

# Bass 확산 모형 계수의 추정치에 대한 민감도 분석 (Sensitivity analysis of the parameter estimates in the Bass Diffusion Model)

홍정식\*, 김영재\*, 안재경\*, 김태구\*\*

\* 서울산업대학교 산업정보시스템공학과

\*\* 서울대학교 산업공학과

## Abstract

신제품이나 서비스의 수요 예측을 Bass 확산 모형을 토대로 수행할 때의 가장 큰 문제점은 모형의 파라미터 추정에 필요한 데이터가 충분치 않다는 것이다. 따라서 Bass 확산 모형의 핵심적인 두 파라미터인 혁신 계수(p)와 모방 계수(q)의 추정을 시도할 때, 어느 정도의 데이터 개수가 요구되는지를 파악하는 것은 매우 현실적인 중요성을 갖는 문제이다.

이제까지의 연구는 주로 기존의 판매 데이터를 토대로 Bass 모형의 파라미터를 추정할 때, 생기는 다양한 문제점 파악에 집중되었다. 시뮬레이션의 경우는 Bass 모형에 랜덤 오차를 추가하여 실시하였다. 이 경우 데이터 개수가 계수추정에 미치는 영향은 도출되나 각 계수별 민감도 분석이 제대로 이루어지지 못하는 한계를 가지고 있다. 따라서 본 논문에서는 시뮬레이션에서 예측치를 발생시킬 때 랜덤 오차 대신, 혁신 계수와 확산 계수의 변동만을 주는 방법을 도입한다. 결과는 다음과 같다. 첫째, p 변동보다는 q 변동이 예측치의 오차에 대해 보다 중요하다. 둘째, 오차가 잠재 수요의 30%이하로 떨어지기 위해서는 수요가 최대 도달하는 시점이  $t^*$ 일 경우,  $t^* + 1$ 까지 데이터가 요구된다.

## 1. 서론

Bass 확산 모형은 신제품이나 신규 서비스의 수요 예측 및 시장 진입 시점 그리고 최적 가격 및 광고 전략에 가장 널리 활용되는 모형이다.[1,11] 본 논문은 주로 신제품이나 신규 서비스의 수요 예측과 관련하여 Bass 확산 모형을 다루고자 한다. Bass 확산 모형이 신제품 수요 예측에 폭넓게 활용되는 이유는, 모형의 단순성과 이론적 설명력에 있다. 즉, Bass 확산 모형은 신제품의 수요 확산 과정을 광고 등에 의해 충동되는 구매자의 내적 충동과 이미 구입한 구매자의 구전 효과에 의한 외적 충동의 산물로 해석하고 이를 수리적으로 정교하게 모형화 하였다. Bass 확산 모형에서 구매자의 내적 충동은 혁신 계수(Coefficient of innovation, p) 그리고 외적 충동은 모방 계수(Coefficient of imitation, q)로 모형화 되었고, 여기에 잠재 수요(potential adopters, m)가 더해져서 세 개의 파라미터가 Bass 모형을 구성하고 있다. 따라서 Bass 모형은 기본적으로 세 개의 수요 데이터가 있으면 모형의 파라미터 추정이 가능하다. 이러한 모형의 단

순성과 그 바탕에 놓여 있는 구매자 행동 논리의 정교함에 의해 Bass 모형은 신제품 수요 예측에 폭넓게 활용되고 있는 것이다.[4,11]

그러나, 다른 한편으로 Bass 확산 모형은 다양한 문제점을 지니고 있다. 가장 빈번하게 지적되는 문제점은 신제품 수요에 결정적인 영향을 미치는 시장 변수인 가격과 광고 등이 모형에서 제외되었다는 것이다.[1,3,5] 따라서 이들 변수를 포함한 다양한 모형이 제시되었으며, Norton과 Bass에 의해 이들 변수가 포함된 일반화된 Bass 모형(GBM, Generalized Bass Model)이 제시되었다. 또 다른 중요한 문제점은, Bass 모형이 실제 데이터에 맞지 않는 사례가 작지 않으며, 특히 데이터가 적을 경우 모형 추정의 불안정성(instability)이 매우 크다는 것이다.[2,6] 특히, Bass 모형의 계수 추정치가 신뢰성을 획득하기 위해서는 수요 최대치가 넘는 기간까지의 데이터가 요구된다는 연구가 제시되었다.[8] 이는 Bass 모형의 신제품 수요 예측에서의 활용 가치를 상당 부분 잠식하는 것이다. [10]은 실제 데이터와 시뮬레이션을 통해 Bass 모형의 계수 추정상의 문제점을 가장 체계적으로 연구한 논문이다. [10]에 의하면 계수 추정치들이 편향되어 있으며 추정에 활용되는 데이터 개수에 따라 일정한 경향을 갖고 변한다는 것이다. 그런데 [10]에서 시뮬레이션은 혁신계수와 모방 계수를 상수로 주고 랜덤 오차를 주는 방법으로 수행되었다. 따라서, 데이터를 통해 추정되는 핵심적인 두 계수인 p와 q의 추정치의 변동이 추정의 신뢰성에 미치는 영향은 제대로 연구되지 못하였다.

