

성장곡선 예측 모형의 특성치 보정에 따른 매개변수의 재추정

박주석¹ · 고영현¹ · 전치혁^{1*} · 이재환² · 홍승표² · 문형돈²

¹포항공과대학교 산업공학과 / ²한국전자통신연구원 IT정보센터

Re-estimation of Model Parameters in Growth Curves When Adjusting Market Potential and Time of Maximum Sales

Joo Seok Park¹ · Young-Hyun Ko¹ · Chi Hyuck Jun¹ · Jae Hwan Lee² · Seung Pyo Hong² · Hyung Don Moon²

¹Department of Industrial Engineering, Pohang University of Science and Technology, Pohang, 790-784

²IT Information Center, ETRI, Daejeon, 305-703

Growth curves are widely used in forecasting the market demand. When there are only a few data points available, the estimated model parameters have a low confidence. In this case, if some expert opinions are available, it would be better for predicting future demand to adjust the model parameters using these information. This paper proposes the methodology for re-estimation of model parameters in growth curves when adjusting market potential and/or time of maximum sales. We also provide the detailed procedures for five growth curves including Bass, Logistic, Gompertz, Weibull and Cumulative Lognormal models. Applications to real data are also included.

Keywords: growth curve, forecasting, market potential, nonlinear least squares, expert opinion

1. 서론

성장곡선(growth curve) 모형이라고도 불리는 확산 모형(diffusion model)은 시장예측 방법론 중 정량적 방법의 하나로서 신제품의 도입으로부터 포화 상태까지의 과정을 선형적으로 규명하여 적용한 모형이다. 이는 생물학 분야에서 세포 증식이나 전염병 확산 등에 대한 예측 목적으로 개발되었으며, 1960년대 이후 예측 능력 및 유용성을 인정받으며 특정 제품이나 서비스의 시장성장 예측 분야에 널리 적용되어 왔다. 이 모형에서는 해당 품목에 대한 누적 시장규모를 시간에 따른 S자 형태의 곡선으로 가정한다. 이는 해당 품목의 시장 도입 초기에는 확산 속도가 가속되다가 S 곡선의 변곡점 이후에는 확산 속도가 감소되어 안정화 상태에 도달하리라는 선형적 분석에 근거한다(Meade, 1984). 그리고 확산 모형은 장기적 안정성 측면에서

시계열 모형(time-series model)이나 계량경제 모형(econometric model)에 비해 우수하다고 말할 수 있다(임종인, 오형식, 1992). 또한 확산 모형에서는 함수의 형태 자체는 과거의 경험에 의해 알려져 있다고 가정하므로 추정해야 할 매개변수(parameters)의 개수가 상대적으로 적으며, 각 매개변수는 정성적 방법에 의해 추정할 수도 있으므로 과거 자료에 전적으로 의존할 필요가 없다. 하지만 모형 구축 과정에 주관적 판단과 같은 비정량적 요소가 개입될 여지가 있다는 점은 단점으로 작용할 수도 있다.

이러한 확산 모형을 기반으로 하여 주어진 데이터에 적합한 모형의 매개변수를 추정한 후 예측 결과를 구할 수 있다. 그러나 제시된 예측 결과에 대해 전문가를 비롯한 사용자가 만족하지 못하는 경우에는 사용자의 판단으로 예측 결과를 부분적으로 수정하거나 예측 모형을 수정해야 할 필요가 있다. 특히

*연락처 : 전치혁 교수, 790-784 경북 포항시 남구 효자동 산31 포항공과대학교 산업공학과, Fax : 054-279-2870

e-mail : chjun@postech.ac.kr

2002년 6월 접수, 2회 수정 후 2003년 1월 게재 확정.

데이터 수가 매우 적을 때는 기존 데이터에 기반한 추정치의 신뢰성이 낮을 수 있으므로 전문가의 의견이 적극 반영된 이러한 노력이 절실히 요구된다. 성장곡선의 경우 최대 포화 판매량(market potential)을 나타내는 매개변수가 가장 중요하다고 할 수 있는데, 과거 데이터에 의해 추정된 값을 바탕으로 전문가의 의견을 반영하여 이를 보정할 수 있을 것이다. 한편, 시간별 판매량의 추이를 보면 증가 양상을 보이다가 어느 시점에서 최대를 이루며 그 이후는 감소하는 종 모양 (bell-shaped)을 이루는데, 최대 판매시점에 대한 전문가의 의견이 유용한 경우 이를 활용하여 모형의 매개변수를 재추정할 수 있을 것이다.

