

# 비선형 최소자승법을 이용한 성장곡선 모형의 매개변수 추정시 초기값 설정 방법에 관한 연구

염세경<sup>1</sup> · 홍승표<sup>2</sup> · 강희일<sup>2</sup> · 김지수<sup>1</sup> · 전치혁<sup>1</sup>

<sup>1</sup>포항공과대학교 산업공학과 / <sup>2</sup>한국전자통신연구원 지식정보센터 정보체계연구팀

## Determination of starting values in estimating growth curves by using non-linear least squares

Se-Kyoung Youm<sup>1</sup> · Seung-Pyo Hong<sup>2</sup> · Hoe-Il Kang<sup>2</sup> · Ji-Soo Kim<sup>1</sup> · Chi-Hyuck Jun<sup>1</sup>

Growth curves including Logistic and Gompertz functions are widely used in forecasting the market demand. To estimate the parameters of those functions, we use the non-linear least square method. However, it is difficult to set up the starting points for each parameter. If a wrong starting point is selected, the result reveals the local optimum or does not converge to a certain value. The purpose of this paper is to resolve the problem of selecting a starting point. Especially, rescaling the market data using the national economic index make it possible to figure out the range of parameters and to utilize the grid search method. Applications to some real data are also included.

### 1. 서론

시장예측 방법은 크게 정성적 예측법과 정량적 예측법의 두 가지로 나눌 수 있다. 설문조사법이나 델파이법 그리고 AHP (analytic hierarchy process) 등으로 대표되는 정성적 예측법은 일반 소비자의 선호도 혹은 전문가의 지식과 의견을 바탕으로 미래의 시장수요를 예측하는 방법이다. 이 방법의 경우 과거 시장자료가 존재하지 않거나, 존재한다 하더라도 이에 대한 수리적 모형화가 불가능한 상황에서 활용 가능하다는 장점을 가지지만, 주관적 판단과 같은 비정량적 요소에 전적으로 의존해야 한다는 번이 단점으로 지적되고 있다. 한편 정량적 예측법은 과거 시장자료에 대한 통계적 분석을 통하여 파악된 수요 경향을 수리적 모형에 반영함으로써 미래의 시장수요를 예측하는 방법으로, 주로 사용되는 모형으로는 시계열(time series) 모형, 계량경제(econometric) 모형, 확산(diffusion) 모형 등이 있다. 시계열 모형이나 계량경제 모형의 경우 안정적 분석 및 예측을 위하여 상당량의 과거 시장자료를 확보해야 하는데 반하여, 확산 모형의 경우에는 비교적 적은 자료만으로도 장기적 안정성을 갖춘 시장예측이 가능하다는 장점을 가진다(임

종인, 오형식, 1992).

확산 모형은 생물학 분야에서 세포 증식이나 전염병 확산 등에 대한 예측 목적으로 개발되었으며, 1960년대 이후 모형의 예측 능력 및 유용성을 인정받으며 다양한 종류의 제품이나 서비스에 대한 시장예측 분야에 널리 적용되어 왔다. 이 모형에서는 해당 상품의 시장 도입 초기에는 확산 속도가 가속되다가 특정 시점 이후에는 확산 속도가 감소되어 안정화 상태에 도달하리라는 선형적 분석에 근거하여 해당 상품에 대한 시장규모를 시간에 따른 S자 형태의 곡선으로 가정한다. 이러한 이유로 확산 모형은 성장곡선(growth curve) 모형이라 불리우기도 한다(Meadc, 1984). 이제까지 많은 연구를 통하여 다양한 형태의 성장곡선 모형이 제안되어 왔는데, 이들은 크게 선형 모형과 비선형 모형으로 구분될 수 있다(Meadc and Islam, 1998; Young, 1993). 선형 성장곡선 모형을 이용한 대부분의 분석에서는 잠재시장규모를 상정하고 분석하는데, 이 경우 장기 예측 시점에서의 잔차가 크게 발생하는 것으로 알려져 있다. 이에 비하여 비선형 성장곡선 모형에서는 잠재시장규모를 하나의 매개변수로 간주하여 추정할 수 있다는 장점을 가진다(황정연, 강병용, 전경표, 1999). 본 연구에서 다루고 있는 성장곡선 모형은 비선형 성장곡선 모형을 의미한다.

시장예측에 적용될 성장곡선 모형의 도출 작업은 과거 시계열 시장자료를 가장 잘 설명할 수 있는 해당 성장곡선의 매개변수를 추정하는 작업이라 할 수 있다. 대표적인 매개변수 추정 방법으로는 최소자승법(OLS: ordinary least squares)(Bass, 1969; Young and Ord, 1989), 최우(ML: maximum likelihood) 추정법(Schmittlein and Mahajan, 1982; Olson, 1982) 그리고 비선형 최소자승법(NLS: nonlinear least squares)(Srinivasan and Mason, 1986) 등이 있다. 먼저 최소자승법의 경우에는 일반적인 다중 회귀분석에서와 동일한 방법으로 매개변수를 추정하므로 가장 간단하다는 장점을 가지지만 일부 매개변수가 알려져 있다는 가정이 필요하다거나(Young and Ord, 1989), 다중공선성 문제 및 매개변수의 표준오차(standard error)를 직접적으로 도출할 수 없다는 등의 단점 또한 지니고 있다(Heeler and Hustad, 1980). 선형변환이 불가능한 성장곡선이나 본 논문에서 이용하고 있는 Logistic 성장곡선과 Gompertz 성장곡선과 같이 선형이 가능하지만 잠재 시장 규모를 알지 못하면 매개변수 추정이 불가능하다. 또한 ML 추정법의 경우에는 최소자승법이 가지는 단점을 어느 정도 보완할 수 있으나 매개변수 추정치의 표준오차를 과소평가함으로써 추정치의 통계적 신뢰도에 대한 잘못된 추론을 제공할 수 있다는 단점을 지니고 있다(Schmittlein and Mahajan, 1982). 또한 우도함수(likelihood function)를 구성하는 것이 매우 어려우며 이러한 경우 역시 초기값을 설정해 주어야 한다는 문제점을 가지고 있다. 한편 비선형 최소자승법은 가장 널리 이용되는 방법으로 최소자승법이나 ML 추정법의 한계를 극복할 수 있는 것으로 알려져 있다(Srinivasan and Mason, 1986).

