

# NLS와 OLS의 하이브리드 방법에 의한 Bass 확산모형의 모수추정

홍정식<sup>1</sup> · 김태구<sup>2</sup> · 구훈영<sup>3\*</sup>

<sup>1</sup>서울과학기술대학교 IT정책전문대학원 산업정보시스템전공 /

<sup>2</sup>서울대학교 산업공학과 / <sup>3</sup>충남대학교 경영학과

## A Parameter Estimation of Bass Diffusion Model by the Hybrid of NLS and OLS

Jungsik Hong<sup>1</sup> · Taegu Kim<sup>2</sup> · Hoonyoung Koo<sup>3</sup>

<sup>1</sup>Department of Industrial and Information Systems Engineering, Seoul National University of Science and Technology

<sup>2</sup>Department of Industrial Engineering, Seoul National University

<sup>3</sup>Department of Business Administration, Chungnam National University

The Bass model is a cornerstone in diffusion theory which is used for forecasting demand of durables or new services. Three well-known estimation methods for parameters of the Bass model are Ordinary Least Square (OLS), Maximum Likelihood Estimator (MLE), Nonlinear Least Square (NLS). In this paper, a hybrid method incorporating OLS and NLS is presented and its performance is analyzed and compared with OLS and NLS by using simulation data and empirical data. The results show that NLS has the best performance in terms of accuracy and our hybrid method has the best performance in terms of stability. Specifically, hybrid method has better performance with less data. This result means much in practical aspect because the available data is little when a diffusion model is used for forecasting demand of a new product.

**Keywords:** Bass model, OLS, NLS, MLE, Hybrid of NLS and OLS

### 1. 서론

Bass(1969)은 신제품의 확산 과정을 설명하는 논문의 준거점 역할을 하는 논문이라 할 수 있다(Chandraekaran, 2004; Mahajan, 1990; Ozkaya, 2008). 제품의 확산 과정에 미치는 요소를 외부적 영향 요소와 내부적 영향 요소로 나누고 신제품 채택 확률의 위험률을 기존 채택자 수에 선형 비례하는 관계식으로 표현하는 Bass 확산 모형은 절약적이면서(parsimonious) 설명력이 뛰어난 모형이라고 할 수 있다(Chandraekaran, 2004). Bass 모형을 기반으로 한 연구는 거의 750편에 이르고 있다(Ozkaya, 2008). Bass 모형을 기반으로 한 이러한 연구들은 크게 네 가지

로 나누어진다. 첫째는 모형의 확장에 관한 연구이다. 이는 다시 두 가지로 나누어 살펴볼 수 있다. 하나는 Bass 모형에 한 두 개의 파라미터를 추가하여 모형의 유연성을 제고하는 연구들이다(Easingwood, 1983; Park, 2005; Hong, 2009). 다른 하나는 소득이나 가격 등의 변수를 추가하여 모형의 설명력을 높이는 연구들이다(Horsky, 1990; Kalish, 1985; Robinson, 1975). 둘째, 모형의 구조를 확장한 연구이다. 주로 2000년 이후에 이루어지는 사회적 관계망을 접목시키거나(Rahmandad, 2007; Goldenberg, 2006; Goldenberg, 2009; Guseo, 2009; Ire, 2006), 채택자를 두 그룹으로 나누어 이분화된 시장(dual market)을 기반으로 확산 모형을 구축한 연구가 이에 해당한다고 하겠다(Vakratsas,

본 연구는 지식경제부 및 정보통신연구진흥원의 IT 산업원천 기술개발 사업의 일환으로 수행하였음(10033098).

\* 연락처: 구훈영 교수, 305-764 대전시광역시 유성구 대학로 99, 충남대학교 경영학과, Tel : 042-821-5531, Fax : 042-821-8718,

E-mail : eloh99@gmail.com

2010년 12월 18일 접수; 2011년 1월 13일 수정본 접수; 2011년 2월 7일 게재 확정.

