, 지수평활법, Winters 방법.

1. 서론

기업은 제품의 수요예측을 통해서 생산계획 및 구매계획을 설정하기 때문에 수요예측의 정확도를 높이기 위해 노력을 한다. 수요예측 방법은 정성적 (qualitative) 수요예측 방법과 정량적 (quantitative) 수요예측 방법으로 구분할 수 있다. 정성적 수요예측 방법은 과거 시장자료가 존재하지 않거나, 존재하더라도 이에 대한 수리적 모형화가 불가능한 상황에서, 일반 소비자의 선호도 혹은 전문가의 지식과 의견을 바탕으로 미래의 수요를 예측하는 방법이다. 정량적 수요예측 방법은 과거시장자료에 대한 통계적 분석을 통하여 미래의 수요패턴을 예측하는 방법으로 시계열 모형, 계량경제 모형 혹은 성장곡선 모형 등에 적합 시켜 그 결과를 이용해서 미래 수요를 예측한다.

신제품 수요예측의 경우는 일반적으로 자료가 충분하지 않기 때문에 다양한 통계적인 방법을 활용하기 보다는 전문가 집단 혹은 특정 고객집단에 대한 설문 조사를 통한 정성적 방법을 많이 활용한다. 하지만 매출 기준으로 어느 정도 시간이 흐르면 정성적인 방법보다 정량적인 방법을 통한 예측 시도가 이루어진다. 이 때 지수평활법 (exponential smoothing) 등을 활용한 직관적인 방법이나 유사제품의 수요패턴을 이용하여 예측력을 보완할 수 있다.

화장품과 같은 제품의 경우는 수명 주기가 2-3년 정도로 매우 짧아 충분한 자료가 축적되지 못하고, 계절성이 존재하기 때문에 지수평활법 등 기존의 통계적 방법을 활용한 수요예측 방법에 한계점이 존재한다. 따라서 신제품의 수요예측에서는 기존의 통계적 방법론이 개선될 필요가 있다고 알려져 있다.

시계열자료를 통계적으로 분석하는 정량적인 수요예측방법은 크게 ARIMA모형을 이용하는 방법과 지수평활법을 이용하는 방법으로 구분할 수 있다. Box-Jenkins의 ARIMA모형은 관측된 시계열자료의

[†] 이 논문은 2007년도 중앙대학교 연구장학기금 지원에 의한 것임.

¹ 교신저자: (156-756) 서울시 동작구 흑석동 중앙대학교 자연과학대학 수학과통계학부, 교수.

E-mail: spark@cau.ac.kr

² (156-756) 서울시 동작구 흑석동 중앙대학교 대학원 통계학과, 석사.

자기 상관성에 대한 분석을 기초로 만들어지는 모형으로 많이 활용되는 계량적인 모형이긴 하지만 높은 정확도를 유지하기 위해서는 충분한 자료가 필요하고, 또 예측성이 높은 만큼 복잡한 이론적 기반을 필요로 한다. 그에 반해 지수평활법은 시계열 자료를 추세, 계절, 순환 및 불규칙 성분으로 구분하는 직관적이고 단순한 통계적 방법으로 최근 수요예측 분야에 가장 많이 활용되고 있다. 지수평활법은 ARIMA모형에 비해 예측 능력이 떨어지지 않으면서 계산량이 적고 수식이 단순하기 때문에 많이 선호되고 있지만 시계열의 내재적 과정에 급격한 수준의 변화 또는 기울기가 발생할 때 그 변화에 신속하지 못하다는 단점을 가지고 있다. 본 연구에서는 계절성이나 추세성이 존재하지만 자료가 충분하지 않을 경우 지수평활법보다 효과적으로 수요를 예측할 수 있는 방법을 제안하고, 실제 화장품 자료를 통해 제안된 통계적 방법의 효율성을 검토하고자 한다.