비선형 최소자승법을 이용하기 위해서는 각 매개변수의 초기값을 설정해 주어야 하는데 초기값을 어떻게 설정하는가에 따라 최종적인 추정 결과가 크게 달라질 수 있으므로 이 작업은 매우 신중히 이루어져야 한다. 이제까지 주로 사용되어 온 초기값 설정 방법은 두 가지로, 하나는 해당 성장곡선을 유사한 특성을 가진 것으로 판단되는 다른 상품에 대한 시장예측에 적용시킨 기존 연구들을 대상으로 메타 분석을 실시하여 그 결과를 이용하는 것이고, 다른 하나는 해당 성장곡선을 선형 변환시킨 후 최소자승법을 이용하여 도출된 각 매개변수의 추정치를 비선형 최소자승법을 위한 초기값으로 설정하는 것이다. Bass 성장곡선 모형과 같은 특수한 형태의 성장곡선에 대해서는 직접적인 선형변환이 가능하지만(Bass, 1969), 그 외 대부분의 비선형 성장곡선 모형을 선형변환시키기 위해서는 잠재시장규모와 같은 일부 매개변수를 상정해야 한다. 따라서 위 두 가지 방법 모두 초기값 설정 과정에 분석자의 주관적 판단이 개입되어야 한다는 문제점을 가지고 있다. 잠재시장규모의 주관적 상정은 관심을 갖는 시장규모의 척도가 보급률 등과 같이 상한값을 가지는 경우가 아닌 매출액이나 보급대수 등일 경우 특히 위험하다.

본 논문에서는 비선형 최소자승법을 이용하여 성장곡선 모형의 매개변수를 추정하는 과정에서 발생하는 초기값 설정 문제의 해결 방법을 다루고 있다. 고려된 성장곡선 모형은 Logistic

성장곡선과 Gompertz 성장곡선의 두 가지이다. 일반적으로 모든 대상 시장이나 상품에 대해 항상 최적의 예측 결과를 보장하는 특정 성장곡선 형태는 존재하지 않는 것으로 알려져 있지만, 가장 널리 이용되어 온 성장곡선으로 위 두 가지를 들 수 있으며 이들만을 고려하더라도 대부분의 상품에 대한 안정적 시장예측이 가능한 것으로 알려져 있다(Young and Ord, 1989). 본 연구의 주된 목적은 분석자의 주관적 판단을 배제하고 정량적 방법만으로 이들 성장곡선의 각 매개변수에 대한 초기값을 설정할 수 있는 절차를 개발하는 데 있다.

본 논문의 구성은 다음과 같다. 2절에서는 비선형 최소자승법 중 하나인 Gauss-Newton 방법에 대해 설명하고, 3절에서는 주어진 시계열 시장자료만으로 초기값을 설정할 수 있는 정량적 방법을 제안하며, 4절에서는 제안된 방법을 국내시장에서의 프린터 매출액과 국내 호스트 컴퓨터 등록대수의 예측에 적용시킨 결과를 나타낸다. 마지막으로 5절에서는 결론 및 추후 연구과제에 대해 기술한다.

## 2. 비선형 최소자승법을 이용한 성장곡선 모형의 매개변수 추정

본 절에서는 비선형 최소자승법을 이용하여 일반적인 성장곡선 모형의 매개변수를 추정하는 방법에 대하여 설명하고자 한다. 다음과 같은 예측식을 생각해 보자.

$$N_t = f(t; \underline{\gamma}) + \epsilon_t \quad (1)$$

여기서  $N_t$ :  $t$  시점에서의 해당 상품의 시장규모

$\underline{\gamma} = (\gamma_1, \gamma_2, \dots, \gamma_p)$ : 매개변수 벡터

$f(t; \underline{\gamma})$ : 시간  $t$ 에 대한 비선형 성장곡선 함수

$\epsilon_t$ :  $t$  시점에서의 오차항

주어진 시계열 시장자료의 개수를  $T$ 라 할 때 비선형 최소자승법에서는 다음 목적식을 최소화하는 매개변수 벡터  $\underline{\gamma}$ 를 추정한다.

$$\text{Min} \sum_{t=1}^T [N_t - f(t; \underline{\gamma})]^2 \quad (2)$$

식 (2)에 주어진 목적식의 해를 구하는 방법으로는 Newton 방법, Gauss-Newton 방법, Marquart 방법 등이 있는데 본 논문에서는 Gauss-Newton 방법을 이용하였다(Neter et al., 1996).