따라서 본 논문에서는 시뮬레이션을 실시하는 데에 있어 혁신 계수 p와 모방 계수 q의 변동을 주고, 이에 따라 생성되는 데이터의 개수를 변화 시키면서 수요 추정의 신뢰성을 분석하고자 한다. 논문은 다음과 같이 구성된다. 2절에서는 Bass 모형의 개요와 모수 추정의 방법론이 기술되고 3절에서는 시뮬레이션 방법이 소개된다. 4절에서는 시뮬레이션 결과를 토대로 Bass 모형 계수의 추정치에 대한 민감도 분석이 실시된다. 마지막으로 5절에서 결론 및 토의가 제시된다.

## 2. 연구배경

### 2.1 Bass 확산모형

Bass 모형은 시간을 통해 상품이나 서비스가 전파되는 수준을 나타내는 것으로 채택자(adopter) 수의 연속적인 증가 형태를 표현하는 방법이다. Bass 모형의 기본 가정 및 기본 개념은 다음과 같다.

$f(t)$  : 채택 확률 밀도함수 (density function)

$F(t)$  : t 까지 채택자의 누적 확률 함수  
(cumulative fraction of adopters)

m : 궁극적 채택자의 잠재적 수  
(potential number of ultimate adopters)

$n(t)$  : t의 신규 채택자 수.  $n(t) = mf(t)$

$N(t)$  : t 까지의 누적 채택자 수.  $N(t) = mF(t)$

p : 혁신 계수 (innovation coefficient).

이전의 채택과는 독립적인 영향을 표현

q : 모방 계수 (imitation coefficient).

구전이나 학습에 의한 영향을 표현.

Bass 모형에서 새로운 가입자는 혁신자와 모방자라는 두 집단으로 이루어진다고 가정된다. 혁신자(innovator)들의 채택은 다른 채택자들과는 무관하게 대중매체(외부영향)에만 의존한다. 혁신자들의 채택 비율은 시간이 지나면 감소한다. 반면, 모방자(imitator)들의 채택은 다른 채택자들의 영향을 받으며, 전파과정은 구전(word-of-mouth)이다.

이와 같은 가정을 가지는 Bass 모형의 식은 위험함수(hazard function)의 정의를 통해서 얻을 수 있다. 위험함수는 가입위험률(subscription hazard rate)이 선형함수로 표현된다고 가정한다. 가정에 따라 정의된 가입위험률은 다음과 같다.

$$\frac{f(t)}{1-F(t)} = p + qF(t) \quad \text{단, } F(0) = 0$$

누적 채택자의 증가율로 Bass 모형을 표현하면 다음과 같다.

$$n(t) = \frac{dN(t)}{dt} = p[m - N(t)] + \frac{q}{m} N(t)[m - N(t)]$$

단,  $n(0) = pm$

이때  $p[m - N(t)]$ 는 혁신자(innovator)의 채택을 의미하고,  $\frac{q}{m} N(t)[m - N(t)]$ 는 모방자(imitator)의 채택을 의미한다.

위 일계미분방정식을 풀어 누적채택자의 함수를 구하면 다음과 같다.[1]

$$N(t) = m \frac{1 - e^{-(p+q)t}}{1 + \frac{q}{p} e^{-(p+q)t}} \quad F(t) = \frac{1 - e^{-(p+q)t}}{1 + \frac{q}{p} e^{-(p+q)t}}$$

$$n(t) = m \frac{p(p+q)^2 e^{-(p+q)t}}{(p+q e^{-(p+q)t})^2} \quad f(t) = \frac{p(p+q)^2 e^{-(p+q)t}}{(p+q e^{-(p+q)t})^2}$$

따라서 당기 가입자가 최대가 되는 시점  $t^*$ 와 채택률(the rate of adoption)이 최대가 되는 시점  $t^{**}$ 는 다음과 같다.[1]

$$t^* = -\frac{1}{p+q} \ln\left(\frac{p}{q}\right)$$

$$t^{**} = -\frac{1}{p+q} \ln\left[(2 + \sqrt{3}) \frac{p}{q}\right]$$

## 2.2 계수 추정 방법론

Bass 모형의 계수를 추정하는 방법은 크게 다음 세 가지가 있다.