따라서 본 논문에서는 성장곡선의 두 가지 특성치인 포화 판매량 및 최대 판매시점에 관련된 다음의 세 가지 경우에 대하여 성장곡선의 매개변수를 재추정하는 문제를 다루고자 한다. 즉, 포화 판매량을 보정하는 경우, 최대 판매시점을 보정하는 경우, 그리고 앞의 두 가지를 모두 보정하는 경우를 다룬다. 관련 이론은 비교적 단순한 편이나 기존 문헌에서 다루어지지 않은 점과, 실제 예측 시스템에 보정 기능을 구현할 때의 활용성 등에 본 논문의 의의를 두고자 한다. 2절에서는 위의 세 가지 각각의 경우에 대한 일반적 방법론을 기술하고, 3절에서는 고려하는 각 성장곡선에 대하여 구체적인 절차를 제시하며, 4절에서는 응용 사례를 예시한다.

2. 특성치 보정과 매개변수 재추정

2.1 대상 모형 및 가정

본 논문에서는 여러 확산 모형 중 가장 널리 사용되는 Bass 모형, Logistic 모형, Gompertz 모형, Weibull 모형, 그리고 Cumulative Lognormal 모형의 다섯 가지 성장곡선 모형을 고려하고자 한다. N_t 를 시점 t 까지의 누적 판매량이라 할 때 성장곡선 모형은 다음 식으로 표현될 수 있다.

$$N_t = F(t, \boldsymbol{y}) + \varepsilon_t \quad (1)$$

위에서 $F(t, \boldsymbol{y})$ 는 매개변수 벡터 \boldsymbol{y} 를 갖는 특정한 성장곡선 식을 나타내며, ε_t 는 오차항을 나타낸다. 누적 판매량 대신 시점 t 에서의 판매량인 S_t 에 대한 데이터가 존재할 때는 당연히 다음과 같이 누적 판매량으로 환산하여 사용할 수 있다.

$$N_t = \sum_{i=1}^t S_i$$

본 연구에서 다룰 다섯 가지의 성장곡선식은 각각 아래와 같다.

· Bass 모형 :
$$F(t; m, a, b) = \frac{m(1 - e^{-(a+b)t})}{1 + \frac{b}{a} e^{-(a+b)t}} \quad (2a)$$

· Logistic 모형 :
$$F(t; m, a, b) = \frac{m}{1 + ae^{-bt}} \quad (2b)$$

· Gompertz 모형 :
$$F(t; m, a, b) = me^{-ae^{-bt}} \quad (2c)$$

· Weibull 모형 :
$$F(t; m, a, b) = m(1 - e^{-at^b}) \quad (2d)$$

· Cumulative Lognormal 모형 :

$$F(t; m, a, b) = m \int_{-\infty}^t \frac{1}{x\sqrt{2\pi b}} e^{-\frac{(\ln x - a)^2}{2b^2}} dx \quad (2e)$$

위의 각 성장곡선 모형은 3개의 양의 매개변수 m, a, b 를 갖고 있다. 여기서 m 은 공히 포화 판매량을 나타낸다. Bass 모형에서 a 는 혁신계수(coefficient of innovation), b 는 모방계수(coefficient of imitation)라 종종 불린다(Bass, 1969).

매개변수의 추정 방법으로는 적절한 함수변환을 동반한 최소자승법, 최우추정법(maximum likelihood estimation) (Schmittlein and Mahajan, 1982; Olson, 1982), 비선형 최소자승법 (nonlinear least squares method; NLS) (Srinivasan and Mason, 1986) 등 여러 가지가 있으나, 본 연구에서는 NLS를 사용한다. 즉, 시점 T 까지의 데이터가 유용할 때 다음 식을 최소로 하는 매개변수 값들을 추정치로 도출한다. NLS에서의 최소화 방법에도 여러 가지가 있으나 Gauss-Newton 방법(Neter *et al.*, 1996)이 널리 사용되며, 본 연구에서도 이 방법을 사용하고 있다.

$$\text{Min } Q = \sum_{i=1}^T [N_i - F(t; m, a, b)]^2 \quad (3)$$

모형별 예측 성능을 비교할 때 또는 한 모형 내에서 서로 다른 매개변수 추정치에 따른 예측 성능을 비교할 때는 통상 오차자승합(sum of squares of errors; SSE)을 활용하는데, 이는 \widehat{N}_t 를 시점 t 까지의 누적 판매량에 대한 추정치라 할 때 다음과 같이 산출된다.

$$\text{SSE} = \sum_{i=1}^T (N_i - \widehat{N}_i)^2 \quad (4)$$

또한 아래와 같은 평균제곱근 오차(root mean squared error; RMS)가 한 시점에서의 평균 오차를 의미하는 것으로 보다 널리 사용된다.

$$\text{RMS} = \sqrt{\text{SSE}/T} \quad (5)$$

데이터 수가 상당수 있는 경우에는 T 개의 데이터를 두 부분으로 나누어 전반부를 모형의 매개변수 추정에 사용하고 후반부를 모형의 타당성 검증에 사용하기도 한다. 이 경우 RMS를 산출할 때 전반부 것을 training RMS, 후반부를 validation RMS라 하기도 한다.