매개변수에 대한 초기값  $\underline{\gamma}^{(0)} (= (\gamma_1^{(0)}, \gamma_2^{(0)}, \dots, \gamma_p^{(0)}))$ 가 주어졌을 때 Gauss-Newton 방법에서는 주어진 초기값을 식 (2)에 나타낸 오차제곱합(Sum of Squared Error)이 최소값을 갖도록 연속적으로 개선시켜 최적의 매개변수 벡터  $\underline{\gamma}$ 를 도출한다.  $i$  번째 iteration에서 얻어진 매개변수 벡터를  $\underline{\gamma}^{(i)} (= (\gamma_1^{(i)}, \gamma_2^{(i)}, \dots, \gamma_p^{(i)}))$ 라 하고 오차제곱합을  $SSE^{(i)}$ 라 하자.  $SSE^{(i)}$ 는 다음

식으로 얻을 수 있다.

$$SSE^{(i)} = \sum_{t=1}^T [N_t - f(t; \underline{\gamma}^{(i)})]^2 \quad (3)$$

(i+1)번째 iteration에서 Taylor 급수를 이용하면  $f(t; \underline{\gamma})$ 에 대한 다음 근사식을 얻을 수 있다.

$$f(t; \underline{\gamma}) \approx f(t; \underline{\gamma}^{(i)}) + \sum_{k=1}^p \left[ \frac{\partial f(t; \underline{\gamma})}{\partial \gamma_k} \right]_{\underline{\gamma} = \underline{\gamma}^{(i)}} (\gamma_k - \gamma_k^{(i)}) \quad (4)$$

따라서 식 (1)의 예측식은 다음과 같이 나타낼 수 있다.

$$N_t \approx f_t^{(i)} + \sum_{k=1}^p D_{tk}^{(i)} \delta_k^{(i)} + \varepsilon_t \quad (5)$$

여기서

$$\begin{aligned} f_t^{(i)} &= f(t; \underline{\gamma}^{(i)}) \\ D_{tk}^{(i)} &= \left[ \frac{\partial f(t; \underline{\gamma})}{\partial \gamma_k} \right]_{\underline{\gamma} = \underline{\gamma}^{(i)}} \\ \delta_k^{(i)} &= \gamma_k - \gamma_k^{(i)} \end{aligned} \quad (5)$$

한편  $M_t^{(i)}$ 를  $N_t - f_t^{(i)}$ 로 놓으면

$$M_t^{(i)} \approx \sum_{k=1}^p D_{tk}^{(i)} \delta_k^{(i)} + \varepsilon_t \quad (6)$$

이 되며, T개의 시계열 시장자료가 주어졌을 때 이들을 다음과 같은 행렬 형태로 나타낼 수 있다.

$$M^{(i)} \approx D^{(i)} \delta^{(i)} + \varepsilon \quad (7)$$

여기서

$$M^{(i)} = \begin{bmatrix} N_1 - f_1^{(i)} \\ N_2 - f_2^{(i)} \\ \vdots \\ N_T - f_T^{(i)} \end{bmatrix}, \quad D^{(i)} = \begin{bmatrix} D_{11}^{(i)} & D_{12}^{(i)} & \cdots & D_{1p}^{(i)} \\ D_{21}^{(i)} & D_{22}^{(i)} & \cdots & D_{2p}^{(i)} \\ \vdots & \vdots & \ddots & \vdots \\ D_{T1}^{(i)} & D_{T2}^{(i)} & \cdots & D_{Tp}^{(i)} \end{bmatrix},$$

$$\delta^{(i)} = \begin{bmatrix} \delta_1^{(i)} \\ \delta_2^{(i)} \\ \vdots \\ \delta_p^{(i)} \end{bmatrix}, \quad \varepsilon = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_T \end{bmatrix}$$

식 (7)에 대해 선형 회귀분석에서의 동일한 방법, 즉 최소자승법을 적용시키면  $\delta^{(i)}$ 의 추정치  $\hat{\delta}^{(i)}$ 를 다음과 같이 도출할 수 있다.

$$\hat{\delta}^{(i)} = (D^{(i)'} D^{(i)})^{-1} D^{(i)'} N^{(i)} \quad (8)$$

식 (5)를 이용하면 각 매개변수의 새로운 추정치를 다음과 같이 얻을 수 있다.

$$\gamma_k^{(i+1)} = \gamma_k^{(i)} + \hat{\delta}_k^{(i)}, \quad k = 1, 2, \dots, p \quad (9)$$

여기서 (i+1)번째 iteration에서의 오차자승합  $SSE^{(i+1)}$ 을  $SSE^{(i)}$ 와 비교하여 뚜렷한 감소 효과를 관측할 수 있으면 새로운 iteration을 시작하고, 그렇지 않으면  $\underline{\gamma}^{(i)}$ 를 최종적인 매개변수 벡터의 추정치로 결정한다. 본 논문에서는 SAS의 NLIN 프로시저를 사용하였다.

### 3. 경제지표 및 구간분할탐색을 이용한 초기값 설정 방법

본 절에서는 비선형 최소자승법을 이용한 성장곡선 모형의 매개변수 추정시 주어진 시계열 시장자료에 기초하여 정량적으로 적절한 초기값을 설정할 수 있는 방법을 제시한다. 분석자의 주관적 판단 등에 근거하여 임의로 설정된 초기값은 성장곡선의 매개변수 추정에 이용할 경우 몇 가지 치명적인 문제점이 나타날 수 있다. 가장 많이 지적되는 문제점은 설정된 초기값에 영향을 받아 식 (2)에 주어진 목적식에 대한 최적해(global optimum)가 아닌 부분최적해(local optimum)를 매개변수의 추정치로 도출한다거나, 특정값으로의 수렴이 이루어지지 않아야 아예 매개변수의 추정치 자체를 도출할 수 없는 경우가 발생할 수 있다는 점이다. 또한 성장곡선 모형을 시장수요 예측에 이용할 경우 가장 중요하게 다루어지는 매개변수는 잠재 시장규모, 즉 S곡선의 포화점(saturation point)이라 할 수 있는데 분석자의 잘못된 판단은 장기 예측 시점에서 현실과 크게 동떨어진 예측치를 제공하게 되는 결과로 이어진다. 특히 관심을 갖는 시장규모의 척도가 보급률 등과 상한값을 가지는 개념이 아닌 매출액이나 보급대수 등인 경우 이러한 위험은 더욱 가중된다고 할 수 있다.