2. 누적자료를 이용한 회귀모형

2.1. 평활모형 (smoothing model)

시계열자료란 연도별, 분기별, 월별 등 시간의 흐름에 따라 규칙적으로 관측된 자료를 의미하며, 시계열의 변화 형태가 평균과 분산을 중심으로 일정한 변동 폭을 띠면 안정적 시계열이라 하고, 그 반대라면 불안정한 시계열이라 한다. 안정적 시계열에서는 다음 기간의 예측치로 평균값이 유용한 통계량이 되지만, 불안정한 시계열에서는 추세나 계절성분 또는 불규칙한 성분들이 존재하기 때문에 이를 고려할 수 있는 예측 방법이 필요한데, 그 대표적인 방법이 평활법이다.

대표적인 평활법으로 이동평균법 (moving average method)과 지수평활법 (exponential smoothing method)이 있다. 이동평균법은 과거자료에 동일한 가중치를 부여하는 방법이고, 지수평활법은 최근 자료에 더 큰 가중치를 주고 과거로 갈수록 가중치를 지수적으로 줄여나가는 방법이다. 주로 최근 자료가 미래 예측에 영향을 많이 줄 것으로 알려져 있기 때문에 지수평활법은 시계열이 생성되는 구조의 변화가 있는 경우 이 변화에 쉽게 대처할 수 있으며, 그 계산법이 쉽고 많은 자료의 저장이 필요 없다는 장점을 가지고 있다. 그렇기 때문에 통계적 기반 위에 발전된 모형이 아닐지라도 현재까지 실무적 차원에서 가장 폭넓게 사용되고 있다 (De Gooijer와 Hyndman, 2006). 하지만 분석기간 중 추세의 형태가 변하는 경우에는 예측에 어려움이 발생할 수 있다.

시계열이 수평형태를 보일 때 적절한 지수평활법과 달리 시계열에 선형추세 (linear trend)가 존재한다면 이를 고려할 수 있는 Brown (1959)의 이중지수평활모형이나 Holt (1957)의 선형지수평활모형도 많이 활용된다. 하지만 이들 모형들은 계절성 (seasonality)을 고려하지 못하는 단점이 있기 때문에 Winters (1960)는 Holt 모형을 확장하여 계절성을 설명할 수 있도록 모형을 제안하였다. Winters 모형은 평균수준, 추세성분, 계절성분 등에 대한 평활방정식으로 구성되어 있으며, 계절성분을 표현하는 방식에 따라 승법모형 (multiplicative model)과 가법모형 (additive model)으로 구별된다. 승법모형은 평균수준과 추세성분을 더한 결과에 계절성분을 곱하여 이를 예측값으로 한다. 가법모형은 평균수준과 추세성분을 더한 결과에 계절성분을 더하여 이를 예측값으로 한다. 승법모형을 활용한 분석 방법이 일반적이다.

2.2. 누적자료모형 (cumulative data model)

시계열 자료가 충분하다면 시계열의 특성을 나타내는 평균수준, 추세성분, 계절성분 등을 추론하여 정확한 예측을 수행할 수 있다. 하지만 신상품과 같이 수요 자료가 충분하지 않은 경우 다양한 통계적 모형을 적용하기 어렵다. 특히 추세나 계절성분이 존재하는 경우 지수평활법에 의한 예측은 부정확할 것이다.

자료가 가지고 있는 변동성 때문에 시계열 자료로부터 직관적인 의미를 얻어내는 것은 쉽지 않다. 특히 시계열 자료가 충분하지 않을 경우는 더욱 그렇다. 하지만 시간의 흐름에 따라 얻어지는 누적자료는 원 시계열 자료보다 해석하기 쉬운 경우가 많다.