2.2 최대 포화 판매량의 보정

잠재시장 규모라고도 불리는 최대 포화 판매량은 확산 모형

에서 m 으로 표기된 매개변수로서 누적 판매량의 최대치를 의미한다. 주어진 데이터와 NLS 추정법을 통하여 식 (3)을 최소화시키는 매개변수 추정치로 m_1, a_1, b_1 을 얻었다고 하자. 추정 한 후에 사용자가 전문가의 의견 등을 청취한 후 주관적인 판단 하에 m_1 을 m_2 로 수정하고자 한다고 하자. 이 경우, 우선 매개변수 m 을 사용자의 의도에 따라 m_2 로 고정시키고 나머지 매개변수인 a 와 b 를 식 (3)에 기반한 NLS 추정법을 통해 다시 추정함으로써 새로운 매개변수 m_2, a_2, b_2 를 구하게 된다. <그림 1>은 이러한 과정을 나타내고 있다. 새로운 매개변수에 대한 RMS는 증가할 수 있으나, 조정된 포화 판매량이 보다 현실에 가깝다고 판정이 될 때 미래의 예측치는 보다 정확할 것으로 기대된다.

2.3 최대 판매시점의 보정

성장곡선식 $F(t; m, a, b)$ 에 대하여 시점 t 에서의 판매량은 다음과 같이 표현할 수 있다.

$$F'(t; m, a, b) = \frac{dF(t; m, a, b)}{dt} \quad (6)$$

최대 판매시점이란 시간별 제품의 판매량이 최대가 되는 시점을 의미하므로, 다음을 만족하는 시점 t^* 로 산출된다.

$$F''(t; m, a, b) \equiv \frac{d^2F(t; m, a, b)}{dt^2} = 0 \quad (7)$$

데이터 및 NLS에서 추정된 매개변수 (m_1, a_1, b_1)에 대응하는 최대 판매시점 t_1^* 을 사용자 또는 전문가의 견해에 따라 t_2^* 로 조정하고자 한다고 하자. 이 때 최대 판매시점에서의 판매량(최대 판매량)에는 변화가 없다고 하면, 조정 후의 매개변수 (m_2, a_2, b_2)는 다음 두 개의 식을 만족하여야 한다.

$$F''(t_2^*; m_2, a_2, b_2) = 0 \quad (8a)$$

$$F'(t_1^*; m_1, a_1, b_1) = F'(t_2^*; m_2, a_2, b_2) \quad (8b)$$

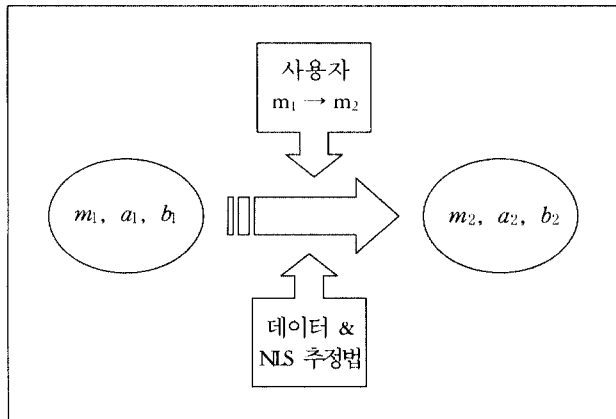


그림 1. 최대 포화 판매량 보정에 따른 매개변수 추정.

여기서 최대 포화 판매량 역시 변화가 없다는 가정 ($m_1 = m_2$)을 추가하면, 식 (8)의 비선형 방정식의 해로서 새로운 매개변수 a_2 와 b_2 를 산출할 수 있다.

<표 1>은 본 논문에서 고려하고자 하는 다섯 가지 성장곡선 모형에 대한 최대 판매시점, 최대 판매량 등을 보여주고 있다. 여기서 판매량은 식 (6)에 의하여 도출된 것이며, 최대 판매시점은 식 (7)의 해로 산출된 것이다.