### 2.2.1 OLS 방법론[7]

연속 모형에서 당기 가입자 수는 다음과 같이 표현된다.

$$n(t) = pm + (q-p)N(t) - \frac{q}{p}N(t)^2$$

단, X는 당기 가입자, N은 누적가입자 수이다. 이를 OLS 추정을 위한 이산모형으로 바꾸어 표현한다.

$$n_t = pm + (q-p)N_{t-1} - \frac{q}{p}N_{t-1}^2 = a + bN_{t-1} + cN_{t-1}^2$$

위의 이산모형을 이용하여 OLS 추정방법으로 a, b, c를 구한 후 p, q, m을 얻는다.

### 2.2.2 NLS 방법론[9]

당기 가입자는 해당 기간 동안의 누적 가입비율의 차로 구할 수 있다.

$$n_t = m(F(t) - F(t-1)) + \epsilon_t$$

이를 기댓값을 취하여 다음과 같이 나타낼 수 있다.

$$E[n_t] = m(F(t) - F(t-1))$$

이 두 값의 차이를 최소화 하는 비선형 최소제곱법(Nonlinear MSE)을 다음과 같이 설정하여 모수를 추정한다.

$$\text{Min}_{p,q,m} \sum \{E[n_t] - n_t\}^2$$

### 2.2.3 MLE 방법론[8]

N(t)를 t기 까지의 누적 가입자, c를 최종적인 가입 확률,  $x_i$ 를 구간  $(t_{i-1}, t_i)$ 동안의 신규 가입자 수, 그리고 전체 인구 중에서 임의로 선택된 한 개인의 누적 가입 확률을  $F(t) = \frac{c(1 - e^{-bt})}{1 + ae^{-bt}}$ 로 둘 수

있다. (단,  $a=q/p$ ,  $b=p+q$ )

이때 우도 함수는

$$L(a,b,c,x_i) = [1 - F(t_{T-1})]^{x_T} \prod_{i=1}^{T-1} [F(t_i) - F(t_{i-1})]^{x_i}$$

로 정의된다. 따라서 우도함수를 최대화 하는 a, b, c, d를 구하여 그로부터 p, q, m을 구할 수 있다.

## 3. 시뮬레이션 방법

### 3.1 시뮬레이션 설계

단계 1. [3]에서 제시된 다양한 상품군의 p, q값을 토대로 p, q의 범위를 결정한다. 관심있는 p, q값을 선택한다.

단계 2. 선택된 p, q로부터 시뮬레이션 데이터를 생성한다. 주어진 누적 가입자 수를 바탕으로 다음 기간의 누적 가입자 수를 구하고자 할 때,  $p^* = p(1+\epsilon)$   $\epsilon \sim n(0, 0.1^2)$ 의 형태를 이용하여 p, q에 변동을 주게 된다. 이렇게 변동이 가미된 p, q를 이용하여 이전 기간의 누적 가입자 수로부터 해당 기간의 누적 가입자를 구하게 된다. 이 과정을 매 기간마다 거쳐 수요 그래프가 완전히 포화 시점에 이를 때까지 충분한 수의 랜덤 데이터를 얻게 된다.

단계 3. 단계 2의 방법으로 얻게 된 수요 그래프에서, 모수 추정을 위해 사용할 n개의 데이터를 선택한다.

단계 4. 단계 3에서 선택된 수요 데이터를 이용하여 OLS방법으로 모수 p, q를 추정한다.

단계 5. 단계 4에서 추정된 모수들을 통해 새로운 수요 데이터를 생성한다. 이 과정에서는 p, q에 변동을 주지 않는다. 이 값들을  $n_i$ 라 하자.

단계 6. 비교를 위하여 단계 1에서 얻은 p, q를 이용하여 역시 수요 데이터를 생성한다. 실제 데이터와는 다른 값이며, 주어진 모수를 Bass 모델에 적용하여 얻은 것이다. 이 값들을  $\hat{n}_i$ 라 하자.

단계 7. 단계 5와 단계 6에서 얻은 두 수요 데이터를 통해서 MSE 값을 구한다.