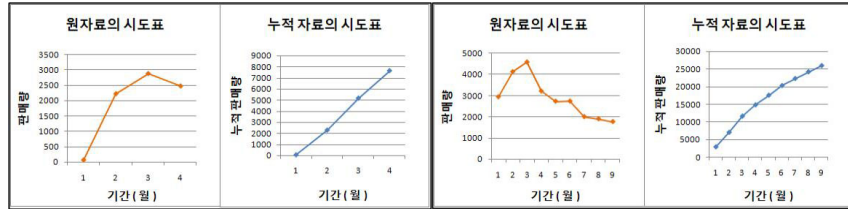


그림 2.1 누적자료를 이용한 회귀모형 적합 결과

일반적으로 누적자료를 설명하기 위하여 이용되는 성장곡선모형 (growth curve model)을 활용할 수 있지만 본 연구에서는 전형적으로 누적자료의 형태가 선형 혹은 포물선 형태를 띤 비선형으로 나타나므로 회귀모형을 통해 예측해보고자 한다. 또한 언제 선형을 사용하고 언제 비선형을 사용하는 기준을 유사제품을 통한 계절성으로 구분하여 예측을 수행하고자 한다.

그림 2.1의 좌측 그림과 같이 누적 자료 형태가 선형인 제품에 대하여 적용되는 모형은 다음과 같다.

$$y_t^* = \beta_0 + \beta_1 t + \epsilon_t, \quad t = 1, \dots, n \quad (2.1)$$

여기서 y_t^* 는 t -번째 시점의 누적 판매량, β_0 와 β_1 은 회귀모수이다. 또한 ϵ_t 는 t -번째 시점에서 발생하는 확률오차를 의미한다. 누적자료를 이용했기 때문에 선형회귀모형 (2.1)의 기울기의 최소제곱추정치가 바로 다음 시점의 예측 값이 된다.

누적자료를 이용한 선형회귀모형 (2.1)은 단기적인 추세를 이해하는데 도움을 주기 때문에 주로 평균에 의존하는 이동평균법이나 지수평활법에 비해 단기간을 예측하는데 효과적일 수 있다. 즉, 매출 자료의 수가 적을 경우 이동평균법이나 지수평활법을 통해 예측을 수행할 때 가장 문제가 되는 점은 추세나 계절효과를 적절하게 반영하지 못한다는 것이다. 반면에 누적자료를 이용한 회귀모형은 최근 추세를 반영하기 쉽고, 계절효과가 나타나는 시점 이후 자료만 사용해서 하거나 혹은 선형이나 비선형모형을 통해 적합한 후 예측치를 제공할 수 있는 장점이 있다.

그림 2.1의 우측 그림과 같이 누적 자료의 시도표가 포물선의 형태를 가질 때, 적용할 수 있는 비선형 모형은 다음과 같다.

$$y_t^{*2} = \beta_0 + \beta_1 t + \epsilon_t. \quad (2.2)$$

누적자료의 형태가 포물선인 비선형회귀모형 (2.2)를 추정하는 방법은 모형 (2.1)을 추정할 때와 마찬가지로 최소제곱추정량을 사용할 수 있고, 다음과 같이 다양한 오차항의 형태에 강건한 Theil-Sen 추정량을 변형하여 사용할 수 있다.

$$S_{nj} = \frac{\hat{y}_n^{*2} - \hat{y}_j^{*2}}{n - j}, \quad T_{nj} = \hat{y}_j^{*2} - j \cdot S_{nj}, \quad j = 1, \dots, n - 1. \quad (2.3)$$

$$\hat{\beta}_1 = \text{median}\{j; S_{nj}\}, \quad \hat{\beta}_0 = \text{median}\{j; T_{nj}\}$$

$S_{ij}\{i > j; i, j = 1, \dots, n\}$ 에 기초한 Theil-Sen 추정량 공식 (2.3)처럼 변형한 이유는 시계열 변동이 큰 상황에서 적은 수의 자료를 활용하여 예측할 때 가장 최근 경향을 많이 반영하려는 의도이다.