본 연구에서는 예측하고자 하는 시장규모의 척도가 매출액이나 보급대수 등인 경우 적절한 경제지표를 도입하여 이들을 표준화시키고, 표준화된 과거 시계열 시장자료를 이용하여 성장곡선 모형의 각 매개변수가 가질 수 있는 값의 범위를 결정한 후, 구간분할탐색(grid search)을 통하여 해당 목적식을 최소화시키는 매개변수 값들의 조합을 초기값으로 설정한다. 본 논문에서 고려된 성장곡선 모형은 Logistic 성장곡선과 Gompertz 성장곡선의 두 가지로 각 모형의 예측식은 다음과 같다.

$$\text{Logistic 성장곡선: } N_t = \frac{m}{1 + \alpha \exp(-\beta t)}, \quad \alpha, \beta > 0 \quad (10)$$

$$\text{Gompertz 성장곡선: } N_t = m \exp(-\alpha \exp(-\beta t)), \quad \alpha, \beta > 0 \quad (11)$$

여기서 m(잠재시장규모),  $\alpha, \beta$ : 추정해야 할 매개변수  
 식 (10)과 식 (11)에서의  $N_t$ 는 t시점에서의 표준화된 시장규모를 의미하는 것으로, 본 연구에서는 매출액이나 보급대수 등으로 주어진 시계열 시장자료의 표준화를 위한 경제지표로 해당 시점에서의 국민 총생산액과 인구수를 도입하였다. 즉

관심을 갖는 시장규모의 척도가 매출액인 경우에는

$$N_t = \frac{N_p(t)}{GDP(t)} \cdot 100, t = 1, 2, \dots, T \quad (12)$$

여기서  $N_p(t)$ :  $t$ 시점에서의 해당 상품의 매출액  
 $GDP(t)$ :  $t$ 시점에서의 국내 총생산액

으로 그리고 관심을 갖는 시장규모의 척도가 보급대수인 경우에는

$$N_t = \frac{N_s(t)}{POP(t)} \cdot 100, t = 1, 2, \dots, T \quad (13)$$

여기서  $N_s(t)$ :  $t$ 시점에서의 해당 상품의 보급대수  
 $POP(t)$ :  $t$ 시점에서의 인구수

으로 주어진 시계열 시장자료를 표준화시킨다. 이와 같은 표준화 개념을 도입한 목적은 잠재시장규모 및 다른 매개변수들이 가질 수 있는 값의 범위를 설정하는데 있다.

식 (12)와 식 (13)에 의해 표준화된 시장자료를 이용하여 먼저 잠재시장규모가 가질 수 있는 값의 범위를 쉽게 결정할 수 있으며 다음과 같다.

$$N_T \leq m^{(0)} \leq 100 \quad (14)$$

또한 잠재시장규모의 범위가 결정되면 이를  $k$ 개의 구간으로 나눔으로써  $(k+1)$ 개의 경계값을 얻을 수 있다. 여기서 잠재시장 규모에 따른  $\alpha$ 와  $\beta$ 의 변화량은 어떻게 되는가를 살펴보기 위하여 이를 임의의  $k$ 개로 나누었으며 이때의  $k$ 는 변화량의 형태를 파악할 수 있는 정도면 충분하다. 이는 각각  $m$ 의 구체적인 값보다는 최대값과 최소값을 파악하여 구간으로 이용하고자 함이다.

식 (10)과 식 (11)을 선형변환 시키면 식 (15), 식 (16)과 같다. 식 (15)와 식 (16)의 우변은  $\alpha + \beta t$ 로 나타낼 수 있으며 이는 선형의 형태로 나타나기 때문에 선형 회귀 분석이 가능하게 된다. 잠재시장규모의  $i$ 번째 경계값  $m_i$ 를 60이라 한다면 좌변은  $m_i$ 를 포함한 시계열 시장자료로 표현할 수 있으므로 최소자승법을 이용하여 나머지 매개변수  $\alpha_i$ 와  $\beta_i$ 의 값을 추정할 수 있다.

선형 변환된 Logistic 성장곡선

$$\ln \frac{m_i - N_t}{N_t} = \ln \alpha_i - \beta_i t, i = 1, 2, \dots, k+1 \quad (15)$$

선형 변환된 Gompertz 성장곡선

$$\ln \left( -\ln \frac{N_t}{m_i} \right) = \ln \alpha_i - \beta_i t, i = 1, 2, \dots, k+1 \quad (16)$$

위의 과정을 통하여 나머지 두개의 매개변수의 범위는 다음과 같이 구해지며 구체적인 계산 과정은 응용사례에서 자세히

보여주고 있다.

$$\min [\alpha_1, \alpha_2, \dots, \alpha_{k+1}] \leq \alpha^{(0)} \leq \max [\alpha_1, \alpha_2, \dots, \alpha_{k+1}] \quad (17)$$

$$\min [\beta_1, \beta_2, \dots, \beta_{k+1}] \leq \beta^{(0)} \leq \max [\beta_1, \beta_2, \dots, \beta_{k+1}] \quad (18)$$

$\alpha$ 와  $\beta$ 의 구간이 결정되면 이 구간 내에서 추천 초기값의 조합을 찾기 위하여 각 매개변수 구간을  $n$ 개로 나누어야 한다.  $n$ 개로 나눈 매개변수는 각각  $(n+1)$ 개의 경계값을 가진다.  $n$ 을 결정하기 위하여 두 가지 성장곡선 모형에 관하여 개수에 따른 오차 자승합의 변화를 관찰하였다.  $n$ 의 수를 증가시켜 그때마다의 오차자승합을 구한다. 구간을 증가시킬 때 오차 자승합이 뚜렷한 감소효과를 나타내지 않을 경우 이전의 구간을 최종 구간의 개수  $n$ 으로 결정한다. 본 논문에서는 오차자승합의 변화가 둔화되는 값을 0.001로 설정하였다.