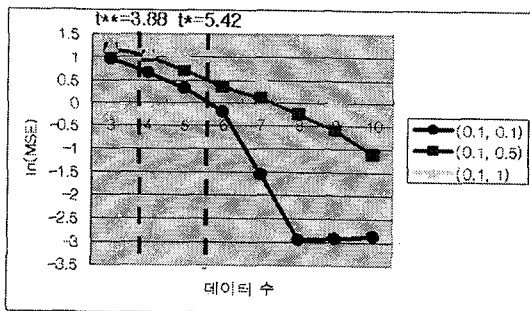
$$MSE = \frac{\sum_i^n (n_i - \hat{n}_i)^2}{m^2}$$

MSE를 구한 후, 시장 규모에 따른 스케일 문제를 없애기 위하여 ln(MSE)를 이용하기로 한다. ln(MSE)의 값이 -2가 나오면 오차는 m의 30%임을 의미한다.

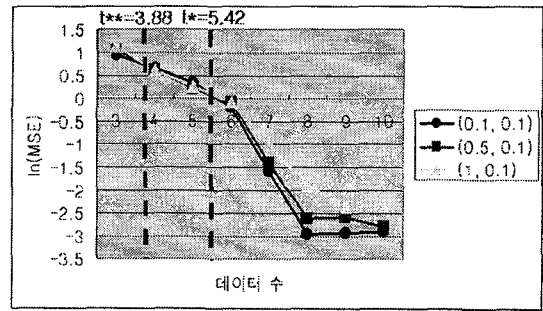
### 3.2 적용대상

본 논문에서는 국내 이동통신 데이터를 1985년~2005년 9월까지 21년치를 수집하여 OLS방식을 이용하여 추정하였다. 추정 결과값은  $p=0.00826$ ,  $q=0.84602$ ,  $m=4000$ 만명의 결과를 얻었다.

### 4. 민감도 분석



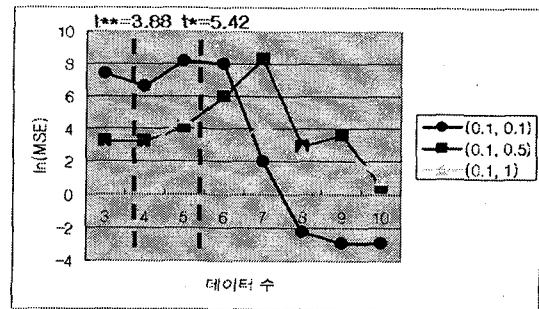
[그림 1] p변동=0.1일 때 q변동의 MSE 변화



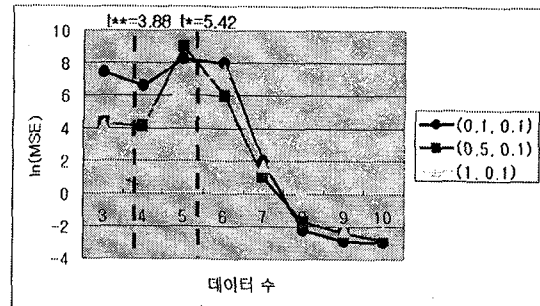
[그림 2] q변동=0.1로 고정 후 p변동의 MSE변화

[그림 1], [그림 2]는 국내 이동통신의 결과이고 m값의 경우는 고정하였으며, [그림 1]과 같은 추세가 p변동을 0.5와 1로 두고 q를 변화시킨 실험과 비슷하였으며, [그림 2]와 같이 q변동을 0.5와 1로 두고 p를 변화시킨 실험도 비슷한 결과를 나왔다. q변동이 p변동보다 MSE에 미치는 영향이 훨씬 크다는 결과를 볼 수 있다.

m값의 경우 p, q, m을 함께 추정할 경우는 매우 불안정하기 때문에, 보통 전문가의 예측을 사용한다. 이 논문에서는 전문가의 예측을 사용한 값과 p, q, m을 추정한 값을 비교하였다. [그림 1], [그림 2]의 경우는 p, q값은 추정하고 m값을 고정하였으며, [그림 3], [그림 4]의 경우는 p, q, m값을 추정하였다.



[그림 3] p변동=0.1일 때 q변동의 MSE 변화



[그림 4] q변동=0.1로 고정 후 p변동의 MSE변화

m값을 추정한 경우도 p변동 보다는 q변동이 MSE의 값에 영향을 크게 미치고 있으며, m값을 고정한 경우와 비교하였을때, MSE의 값이 더 큰

것으로 나타났다. 또한 p, q 변동이 각각 0.5, 1값을 취하며 변할 때에도 [그림 3], [그림 4]와 같은 추세를 보였다.

<표1>은 p, q의 다양한 값에 따라 ln(MSE)의 값의 변화를 보여준다. 이 경우, p, q 절대치의 영향을 보기 위해 p변동, q변동은 0.1로 고정하였다.