누적자료를 이용한 비선형회귀모형 (2.2)는 모형 (2.1)의 선형회귀모형과 달리 다음 시점의 예측이 항상 전 시점의 주문수량보다 감소하는 형태를 띤다. 따라서 주문량이 감소하는 경향을 갖는 제품에 대하여 효과적인 추정을 수행할 수 있다.

누적자료를 활용해서 회귀모형을 통해 예측을 수행하려 할 때 선형회귀모형 (2.1)과 비선형회귀모형 (2.2)를 사용하는 기준은 추세가 상승이냐 감소냐에 따라 선택할 수 있다. 이 판단은 유사제품에서 알려진 계절효과를 참고해서 판단할 수 있다.

3. 사례 : 화장품 신상품 수요 예측

3.1. 자료

본 연구에서는 제품 특성상 수명주기가 짧고 신상품 출시가 많은 화장품을 대상으로 2장에서 논의한 예측 방법론의 성과를 평가해 보고자 한다. 화장품은 1년의 계절성을 보이는 경우가 많아 2년 이상의 자료가 존재할 때는 Winters 방법 등 계절성을 조정한 지수평활법을 이용할 수 있지만 신상품과 같이 매출 자료의 수가 많지 않을 경우는 적절한 수요 예측이 어렵다고 알려져 왔다.

본 연구에서 사용하는 자료는 화장품 전문사인 A의 2006년-2008년도 매출 자료이다. 특히 2008년 새롭게 출시된 신제품 60개의 수요 예측을 시도하고자 한다.

3.2. 자료 분석

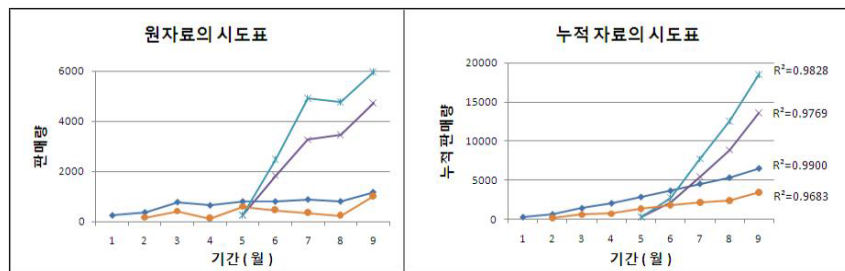


그림 3.1 누적자료를 이용한 선형회귀모형에 적합한 제품들의 시도표

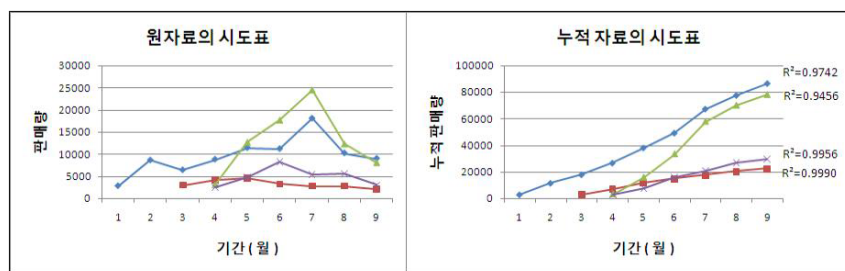


그림 3.2 누적자료를 이용한 비선형회귀모형에 적합한 제품들의 시도표

그림 3.1과 3.2는 최근 출시된 9개월 이하의 매출 자료만 존재할 경우 누적자료를 이용하여 회귀모형 (2.1)과 (2.2)를 적합할 수 있는지를 탐색한 그림이다. 그림 3.1은 누적자료를 이용한 선형회귀모형

(2.1)이 가능한 형태로 볼 수 있고, 그림 3.2는 누적자료가 포물선 형태인 비선형회귀모형 (2.2)가 가능한 경우로 볼 수 있다. 그림 3.1과 그림 3.2를 선형모형과 비선형으로 구분하는 근거는 직전 자료의 매출 증가 여부로 판단할 수 있을 것이다.