2.4 최대 포화 판매량 및 최대 판매시점의 동시 보정

본 절에서는 사용자의 판단에 따라 최대 포화 판매량과 최대 판매시점을 동시에 보정하는 경우에 대하여 성장곡선식의 매개변수를 새롭게 추정하는 방안을 다룬다. 동시 보정이라는 표현을 사용하지만, 실제로는 최대 포화 판매량 보정이 이루어진 후에 최대 판매시점 보정을 그 다음에 수행함으로써 진행된다. 즉, 주어진 데이터로부터 NLS를 통하여 매개변수 (m_1, a_1, b_1)을 추정 한 후에 첫 단계로 m_1 을 m_2 로의 조정에 대하여 2.2절에서 기술한 방안 에 따라 새로운 매개변수 (m_2, a_1, b_1)를 얻고, 두 번째 단계로 최대 판매시점 보정에 대하여 2.3절의 방안 에 따라 최종적 매개변수 (m_2, a_2, b_2)를 얻는 것이다. <그림 2>는 최대 포화 판매량 및 최대 판매시점의 동시 보정 절차를 보여준다.

3. 모형별 보정 절차

본 절에서는 앞 절에서 설명한 보정 절차를 모형별로 자세히 설명한다. 단, 최대 포화량 보정의 경우는 2.2절에서 설명한 일반적 방법론이 각 모델에 동일하게 적용되므로 추가적인 설명은 불필요하다. 따라서 본 절에서는 2.3절에 제시된 최대 판매시점 보정 과정이 각 모형에 대해 어떻게 이루어지는지 살펴 보도록 한다.

3.1 Bass 모형

사용자가 최대 포화시점을 t_1^* 에서 t_2^* 로 수정하게 되면 매개변수 a_1, b_1 도 새로운 매개변수 a_2, b_2 로 수정되어야 하며, 이러한 a_2, b_2 는 식 (8)에 의하여 다음 두 식을 만족해야 한다.

$$t_2^* = \frac{1}{a_2 + b_2} \ln \frac{b_2}{a_2} \quad (9a)$$

$$\frac{(a_1 + b_1)^2}{b_1} = \frac{(a_2 + b_2)^2}{b_2} \quad (9b)$$

식 (9a)와 (9b)를 동시에 만족시키는 a_2, b_2 를 추정하기 위해서는 연립 비선형 방정식을 풀어야 한다. 하지만 다음과 같은

표 1. 성장곡선 모형별 특성치

모형	누적 판매량 $F(t; m, a, b)$	판매량 $F'(t; m, a, b)$	최대 판매시점 t^*	최대 판매량 $F'(t^*; m, a, b)$
Bass	$\frac{m(1 - e^{-(a+b)t})}{1 + \frac{b}{a}e^{-(a+b)t}}$	$\frac{ma(a+b)^2 e^{-(a+b)t}}{(a + be^{-(a+b)t})^2}$	$\frac{1}{a+b} \ln \frac{b}{a}$	$\frac{m}{4b} (a+b)^2$
Logistic	$\frac{m}{1 + ae^{-bt}}$	$\frac{mabe^{-bt}}{(1 + ae^{-bt})^2}$	$\frac{\ln a}{b}$	$\frac{mb}{4}$
Gompertz	$me^{-ae^{-bt}}$	$mabe^{-(bt+ae^{-bt})}$	$\frac{\ln a}{b}$	$mabe^{-(1+\ln a)}$
Weibull	$m(1 - e^{-at^b})$	$mabt^{b-1}e^{-at^b}$	$\left(\frac{b-1}{ab}\right)^{\frac{1}{b}}$	$m(ab)^{\frac{1}{b}}(b-1)^{\frac{b-1}{b}}e^{\frac{1-b}{b}}$
Cumulative Lognormal	$m \int_{-\infty}^t \frac{1}{x\sqrt{2\pi b}} e^{-\frac{(\ln x-a)^2}{2b^2}} dx$	$\frac{m}{t\sqrt{2\pi b}} e^{-\frac{(\ln t-a)^2}{2b^2}}$	e^{a-b^2}	$\frac{m}{\sqrt{2\pi b}} e^{\frac{b^2}{2}-a}$

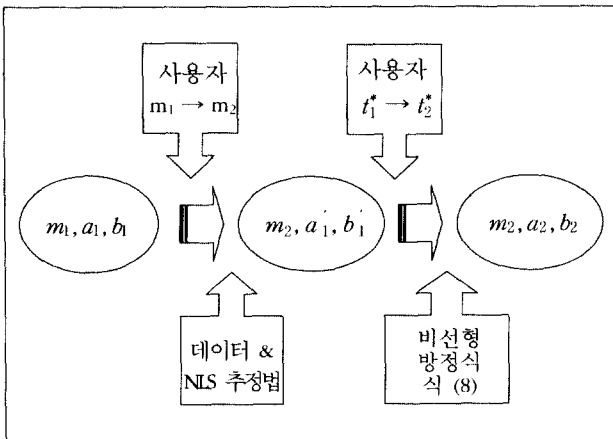


그림 2. 최대 포화 판매량 및 최대 판매시점의 동시 보정 절차.