<표 1> p, q값 변화에 따른 ln(MSE) 값

p \ q	0.2	0.3	0.4
0.01	-1.60	-0.43	-0.05
0.02	-1.25	-0.85	-0.70
0.03	-1.27	-0.72	-1.05
0.04	-0.89	-0.78	-0.94
0.05	-0.80	-1.22	-0.88

p \ q	0.5	0.6	0.7
0.01	-0.50	0.20	0.31
0.02	-0.55	-0.23	0.02
0.03	-0.77	-0.38	-0.40
0.04	-0.72	-0.77	-0.40
0.05	-1.10	-0.63	-0.73

<표 2> p, q값 변화에 따른 t\* 값

p \ q	0.2	0.3	0.4
0.01	14.27	10.97	9.00
0.02	10.47	8.46	7.13
0.03	8.25	6.98	6.02
0.04	6.71	5.93	5.23
0.05	5.55	5.12	4.62

p \ q	0.5	0.6	0.7
0.01	7.67	6.71	5.98
0.02	6.19	5.49	4.94
0.03	5.31	4.76	4.31
0.04	4.68	4.23	3.87
0.05	4.19	3.82	3.52

p, q값에 따라 t\* 값이 정해지고 오차가 30%이 내가 되기 위해서는 t\* + 1 까지의 데이터 개수가 필요하며, t\*가 4이하일 경우, 추가적인 시뮬레이션 결과, 안정적인 계수 추정을 위해서는 최소한 5개까지의 데이터가 필요하다는 결론이 도출되었다.

## 5. 결론 및 추후연구

일반적으로 상품 또는 서비스의 수요 예측은 변화 가능성이 큰 시장에서 중요한 기초자료로 사용된다. 특히, 신제품이나 신규 서비스에서 수요 예측의 경우 데이터가 부족하거나 없는 경우는 결국 수요 예측의 신뢰도를 하락시키게 된다.

따라서 본 논문에서는 Bass 확산 모형을 사용하여 신뢰성 있는 수요 예측을 위한 적절한 데이터 개수를 알아보고자 시뮬레이션을 통한 민감도 분석을 하였다. 파라미터 q변동이 p변동보다 오차에 더 중요한 요소이며, q변동이 커질수록 데이터의 개수는 더 많이 필요하다는 결론을 얻었다. 오차가 잠

재수요의 30%이하로 떨어지기 위해서는 수요가 최대도 도달하는 시점이 t\*일 경우, t\* + 1 까지 데이터가 요구된다.

하지만 연구의 한계 또한 존재하는데 일반적인 데이터로 p변동, q변동의 값을 추정하는 방법론 연구가 추후에 필요할 것이며, 변동을 정규분포의 표준편차로 가정한 것에 대해서는 적합한지에 대한 검증의 필요성이 요구된다.

## 참고문헌

- [1] Bass F.M, T.V. Krishnan, D.C. Jain, "Why the Bass model fits without decision variables", *Marketing Science*, 13(3), 203-223, 1994.
- [2] Bernhardt, Irwin, K.M Mackenzie, "Some Problems in using diffusion models for new products", *Management science*, 19, 197-200, 1972.
- [3] Chatterjee, Rabikar, J. Fliashberg, "The innovation diffusion process in a heterogeneous population: a micro modeling approach", *Management science*, 36, 1057-1079, 1990.
- [4] Easingwood, Christopher, V. Mahajan, E. Muller, "A non-uniform influence innovation diffusion model of new product acceptance", *Marketing Science*, 2(3), 273-295, 1983.
- [5] Golder, Peter N. G.J. Tellis, "Beyond diffusion: an affordability model of the growth of new consumer durables", *Journal of Forecasting*, 17(3/4), 259-280, 1998.
- [6] Heeler R, T. Hustad, "Problems in predicting new product growth for consumer durables", *Management science*, 26(10), 1007-1020, 1980.
- [7] Putsis, William P Jr, V. Srinivasan, "Estimation techniques for macro diffusion models", in V. Mahajan, E. Muller, Yoram Wind(des). *New product diffusion models*, Kluwer Academic Publisher, Boston.
- [8] D. Schmittlein, V. Mahajan, "Maximum likelihood estimation for an innovation diffusion model of new product acceptance", *Marketing Science*, 1(1), 57-58, 1982.
- [9] Srinivasan V, C. Mason, "Nonlinear least squares estimation of new product diffusion models", *Marketing Science*, 5(2), 169-178, 1986.
- [10] V.D. Bulte, Christophe, G. Lilien, "Bias and systematic in the parameter estimates of macro-level diffusion models", *Marketing Science*, 16(4), 338-353, 1997.
- [11] D. Chandrasekaran, G.J. Tellis, "Diffusion and Growth of New Products: A Critical Review of Models and Findings", *Diffusion Review*, Nov 30, 2004.