제안된 누적자료를 이용하는 방법으로 예측치를 제공하기 위해 모형 (2.1)과 (2.2)를 실제자료에 적합하였다. 모형 (2.1)의 모수를 추정하는데 오차항의 다변량 구조를 설명할 수 있는 일반화 최소제곱추정량 (generalized least square estimator)을 사용할 수도 있지만 최소제곱추정량을 사용할 때와 예측치에 거의 차이가 없음을 확인할 수 있었다. 모형 (2.2)의 경우는 최소제곱추정량이나 일반화 최소제곱추정량으로 예측하였을 경우 예측치가 좋지 않아 조정된 Theil-Sen 추정량을 사용하였다.

표 3.1은 2008년에 출시된 신제품 20개를 대상으로 모형을 적합한 후 결정계수를 계산한 표이다. 모형 적합도를 나타내는 결정계수가 대부분의 제품에서 90%를 넘고 있어 제안된 방법의 타당성과 유용성을 볼 수 있다.

계절성이 존재하면서도 자료의 수가 충분하지 않을 경우 2.2장에서 제안한 누적자료를 이용한 수요 예측 방법은 두 단계로 적용시켜볼 수 있다. 첫 단계는 유형별 혹은 경로별 제품군이나 유사제품군 판매 추이로부터 계절성에 대한 정보를 얻는 단계이고, 두 번째 단계는 계절성 정보를 이용해서 누적자료에 대하여 선형회귀모형이나 비선형회귀모형을 적용하는 단계이다.

표 3.1 2008년 출시된 신제품 20개의 모형 결정계수

번호	결정계수		번호	결정계수	
	모형 (2.1)	모형 (2.2)		모형 (2.1)	모형 (2.2)
1	0.9985	0.9689	11	0.9973	0.9644
2	0.9898	0.9254	12	0.9985	0.9115
3	0.9517	0.9830	13	0.9212	0.9516
4	0.9540	0.9630	14	0.9290	0.8835
5	0.9889	0.9692	15	0.9726	0.9573
6	0.9894	0.8861	16	0.9988	0.9758
7	0.9826	0.9841	17	0.9962	0.9212
8	0.8708	0.8061	18	0.9877	0.9949
9	0.9417	0.9387	19	0.9669	0.9434
10	0.9933	0.9853	20	0.9807	0.9411

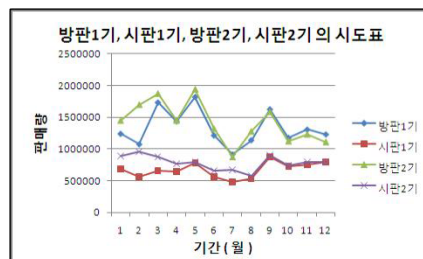


그림 3.3 방판 (시판)1기 : 방판 (시판) 경로의 첫 해 주문 수량, 방판 (시판)2기 : 방판 (시판) 경로의 두 번째 해 주문 수량

계절성에 대한 정보를 얻기 위해 최근 2년간 월별 자료가 존재하는 제품의 경로별 판매 패턴을 보면 다음 그림 3.3과 같이 나타난다. 이는 방판경로³의 제품 76개, 시판 경로의 제품 53개를 대상으로 1년씩

³ A사의 제품 판매 경로로, 방판은 화장품 상담원의 방문 판매나 직영대리점을 통한 판매를 의미하고, 시판은 대형마트나 온라인 (홈쇼핑)을 통한 판매 경로를 말한다.

주문량을 잘라서 본 시도표이다. 상반기에는 약간의 변동성이 존재하지만, 하반기는 전년도와 거의 동일한 패턴으로 판매량의 증감이 나타나고 있음을 알 수 있다.