치환을 통해 하나의 변수를 가지는 비선형 방정식 문제로 변환할 수 있다. 즉, 다음과 같이 새로운 변수들을 정의하면

$$A_1 = \frac{b_1}{a_1}, \quad A_2 = \frac{b_2}{a_2} \tag{10a}$$

$$B_1 = a_1 + b_1, \quad B_2 = a_2 + b_2 \tag{10b}$$

식(10a)와(10b)를 통해 다음과 같이 식(11a)와(11b)가 얻어지고,

$$t_2^* = \frac{1}{B_2} \ln A_2 \tag{11a}$$

$$B_1 \left(1 + \frac{1}{A_1}\right) = B_2 \left(1 + \frac{1}{A_2}\right) \tag{11b}$$

이는 식(11a)를(11b)에 대입함으로써 하나의 식으로 변환된다.

$$A_2^{(1+1/A_2)} = e^{t_2^* B_1 (1+1/A_1)} \tag{12}$$

이렇게 얻어진 식 (12)는 A_2 하나만을 미지수로 가지는 비선형 방정식이 되고, A_1 을 초기치로 사용하는 Newton-Rapson Method (Haberman, 1978) 를 적용하여 쉽게 구할 수 있다. A_2 가 얻어지면 식 (11a)에 의해 B_2 가 얻어지고 식 (10a), (10b)의 역치환을 통해 a_2, b_2 를 쉽게 얻을 수 있다.

3.2 Logistic 모형

본 모형에 대한 새로운 매개변수 a_2, b_2 를 추정하기 위해 필요한 두 가지 식은 식 (8)에 의하여 아래와 같이 도출된다.

$$a_2 = e^{b_2 t_2^*} \tag{13a}$$

$$b_1 = b_2 \tag{13b}$$

따라서 식 (13b)를 (13a)에 대입함으로써 a_2 가 간단히 구해진다.

3.3 Gompertz 모형

새로운 매개변수 a_2, b_2 를 추정하기 위해 필요한 두 가지 식은 식 (14a)와(14b) 이다.

$$b_2 = \frac{\ln a_2}{t_2^*} \tag{14a}$$

$$a_1 b_1 e^{-(1+\ln a_1)} = a_2 b_2 e^{-(1+\ln a_2)} \tag{14b}$$

식 (14a)를 (14b)에 대입하여 정리하면 식 (15)와 같음을 알 수 있다. 매개변수 a_1, b_1, t_2^* 는 주어진 값이므로, 식 (15)의 우변은 상수로 치환된다.

$$a_2 e^{-(1+\ln a_2)} \ln a_2 = a_1 b_1 t_2^* e^{-(1+\ln a_1)} \quad (15)$$

결국 Bass 모형과 마찬가지로 식 (15)는 a_2 만을 미지수로 가지는 비선형 방정식이 되고 a_1 을 초기치로 사용하는 Newton-Rapson Method를 적용하여 쉽게 추정할 수 있다. 이렇게 얻어진 a_2 를 식 (14a)에 적용하면 b_2 가 얻어진다.

3.4 Weibull 모형

새로운 매개변수 a_2 , b_2 를 추정하기 위해 필요한 두 가지 식은 식 (16a)와 (16b)이다.

$$a_2 = \frac{b_2 - 1}{b_2 t_2^{b_2}} \quad (16a)$$

$$(a_1 b_1)^{\frac{1}{b_1}} (b_1 - 1)^{\frac{b_1 - 1}{b_1}} e^{-\frac{1-b_1}{b_1}} = (a_2 b_2)^{\frac{1}{b_2}} (b_2 - 1)^{\frac{b_2 - 1}{b_2}} e^{-\frac{1-b_2}{b_2}} \quad (16b)$$

식 (16a)를 (16b)에 대입하여 정리하면 식 (17)과 같음을 알 수 있다. 매개변수 a_1 , b_1 , t_2^* 는 주어진 값이므로, 식 (17)의 우변은 상수로 치환된다.

$$(b_2 - 1) e^{\frac{1-b_2}{b_2}} = (a_1 b_1)^{\frac{1}{b_1}} (b_1 - 1)^{\frac{b_1 - 1}{b_1}} e^{-\frac{1-b_1}{b_1}} t_2^* \quad (17)$$

결국 식 (17)은 b_2 만을 미지수로 가지는 비선형 방정식이 되고 b_1 을 초기치로 사용하는 Newton-Rapson Method를 적용하여 쉽게 추정할 수 있다. 이렇게 얻어진 b_2 를 식 (16a)에 적용하면 a_2 가 얻어진다.