구간의 개수가 결정되면  $n$ 으로 나눈 매개변수의 경계값에 대한  $(n+1)^3$ 개의 조합이 얻어진다. 최종적인 초기값은 오차자승합이 가장 작은 매개변수들의 조합으로 설정된다.

$j$ 번째 조합의 오차자승합  $CSSE^{(m_i, \alpha_j, \beta_j)}$ 는 다음과 같이 얻어진다.

$$CSSE^{(m_i, \alpha_j, \beta_j)} = \sum_{t=1}^T (N_t^{(j)} - N_t)^2, j = 1, 2, \dots, (n+1)^3 \quad (19)$$

여기서

$$N_t^{(j)} = \frac{m_j}{1 + \alpha_j \exp(-\beta_j t)} \quad (\text{Logistic 성장곡선})$$

또는

$$N_t^{(j)} = m_j \exp(-\alpha_j \exp(-\beta_j t)) \quad (\text{Gompertz 성장곡선})$$

초기값 설정을 포함한 전체적인 비선형 최소자승법을 이용한 성장곡선 모형의 매개변수 추정 절차를 <그림 1>에 나타내었다.

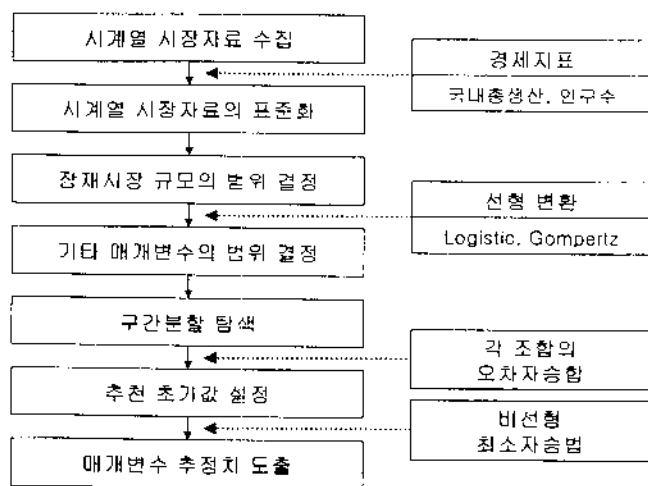


그림 1. 비선형 최소자승법을 이용한 성장곡선 모형의 매개변수 추정 절차.

4. 적용사례

3절에서 제안한 비선형 최소자승법에서의 초기값 설정방법을 국내시장에서의 프린터 매출액과 국내 호스트 컴퓨터 등록대수의 예측에 적용하였다.

4.1 국내시장에서의 프린터 매출액

1986년부터 1998년까지의 국내시장에서의 프린터 연간 매출액과 표준화된 연간 매출액을 <표 1>에 나타내었다. 표준화된 연간 매출액은 연간 매출액을 해당 연도의 국내 총생산에 대한 백분율로 계산되었다.

<표 1>의 자료 중 1986년부터 1993년까지의 자료를 이용하여 Logistic 성장곡선과 Gompertz 성장곡선의 매개변수를 추정하였으며, 그 이후의 자료는 각 모형의 예측력 검증에 위하여 사용되었다. 도출된 잠재시장규모의 범위는 [0.629, 100]이며 이를 10개의 구간으로 나눔으로써 얻어진 11개의 경계값에 대해 식 (15)와 식 (16)을 이용한 선형변환을 거친 후 선형 회귀분석을 실시하여 매개변수  $\alpha$ 와  $\beta$ 의 범위를 도출하였다. <표 2>는 Logistic 성장곡선에 대하여 변형된 데이터를 이용한  $t$ 에 관한 단순회귀분석 과정을 나타내고 있다.

선형회귀분석으로 구한  $a$ 와  $b$ 의  $p$ -value는 모두 통계적으로 유의하였다. 표에서 구한  $a$ 와  $b$ 에 대한 방정식을 풀어  $\alpha$ 와  $\beta$ 의 값을 구한 후 이들 중 최대값과 최소값을 매개변수의 변화 범위로 나눌 수 있다. 각각의  $m$ 에 대하여  $\alpha, \beta$ 를 계산한 다음 매개변수의 범위를 도출하면 <표 3>과 <표 4>와 같다.