표 3.2 3개월 예측값에 대한 MAPE (%)

번호	MA	CR	예측력	번호	MA	CR	예측력	번호	MA	CR	예측력
1	73.11	26.51	항상	21	7.25	6.65	항상	41	13.48	13.72	-
2	28.34	19.56	항상	22	19.41	14.90	항상	42	24.06	25.94	-
3	12.51	15.65	-	23	52.48	4.24	항상	43	19.01	19.37	-
4	12.69	8.36	항상	24	210.06	80.55	항상	44	3100.27	975.45	항상
5	27.12	23.13	항상	25	1585.76	859.12	항상	45	80.65	55.91	항상
6	14.32	5.44	항상	26	55.23	51.83	항상	46	133.33	64.24	항상
7	42.14	5.93	항상	27	5.63	5.83	-	47	100.08	61.69	항상
8	8.81	12.35	-	28	78.28	16.08	항상	48	83.76	61.05	항상
9	94.43	32.56	항상	29	23.18	22.99	항상	49	90.48	63.76	항상
10	281.01	120.07	항상	30	11.83	12.78	-	50	72.53	55.96	항상
11	57.65	50.61	항상	31	15.30	10.37	항상	51	66.71	56.03	항상
12	46.43	26.46	항상	32	43.36	8.70	항상	52	61.21	57.59	항상
13	123.43	52.09	항상	33	129.92	71.01	항상	53	73.33	55.20	항상
14	85.74	38.35	항상	34	69.91	4.43	항상	54	93.26	53.74	항상
15	46.19	31.17	항상	35	4.95	7.80	-	55	177.69	81.46	항상
16	27.69	10.28	항상	36	51.76	16.83	항상	56	44.61	42.16	항상
17	408.09	213.98	항상	37	46.77	12.69	항상	57	74.98	56.02	항상
18	30.97	23.06	항상	38	111.69	35.25	항상	58	35.25	43.08	-
19	49.70	4.16	항상	39	17.50	2.71	항상	59	308.99	105.22	항상
20	613.01	392.38	항상	40	21.03	18.03	항상	60	33.36	25.90	항상

상반기에 변동성이 나타난 주원인은 설날로 인한 명절효과가 2006년에는 1월에 2007년에는 2월에 발생한 점, 이것이 신학기 등으로 인해 전통적으로 매출 상승요인이 존재하는 3월에 영향을 주었기 때문인 것으로 분석되고 있다. 일반적으로 추석이 있는 9월의 경우 역시 매출 상승이 일반적인 것으로 알려져 있다. 또한 2년 연속 7월에 매출이 저점을 이루고 있고, 5월과 9월에 매출이 큰 폭으로 증가하는 경향을 보이고 있다. 또한 매출 강세가 있던 그 전달과 그 다음 달에는 감소세, 매출 약세가 있던 그 전달과 그 다음 달에는 증가세가 나타나고 있다. 감소추세가 예상되는 6월, 7월과 명절 효과가 예상되는 달 이후를 예측하는 경우 누적자료가 포물선 형태인 비선형회귀모형 (2.2)를 적용시키고, 그 외의 경우는 누적자료를 이용한 선형회귀모형 (2.1)을 적용시키는 전략을 선택하면 예측 정확도를 높일 수 있다.

표 3.2는 2008년에 출시된 신제품 60개를 대상으로 단순이동평균법 (MA)과 본 연구에서 제안된 통계적 방법 (CR)을 사용하여 최근 3개월을 예측한 후 그 결과를 실제 결과와 비교하여 MAPE⁴를 계산한 표이다. 시계열 자료가 충분하지 않을 때 기존에 사용되고 있는 평활법 중 단순이동평균법을 사용한 이유는 지수평활법과 동시에 고려했지만 단순이동평균법이 MAPE가 더 작아 선택하였다.