3.5 Cumulative Lognormal 모형

새로운 매개변수 a_2 , b_2 를 추정하기 위해 필요한 두 가지 식은 식 (18a)와 (18b)이다.

$$a_2 = b_2^2 + \ln t_2^* \quad (18a)$$

$$\frac{1}{b_1} e^{\frac{b_1^2}{2} - a_1} = \frac{1}{b_2} e^{\frac{b_2^2}{2} - a_2} \quad (18b)$$

식 (18a)를 (18b)에 대입하여 정리하면 식 (19)와 같음을 알 수 있다. 매개변수 a_1 , b_1 , t_2^* 는 주어진 값이므로, 식 (19)의 우변은 상수로 치환된다.

$$\frac{1}{b_2} e^{-\frac{b_2^2}{2}} = \frac{t_2^*}{b_1} e^{\frac{b_1^2}{2} - a_1} \quad (19)$$

결국 식 (19)는 b_2 만을 미지수로 가지는 비선형 방정식이 되고 b_1 을 초기치로 사용하는 Newton-Rapson Method를 적용하여 쉽게 추정할 수 있다. 이렇게 얻어진 b_2 를 식 (18a)에 적용하면 a_2 가 얻어진다.

4. 응용 사례

<표 2>는 국내 시스템 소프트웨어의 연도별 매출액을 보이고 있다. 여기서 매출액 데이터는 각 해당년도에서의 매출액이므로 시점 t 에서의 판매량을 S_t 의미한다.

<표 2>의 시스템 소프트웨어 매출액 데이터에 본 논문에서 제안한 시장예측 모형 보정 방법론을 적용해 보고자 한다. 여기서 1996년의 데이터가 다소 추세를 이탈한 값을 지니고 있음을 알 수 있고, 이는 국내 경기 및 외부적 요인에 의한 결과라 할 수 있다. 그러한 데이터는 잘못된 모형을 추정하는 결과를 낳게 하고, 따라서 본 연구에서 제시하는 방법과 같은 전문가의 의견을 반영한 예측 모형의 보정이 필요할 것이다. 여기에서는 1987년부터 1996년까지의 데이터를 매개변수 추정에 사용하며, 1997년부터 1999년까지의 데이터는 모형 검증에 사용한다.

우선 다섯 가지 성장곡선 모형별로 위의 데이터와 NLS에 의한 매개변수를 추정한 결과는 <표 3>과 같다. NLS는 그 특성상 매개변수 추정시 초기값에 따라 결과가 다를 수 있다. 본 연구에서는 염세경 외 (2001)에서 제시한 방법에 의하여 초기값을 설정하였다. <표 3>의 추정 결과를 보면 모형에 따라 최대 포화 판매량이 매우 상이하게 추정됨을 알 수 있다. <표 3>에서 RMS는 전반부 데이터(1987~1996)에 대한 모형 추정관련 RMS를, vRMS는 후반부 데이터(1997~1999)에 대한 모형 검증에서의 RMS를 의미한다. 이상적으로는 이 두 가지의 RMS가 작을수록 더 좋은 예측 모형이라고 할 수 있을 것이다.

4.1 최대 포화 판매량 (m) 보정 적용 결과

<표 4>는 시스템 소프트웨어의 최대 포화 판매량(누적 매출액)이 50,000 (5조원)일 것이라는 전문가의 의견을 각 모델에 반영해 새로운 매개변수들을 추정한 결과를 보여준다. 이러한

표 2. 시스템 소프트웨어 매출액

(단위 : 억원)

연도	'87	'88	'89	'90	'91	'92	'93	'94	'95	'96	'97	'98	'99
매출액	35.8	97.6	145.3	251.6	353.3	535.5	1076.0	1204.6	2726.9	2369.7	2969.9	4046.2	4234.5

자료출처 : 통계청

보정 후 모형의 예측 능력인 $vRMS$ 의 값을 <표 3>의 당초 결과와 비교할 때, Lognormal 모형을 제외한 모든 모형에 있어서 상당 부분 줄어들어 향상되고 있음을 알 수 있다. 한편, 보정 전에는 최대 판매시점 등이 모형별로 상당한 차이를 보이고 있는 것에 반해, 보정 후에는 모형별 차이가 크게 축소되었음을 볼 수 있다. 특히 Gompertz, Weibull 모형의 경우 보정 전에 추정된 최대 포화 판매량은 전문가의 의견인 50,000에 비해 상당히 큰 것을 알 수 있고, 이는 예측에 있어서도 잘못된 결과를 주고 있음을 시사해 주는 결과라 하겠다. 또한 이 결과는 데이터에 의한 모형의 절대적 신뢰는 불확실한 예측 결과를 줄 수 있고, 전문가의 의견이 어느 정도 필요함을 확인시켜주는 결과라 할 수 있다.