표 1. 국내시장에서의 프린터 연간 매출액 및 표준화된 매출액

연도	프린터 연간 매출액(×100)	국내 총생산액(×100)	표준화된 프린터 연간 매출액(%)	
1986	1	75,063	94,861,700	0.079
1987	2	153,328	111,197,700	0.138
1988	3	285,186	132,111,800	0.216
1989	4	457,748	148,197,000	0.309
1990	5	678,863	178,796,800	0.380
1991	6	962,740	216,510,900	0.445
1992	7	1,326,376	245,699,600	0.540
1993	8	1,744,939	277,496,500	0.629
1994	9	2,252,784	323,407,100	0.697
1995	10	2,937,983	377,349,800	0.779
1996	11	3,467,023	418,479,000	0.828
1997	12	3,837,795	453,276,400	0.847
1998	13	4,088,024	444,366,500	0.920

\* 자료출처: 통계청

표 3. Logistic 성장곡선의 각 매개변수 범위(프린터)

매개변수	매개변수별 변화 범위	
	최소값	최대값
$m$	0.629	100
$\alpha$	14.031	1256.399
$\beta$	0.283	0.409

표 2.  $m$ 의 구간에 따른 Logistic 성장곡선  $a, b$  및  $\alpha, \beta$ 의 변화(프린터)

	년도	$m_1$	$m_2$	$m_3$	$m_4$	$m_5$	$m_{10}$	$m_{11}$
		0.6288	10	25	...	75	85	100
각 $m$ 에 따라 선형 변환된 시장자료	1	2.454	4.831	5.752	...	6.853	6.978	7.141
	2	1.833	4.270	5.195	...	6.279	6.422	6.586
	3	1.290	3.814	4.743	...	5.848	5.973	6.136
	4	0.805	3.446	4.381	...	5.488	5.614	5.777
	5	0.491	3.232	4.172	...	5.218	5.407	5.570
	6	0.222	3.068	4.011	...	5.122	5.248	5.411
	7	-0.160	2.846	3.814	...	4.927	5.053	5.216
	8	-0.527	2.702	3.657	...	4.733	4.899	5.063
$Z_i = a_i + b_i t$	$a$	2.641	4.835	5.750	...	6.848	6.974	7.136
	$b$	-0.409	-0.290	-0.285	...	-0.280	0.283	-0.283
$\ln\left(\frac{m_i - T(t)}{T(t)}\right)$ = $\ln \alpha_i - \beta_i t$ = $a_i + b_i t$	$\alpha$	14.031	125.839	314.241	...	942.342	1067.954	1256.399
	$\beta$	0.409	0.290	0.285	...	0.283	0.283	0.283
RMSE		0.0014	0.0084	0.0091	...	0.0095	0.0095	0.0095

표 4. Gompertz 성장곡선의 각 매개변수 범위(프린터)

매개변수	매개변수별 변화 범위	
	최소값	최대값
$m$	0.629	100
$\alpha$	3.162	6.929
$\beta$	0.047	0.236

표 5. 구간 분할에 따른 오차자승합 및 변화량

$n$	Logistic SSE	SSE변화량	Gompertz SSE	SSE변화량
5	0.009891		0.0105740	
10	0.005203	0.0046880	0.0077770	0.0027970
15	0.004086	0.0011170	0.0041900	0.0035870
20	0.004000	0.0000860	0.0036600	0.0005300

매개변수의 적절한 탐색 구간을 설정하기 위하여 구간의 개수에 따른 오차자승합의 변화를 관찰하였다.

<표 5>에서  $n$ 을 5씩 증가하면서 변화시킨 경우 구간을 15 단계로 나눈 경우 SSE와 20단계로 나눈 경우 SSE의 차이가 0.001 이하로 변화하는 것을 알 수 있다.

<그림 2>에서와 같이 두 가지 성장곡선 모두 구간 분할에 따른 오차자승합의 차이가 15개 이상의 구간분할에서는 거의 없었으므로 탐색수는  $(15+1)^3$ 개의 매개변수 조합 중 가장 작은 오차자승합을 가지는 매개변수들의 조합을 추천 초기값으로 설정하였다. 구해진 추천 초기값을 비선형 최소자승법의 초기값으로 설정하여 각각의 성장곡선 모형의 매개변수 추정치를 <표 6>과 같이 도출하였다.

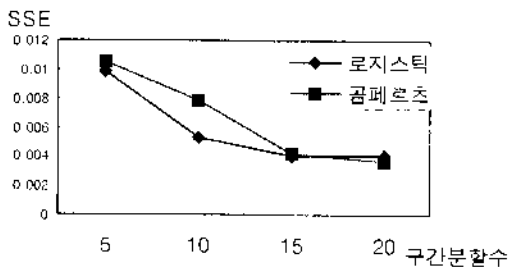


그림 2. 구간 분할에 따른 오차자승합의 변화.

표 6. 성장곡선 추천 초기값 및 매개변수 추정치(프린터 예)

성장곡선 모형		$m$	$\alpha$	$\beta$	SSE
Logistic 성장곡선	추천 초기값	1.000	14.031	0.403	0.0052
	최종 매개변수	0.778 (0.075)	11.220 (1.041)	0.469 (0.042)	0.0016
Gompertz 성장곡선	추천 초기값	1.000	3.162	0.227	0.00419
	최종 매개변수	1.038 (0.137)	3.115 (0.109)	0.225 (0.030)	0.00078

\* ( )안은 표준오차

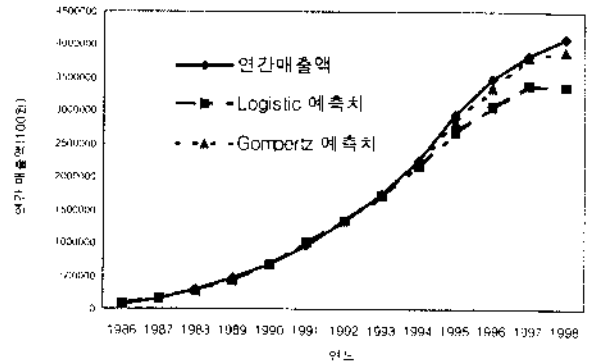


그림 3. 실제 프린터 연간 매출액과 성장곡선별 예측치.