2008; Goldenberg, 2006; Van den Bulte, 2007). 셋째, 모형의 모수추정방법에 관한 연구이다(Boswijk, 2002; Venkatesan, 2004; Satoh, 2001; Schmittlein, 1982; Srinivasan, 1986; Ko, 2008). 넷째, 실제의 한두 가지 혁신의 확산 사례를 대상으로 Bass 모형을 적용한 경험적 연구들이다(Bayus, 1987; Jun, 1999; Mahajan, 1990; Sawhney, 1996; Hong, 2008). 이러한 연구 중 90년 이후로는 충분한 연구가 이루어지지 않은 분야가 모수추정 방법이다. 현재 논문에서 주로 언급되고 사용되는 모수 추정방법은 1986년에 나온 Srinivasan과 Mason의 Nonlinear Least Square (NLS) 방법이다(Srinivasan, 1986). 그러나 이 방법은 삼차원 탐색을 시도해야 하는 번거로움이 있고, 초기치에 크게 의존하는 단점을 지니고 있다. 또한 추정에 사용되는 데이터 수가 작을수록 추정의 편의(bias)가 생기고 안정성이 급격히 떨어지는 한계를 가지고 있다. 신뢰성 성장분야 연구자인 Satoh(2000)에 의해 새로운 Ordinary Least Square(OLS) 모수 추정방법이 나왔지만 시뮬레이션에 따른 성능분석이 미흡하여 모수추정관련 연구에서 제대로 언급이 되지 않고 있다. 최근에 NLS 방법의 초기치 문제를 해결하기 위한 시도가 두 가지 이루어졌다(Venkatesan, 2004; Ko, 2008). Venkatesan(2004)는 최적해를 효율적으로 찾는 Genetic Algorithm(GA)을 NLS 방법의 삼차원 탐색에 적용한 연구이고, 또 Ko(2008)는 그리드 탐색기법(grid search)을 부분적으로 활용한 연구이다. 전자는 NLS 방법의 모수추정의 효율성을 높여주지만, GA 소프트웨어를 사용하여야 하고, 휴리스틱 프로그램인 만큼 효율성 또한 사례에 따라 다르게 나타나고, 또한 실무자들이 사용하기에 복잡한 단점을 지니고 있다. Ko(2008)의 경우, Bass 모형의 경우 로지스틱과의 근사 비교에 의존하고, 낮은 계산 효율로 근사해를 얻는 그리드 탐색기법을 사용하는 한계를 지니고 있다. 본 논문은 Bass 모형의 위험률 표현법을 활용하여 기존 방법과는 다른 새로운 모수 추정 방법을 다룬다. 이 방법에는 OLS 방법과 NLS 방법이 모두 사용되므로 우리는 하이브리드 방법이라는 명칭으로 본 논문의 아이디어를 소개한다. 본 논문이 제시한 하이브리드 방법의 성능 분석을 위해 시뮬레이션을 실시하여 모수추정의 안정성(stability)과 정확도(accuracy)를 기존의 방법과 비교하고, 실제 데이터 분석을 통하여 하이브리드 방법의 우수성을 제시한다. 또한 확장된 Bass 모형의 모수추정의 경우에도(hong, 2009), 하이브리드 방법이 적용될 수 있음을 간략하게 제시한다.

본 논문은 다음과 같이 구성되어 있다. 제 2장에서는 기존의 모수추정 방법이 간략히 소개된다. 제 3장에서는 Bass 모형의 위험률 표현과 이를 활용한 모수추정방법이 기술된다. 제 4장에서는 시뮬레이션과 실제 사례 분석이 이루어지고 마지막으로 제 5장에서 결론이 제시된다.

## 2. Bass 확산 모형의 모수 추정방법

Bass 확산 모형은 신제품의 채택률이 두 가지 요소에 의해 영

향을 받는다고 가정한다. 하나는 매스 미디어나 잡지 등에 의한 외부적 영향이고 다른 하나는 기존 채택자의 구전효과에 의한 내부적 영향이다. 그리고 이 두 요소의 영향은 선형적 결합으로 작용한다고 가정한다. 잠재 수요를  $m$ 이라고 가정할 때, 이는 다음과 같은 수식으로 표현된다(Bass, 1969).

$$\frac{dN(t)}{dt} = (m - N(t)) \cdot \left( p + q \cdot \frac{N(t)}{m} \right)$$

여기서  $p$ 는 외부 영향을 나타내는 요소이고  $q$ 는 내부 영향을 나타내는 요소이며,  $N(t)$ 는  $t$ 시점까지의 누적채택자 수이다. 따라서 Bass 확산 모형에서 추정해야 할 모수의 개수는 세 개다. 이러한 모수의 추정방법으로는 크게 세 가지가 있다(Chandraekaran, 2004).