4.2 최대 판매시점(t_1^*) 보정 적용 결과

<표 5>는 시스템 소프트웨어의 최대 포화 판매량은 변화가 없다는 가정하에 최대 판매(매출) 시점이 2000년 일 것이라는 전문가의 의견을 각 모델에 반영해 새로운 매개변수들을 추정

한 결과를 보여준다. 이 경우 모형별로 RMS가 크게 증가함을 볼 수 있으며, 수정 정도가 적은 Logistic과 Bass 모형의 경우에는 $vRMS$ 의 측면에서 소폭 향상이 있음을 알 수 있지만 수정 정도가 비교적 큰 Gompertz, Weibull, Lognormal의 경우에는 오히려 예측 능력이 상당히 안 좋아지는 결과를 보여 주고 있다. 이는 데이터에 의해 예측된 포화 시점을 인위적으로 보정한 결과로서, 이 방법의 경우 최대 포화 판매량이 수정 전 값과 동일할 것이라는 가정에 문제가 있음을 알 수 있다. 따라서 최대 판매시점의 보정시에는 최대 포화 판매량을 동시에 수정함이 필요하다 하겠다.

4.3 최대 포화 판매량(m) 및 최대 판매시점(t_1^*) 동시 보정 적용 결과

<표 6>은 최대 포화 판매량이 50,000이며 최대 판매시점이 2000년 일 것이라는 전문가의 의견이 있다고 할 때 이 두 특성치를 동시에 보정한 결과를 보여 준다. 이 결과는 RMS 및 $vRMS$ 로 판단할 때 최대 포화 판매량만을 보정한 결과 보다는 다소

표 3. 성장곡선 모형별 매개변수 추정 결과

	Logistic	Bass	Gompertz	Weibull	Lognormal
m_1	19010.8	20154.5	308777.2	222340.4	1930985.9
a_1	436.73311	0.00145	9.85794	0.00001	4.62239
b_1	0.59435	0.57278	0.10225	3.57113	0.88751
t_1^*	1997년 2.7월	1997년 4.9월	2009년 4.6월	2009년 11.0월	2033년 3.4월
RMS	129.249	132.804	152.278	176.279	210.530
$vRMS$	1662.509	1500.584	921.023	641.865	223.476

표 4. 성장곡선 모형별 최대 포화 판매량 보정 결과

	Logistic	Bass	Gompertz	Weibull	Lognormal
m_2	50000	50000	50000	50000	50000
a_2	582.24027	0.00095	10.52732	0.00004	2.77890
b_2	0.48536	0.46890	0.17990	3.72630	0.50715
t_2^*	2000년 1.4월	2000년 2.4월	2000년 1.0월	2000년 11.1월	1999년 5.4월
RMS	183.168	167.375	188.661	177.802	254.763
$vRMS$	1357.953	1184.062	751.038	220.375	1148.156

표 5. 성장곡선 모형별 최대 판매시점 보정 결과

	Logistic	Bass	Gompertz	Weibull	Lognormal
m_2	19010.8	20154.5	308777.2	328118.9	2780266.9
a_2	2267.90457	0.00032	3.77802	0.00020	8.93669
b_2	0.59435	0.57504	0.10225	3.16932	2.52423
t_2^*	2000년	2000년	2000년	2000년	2000년
RMS	2711.394	2584.087	42024.702	34389.507	3840.931
$vRMS$	1505.378	1417.546	7680.571	22286.73	2418.782

표 6. 모형별 최대 포화 판매량 및 최대 판매시점의 동시 보정 결과

	Logistic	Bass	Gompertz	Weibull	Lognormal
m_2	50000	50000	50000	50000	50000
a_2	549.85596	0.00105	10.36750	0.00005	4.25500
b_2	0.48536	0.46853	0.17990	3.72630	1.30002
t_2^*	2000년	2000년	2000년	2000년	2000년
RMS	262.640	368.630	223.308	1411.035	2121.772
vRMS	1446.146	1330.752	742.186	921.743	3293.398

안중지만 최대 판매시점만을 보정한 결과 보다는 다소 좋은 결과를 보여 주고 있음을 알 수 있고, 보정을 하지 않은 결과보다도 다소 좋은 결과를 보여준다. 보정을 하지 않은 경우에 있어서, Gompertz, Weibull, Lognormal의 경우 추정된 모형 특성치들이 전문가의 의견과는 상당히 괴리가 있는 것으로 보아 보다 먼 미래에 대한 예측에 있어서는 보정 결과가 더 적합할 것이라는 것을 추측할 수 있다.