<표 6>에서 탐색을 거친 추천 초기값과 비선형 최소자승법을 거친 매개변수 추정치 사이에는 약간의 차이가 있으며 오차자승합 역시 비선형 최소자승법을 거친 값이 더 작게 나타나고 있다. 실제 프린터 연간 매출액과 각 성장곡선을 통하여 도출된 프린터 연간 매출액의 예측치를 <그림 3>에 나타내었다.

<그림 3>에서 볼 수 있듯이 Gompertz 곡선이 Logistic 곡선에 비해 비교적 좋은 추정치를 나타내고 있다. 1993년 이후 실제 연간 매출액 Logistic 곡선에서는 약간 작게 나타나고 있다. 그러나 두 곡선 모두 표준 오차값이 작으며 구간 탐색 방법을 사용하여 구한 비선형 최소자승법의 초기값은 두 곡선 모두 최종 매개변수값을 도출하는 데 문제가 없었다. 비선형 최소자승법에서 각 매개변수 추정시 마이너스값이 도출되거나 매개변수값이 수렴되지 않는 문제점들은 발생하지 않았다.

#### 4.2 국내 호스트 컴퓨터 등록대수

호스트 컴퓨터는 인터넷의 양적 팽창을 나타내는 대표적인 지수로 IP 주소를 가지고 있으며 호스트명이 네임서버에 등록되어 있는 컴퓨터를 의미한다. 1993년부터 2000년까지의 국내 호스트 컴퓨터의 누적 등록대수와 조정된 누적 등록대수를 <표 7>에 나타내었다. 표준화된 누적 등록대수를 해당 연도의 총 인구수에 대한 백분율로 계산되었다.

표 7. 호스트 등록대수 및 변형된 자료

연도	호스트 컴퓨터 등록대수(대)	총 인구수(명)	인구대비 호스트 컴퓨터 등록대수(%)	
1993	1	7,650	44,195,000	0.017
1994	2	13,856	44,642,000	0.031
1995	3	36,644	45,093,000	0.081
1996	4	73,191	45,545,000	0.161
1997	5	131,005	45,991,000	0.285
1998	6	202,510	46,430,000	0.436
1999	7	296,300	46,858,000	0.632
2000	8	460,974	47,293,000	0.975

\*데이터 출처: 통계청

표 8. Logistic 성장곡선의 각 매개변수 범위(호스트 컴퓨터 등록대수)

매개변수	매개변수 변화구간	
	최소값	최대값
m	0.285	100
$\alpha$	134.730	12141.111
$\beta$	0.725	0.802

표 9. Gompertz 성장곡선의 각 매개변수 범위(호스트 컴퓨터 등록대수)

매개변수	매개변수 변화구간	
	최소값	최대값
m	0.285	100
$\alpha$	5.902	9.686
$\beta$	0.101	0.299

<표 7>의 자료중 1993년부터 1997년까지의 자료를 이용하여 Logistic 성장곡선과 Gompertz 성장곡선의 매개변수를 추정하였으며 그 이후의 자료는 각 성장곡선 모형의 예측력 검증에 위하여 사용되었다. 도출된 잠재시장규모의 범위는 [0.285, 100]이며 이를 10으로 나눔으로써 얻어진 11개의 경계값에 대해 식 (15)와 식 (16)를 이용한 선형변환을 거친 후 선형회귀 분석을 실시하여 매개변수  $\alpha$  와  $\beta$  범위를 도출하였다(<표 8> 과 <표 9> 참조).

<표 8>과 <표 9>에 나타난 매개변수의 범위를 15개로 분할하였으며 이 역시 오차자승합의 변화가 0.001 이하가 되는 시점의 개수이다. 그러므로 총 탐색 구간 역시  $(15+1)^3$ 개이다. 탐색을 거쳐 오차자승합이 가장 작은 값을 가지는 매개변수 조합을 최적 초기값으로 정의하고 이를 비선형 최소자승법에서 초기값으로 설정하여 각각의 성장곡선 모형의 매개변수 추정치를 <표 10>과 같이 도출하였다.

실제 호스트 컴퓨터의 누적 등록대수와 각 성장곡선을 통하

표 10. 성장곡선별 최적 초기값 및 매개변수 추정치(호스트 컴퓨터 등록대수 예)

성장곡선 모형		m	$\alpha$	$\beta$	SSE
Logistic 성장곡선	최적 초기값	1.000	134.730	0.802	0.000113
	최종 매개변수	0.823 (0.438)	116.107 (27.817)	0.823 (0.096)	0.0000823
Gompertz 성장곡선	최적 초기값	6.000	7.502	0.181	0.000075
	최종 매개변수	5.296 (5.679)	7.186 (0.699)	0.180 (0.054)	0.0000361

\* ( ) 안은 표준오차

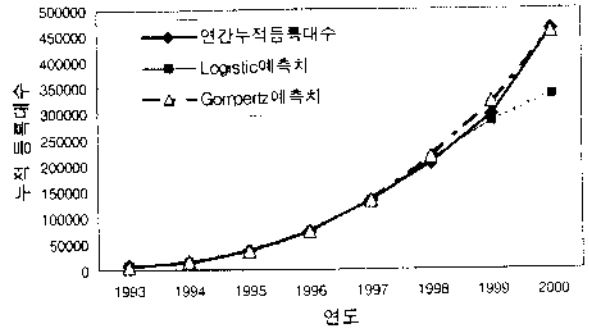


그림 4. 실제 호스트 컴퓨터 누적 등록대수와 성장곡선별 예측치.

여 도출된 호스트 컴퓨터 등록대수의 예측치를 <그림 4>에 나타내었다.