표 3.2를 보면 단순이동평균법에 비해 누적자료를 이용해서 회귀모형을 이용하는 방법이 60개 중 9개를 제외하고는 향상된 예측치를 제공하여 주고 있음을 볼 수 있다. 또한 예측치의 정확도 기준으로 활용되는 MAPE가 30% 이하를 유지하는 경우가 52%로 나타나 단순이동평균법을 사용하는 경우의 32%보다 월등한 예측력을 보임으로 누적자료를 이용하는 통계적 방법의 유효성이 매우 높은 것으로 나타나고 있다.

⁴ 평균절대백분율오차 (MAPE: mean absolute percent error)는 수요의 크기에 상대적인 예측오차를 측정하며, 예측성과를 비율의 관점에서 측정한다. $MAPE = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i| / y_i \times 100$ (단위:%)

4. 결론

자료가 많지 않은 상황에서 신상품의 수요예측은 늘 많은 문제점을 발생시킨다. 본 연구에서 제안된 누적자료를 이용한 회귀모형을 통한 수요예측방법은 직관적이면서도 쉽게 적용할 수 있어 실제 현장에서 많은 활용이 기대된다. 화장품 매출 자료를 통해 실증적으로 분석해 본 결과 기존에 많이 활용되고 있는 단순이동평균법이나 지수평활법에 비해 비교적 높은 정확도를 보이고 있음을 발견할 수 있었다. 특히 계절성이 존재하면서도 자료가 충분하지 않을 경우 유사상품의 매출패턴이나 과거자료를 통해 계절성을 예측한 후 선형과 비선형 회귀모형을 이용하여 예측력을 높일 수 있었다.

누적자료를 이용한 회귀모형에 의한 예측방법은 성장곡선모형을 통한 시계열 예측 분야의 한 응용사례로 볼 수 있다. 하지만 성장곡선모형이 주로 S-자형 형태를 보일 때 효과적으로 알려져 있고, 계절성 등을 반영하기 어렵기 때문에 제안된 선형 혹은 비선형회귀모형을 통한 방법은 실무 분야 등에서 많은 응용이 기대된다.

참고문헌

- Brown, R. G. (1959). *Statistical forecasting for inventory control*, McGraw Hill, New York.
- Holt, C. C. (1957). *Forecasting seasonal and trends by exponentially weighted moving averages*, Prentice-Hall, Pittsburgh, Penn.
- De Gooijer, J. G. and Hyndman, R. J. (2006). 25 years of time series forecasting. *International Journal of Forecasting*, **22**, 3, 443-473.
- Sen, P. K. (1968). Estimate of the regression coefficient based on Kendall's tau. *Journal of the American Statistical Association*, **63**, 1379-1389.
- Theil, H. (1950). A rank-invariant method of linear and polynomial regression analysis. *Indagationes Mathematicae*, **12**, 85-91.
- Winters, P. R. (1960). Forecasting sales by exponentially weighted moving averages. *Management Science*, **6**, 324-342.

Regression models based on cumulative data for forecasting of new product[†]

Sang-Gue Park¹ · Junghyun Oh²

¹² Department of Statistics, Chung-Ang University

Received 30 December 2008, revised 14 January 2009, accepted 17 January 2009

Abstract

If time series data with seasonal effect exist, various statistical models like winters for successful forecasts could be used. But if the data are not enough to estimate seasonal effect, not much methods are available. This paper proposes the statistical forecasting method based on cumulative data when the data are not enough to estimate seasonal effect. We apply this method to real cosmetic sales data and show its better performance over moving average method.

Keywords: Cumulative data, exponential smoothing, moving average, seasonal effect, Winters method.

[†] This research was supported by the Chung-Ang University Research Scholarship Grants in 2007.

¹ Corresponding author: Professor, Department of Statistics, Chung-Ang University, Seoul 156-756, Korea. E-mail: spark@cau.ac.kr

² Graduate Student, Department of Statistics, Chung-Ang University, Seoul 156-756, Korea.