5. 결론

본 논문에서 제시하고 있는 보정 방법은 특히 시계열 데이터 수가 적은 상황에서 성장곡선을 예측 모형으로 사용하고자 할 때 유용할 것으로 사료된다. 이는 적은 데이터에만 의존하여 추정된 매개변수 값들은 그 신뢰도가 매우 낮을 수 있으며, 최대 포화 판매량 또는 최대 판매시점에 대하여서는 어느 정도 전문가로부터 정보를 얻을 수 있기 때문이다. 특히 데이터가 적은 상황에서는 그 예측 결과가 사용하는 모형에 따라 크게 차이가 날 수 있는데, 이러한 전문가 의견을 반영하여 보정할 때에는 모형간 차이가 상당히 줄어들 수 있을 것이다.

한 가지 유의할 점은 본 보정 방안이 자칫하면 인위적으로 될 소지가 있다는 것이다. 즉, 유용한 과거 데이터에만 과도하게 적합(overfitting)시키는 특성치의 보정이 이루어지지 않도록

유의하여야 할 것이다. 또한 보정이 이루어진 경우 예측 결과와 실적치를 계속적으로 비교 검토하여 그 타당성을 관찰함이 필요할 것이며, 상당량의 시계열 데이터가 축적될 때까지 이러한 보정 과정이 수반되어야 할 것이다.

참고문헌

엄세경, 홍승표, 강희일, 김지수, 전치혁(2001), 비선형 최소자승법을 이용한 성장곡선 모형의 매개변수 추정시 초기값 설정방법에 관한 연구, *산업공학*, 14(2), 190-197.
 임종인, 오형식(1992), 신제품 수요예측 방법론 연구, *대한산업공학회지*, 18(2), 51-63.
 Bass, F.M. (1969), A new product growth model for consumer durables, *Management Science*, 15, 215-227.
 Haberman, S.J. (1978), *Analysis of Qualitative Data*, Academic Press.
 Meade, N. (1984), The use of growth curves in forecasting market development - a review and appraisal, *Journal of Forecasting*, 3, 429-451.
 Neter, J. et al. (1996), *Applied Linear Statistical Models*, 4th edition, Irwin Press.
 Olson, J.A. (1982), Generalized least squares and maximum likelihood estimation of the logistic function for technology diffusion, *Technical Forecasting and Social Change*, 21, 241-249.
 Schmittlein, D.C. and V. Mahajan (1982), Maximum likelihood estimation for an innovation diffusion model of new product acceptance, *Marketing Science*, 1, 57-78.
 Srinivasan, V. and C.H. Mason (1986), Nonlinear least squares estimation of new product diffusion models, *Marketing Science*, 5, 169-178.



박주석
 한국과학기술원 물리학 학사
 현재: 포항공과대학교 산업공학 석사과정 재학중
 관심분야: 투자분석, 응용통계



고영현
 포항공과대학교 산업공학 학사
 포항공과대학교 산업공학 석사
 현재: 포항공과대학교 산업공학 박사과정 재학중
 관심분야: 응용통계 및 Data mining



전치혁

서울대학교 자원공학 학사
KAIST 산업공학 석사
미국 캘리포니아 대학(버클리) 산업공학 박사
현재: 포항공과대학교 산업공학과 교수
관심분야: 확률/통계모형, 데이터마이닝



홍승표

동국대학교 산업공학과 학사
포항공과대학교 산업공학 석사
현재: 한국전자통신연구원 정보체계연구팀
연구원
관심분야: 정보통신 산업분석, 무선인터넷 서비스 산업



이재환

경희대학교 전자공학과 학사
경희대학교 전자공학과 석사
현재: 한국전자통신연구원 정보조사분석팀
팀장
관심분야: 정보통신 기술/시장/산업동향, IT
신분야 기술/산업 전망



문형돈

아주대학교 산업공학과 학사
아주대학교 산업공학 석사
현재: 한국전자통신연구원 정보조사분석팀
연구원
관심분야: 정보통신 산업분석, 차세대 이동통신 시스템