<그림 4>에서와 같이 1997년 이후 Gompertz 성장곡선과 실제 시장자료를 비교해 볼 때 비교적 정확한 추세를 그리고 있다. 반면 Logistic 성장곡선은 실제보다 작은 추정치를 나타내고 있으며 매개변수의 표준오차에서 역시 큰 수치를 보이고 있음을 알 수 있다. 그러나 두 곡선 역시 프린터 매출액 분석에서와 같이 매개변수값이 마이너스가 나오거나 수렴하지 않는 경우는 없었으며 Logistic 성장곡선이 과추정치를 보이는 것은 호스트 컴퓨터 등록대수의 시장자료에 Logistic 성장곡선이 잘 맞지 않음을 나타내준다.

### 5. 결론 및 추후 연구과제

본 논문에서는 확산 모형의 매개변수를 추정하기 위하여 비선형 최소자승법을 이용하였으며 이 과정에서 발생하는 초기값 설정 문제에 대한 해결 방안을 제시하였다. 매개변수 추정시 발생하는 초기값 문제를 해결하기 위하여 잠재시장 규모를 경제지표로 환산하였다. 이를 통하여 발생 가능한 구간을 모두 고려하는 초기값 구간을 설정하였다.

제안된 방법을 이용하여 매개변수를 추정한 수치에 대해서는 기존에 초기값을 잘못 설정하여 발생하는 문제점인 수렴되지 않는 경우는 발생하지 않았으며 기존의 방법과 같이 매개변수의 추정값이 마이너스값이 발생하는 경우도 없었다.

그러나 본 논문에서는 비교적 단순한 모형에 대해서만 분석한 한계점을 가지고 있다. 실제 정보통신상품과 같이 기술성장이 급변하는 제품은 그 단일 제품뿐만 아니라 경쟁 관계나 상호 보완 관계에 있는 타제품과 같이 존재하고 있는데, 본 논문에서는 이러한 경우는 고려되지 않고 있다. 그러므로 추후에는 다세대(multi-generation) 확산 모형이나 반복구매(repeat purchase)를 고려한 확산 모형, 다양한 시장 변수를 고려한 모형과 같은 여러 가지 모형의 복합된 형태에 대하여 본 논문에서 제시한 방법을 적용 할 수 있는가에 대한 검증이 필요할 것이다. 또한 선형 변환이 어려운 곡선에 대하여 어떠한 방법으로

매개변수 구간을 설정할 것인가에 대한 연구가 더 이루어져야 할 것이다.

**참고문헌**

국민총생산(GDP), 통계청, <http://www.nso.go.kr/stat/majorecono/econoDB.htm>, 2000  
 임종인, 오형식 (1992), 신제품 수요예측 방법론 연구, *산업공학회지*, 18(2), 51-63  
 주민등록 인구통계 (2000), *통계청*  
 황정연, 강병용, 전경표 (1999), 음성 트래픽과 인터넷 트래픽 추정에 관한 연구, *IE Interfaces*, 12(4), 625-634.  
 Bass, F. M. (1969), A New Product Growth Model For Consumer Durables, *Management Science*, 15, 215-227.  
 Heeler, R. M. and Hustad, T. P. (1980), Problems in Predicting New Product Growth for Consumer Durables, *Management Science*, 26, 1007-1020.

Meade, N. and Islam, T. (1998), Technological Forecasting: Model Selection, Model Stability, and Combining Models, *Management Science*, 44, 1115-1130.  
 Meade, N. (1984), The Use of Growth Curves in Forecasting Market Development A Review and Appraisal, *Journal of Forecasting*, 3, 429-451.  
 Neter, J., Kutner, M. H., Nachrsheim, C. G. and Wasserman, W. (1996), *Applied Linear Statistical Models*, 4th edition, IRWIM Press, 539-546.  
 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 Mahajan, V. (1982), Maximum Likelihood Estimation for an Innovation Diffusion Model of New Product Acceptance, *Marketing Science*, 1, 57-78.  
 Srinivasan, V. and Mason, C. H. (1986), Non-Linear Least Squares Estimation of New Product Diffusion Models, *Marketing Science*, 5, 169-178.  
 Young, P. and Ord, J. K. (1989), Model Selection and Estimation for Technological Growth Curves, *International Journal of Forecasting*, 5, 501-513.  
 Young, P. (1993), Technological Growth Curve: A Competition of Forecasting Models, *Technological Forecasting and Social Change*, 44, 375-389.



**염세경**  
 동국대학교 산업공학 학사  
 포항공과대학교 산업공학 석사  
 현재: 포항공과대학교 위촉연구원  
 관심분야: 수요예측 및 Data mining



**김지수**  
 포항공과대학교 산업공학 학사  
 포항공과대학교 산업공학 석사  
 포항공과대학교 산업공학 박사  
 현재: Purdue University Post Doc.  
 관심분야: 통신 네트워크 및 ATM



**홍승표**  
 동국대학교 산업공학 학사  
 포항공과대학교 산업공학 석사  
 현재: 전자통신연구원 정보체계연구팀 연구원  
 관심분야: 전자상거래 및 e-Business



**전치혁**  
 서울대학교 산업공학 학사  
 한국과학기술원 산업공학 석사  
 U.C. Berkeley 산업공학 박사  
 현재: 포항공과대학교 산업공학과 주임교수  
 관심분야: 응용통계 및 Data mining



**강희일**  
 한국외국어대학 일어학과 학사  
 충남대학교 경영학 석사  
 충남대학교 경영학 박사과정 수료  
 현재: 한국전자통신연구원 정보체계연구팀 선임연구원  
 관심분야: Service(B2C)/Industry Marketing(B2B) 과 Internet Marketing의 접목 분야