첫째 방법은 회귀분석의 모수추정 방법을 활용한 추정방법이다. 이는 OLS 방법으로 불린다(Bass, 1969). 둘째 방법은 우도추정함수를 활용한 MLE 방법이다(Schmittlein, 1982). 세 번째 방법은 추정치와 실측치 차이의 제곱합을 최소화하는 방법으로, 추정치에 비선형 함수식이 사용되는 것을 고려하여 NLS 방법으로 불린다(Srinivasan, 1986). 이들 방법 중에서 NLS 방법이 가장 우수하다고 알려져 있다(Srinivasan, 1986). 그런데, 2000년 이후 몇 개의 모수 추정에 관한 연구가 이루어졌다(Venkatesan, 2004, Satoh, 2001, Ko, 2008). 이들 연구 중, Venkatesan(2004)는 기존 OLS 방법을 개선한 수식을 제시하여 모수 추정방법을 개선시킨 것(Satoh, 2001)과 기존의 NLS 방법을 이용하되 최적해 탐색의 효율성을 향상시킨 것으로 나누어진다(Venkatesan, 2004; Hong, 2009). 결과적으로, 모수 추정방법의 관점에서 보면, 2000년 이전에 제시된 방법 중 가장 성능이 우수한 방법인 NLS 방법과 2000년 이후 개발된 새로운 OLS 방법 두 가지가 가장 주요한 모수 추정방법이라 할 수 있다. 먼저 새로운 OLS 방법을 살펴보자. 그런데, 새로운 OLS 방법을 설명하기 위해, 논의의 자족성을 위해 기존의 OLS 방법부터 기술하기로 한다.

### 2.1 OLS 방법

Ordinary Least Square 추정법은 선형 방정식의 모수를 추정하는데 쓰인다. OLS 추정을 위한 선형 방정식의 형태는 다음과 같다(Bass, 1969).

$$y = bX + \epsilon \quad (1)$$

식 (1)에서  $X$ 는  $y$ 절편을 의미하는 1값의 열과  $n$ 개의 독립변수  $x$ 의 시계열 값을 포함하는  $n+1$ 열 크기를 가지는 행렬이다.  $y$ 절편의 값과 각 독립변수의 계수가 포함된  $b$ 는  $n+1$ 의 행 크기를 가지는 벡터가 된다. 이때 모수  $b$ 는 다음과 같은 행렬 연산을 통해 추정할 수 있다.

$$\hat{b} = (X'X)^{-1}X'y \quad (2)$$

### 2.1.1 기본적인 OLS 추정방법

Bass 모형을 OLS 방법으로 추정하는 전통적인 방식은 다음과 같다. 당기 수요  $n(t)$ 는 다음과 같은 형태로 누적 수요  $N(t)$ 의 선형 식으로 나타낼 수 있다.

$$n(t) = \frac{dN(t)}{dt} = p[m - N(t)] + \frac{q}{m}N(t)[m - N(t)] \quad (3)$$

이를 OLS 추정을 위해 이산화(discretize)하면 다음과 같은 형태를 얻는다.

$$\begin{aligned} X(i) &= pm + (q-p)N(t_{i-1}) - (q/m)N^2(t_{i-1}) + \epsilon \\ &= \alpha_1 + \alpha_2 N(t_{i-1}) + \alpha_3 N^2(t_{i-1}) + \epsilon \end{aligned} \quad (4)$$

위 식을 OLS 추정하면 식 (1)에 의해  $\hat{\alpha}_1$ ,  $\hat{\alpha}_2$  그리고  $\hat{\alpha}_3$  가 얻어지고, 이를 토대로 식 (4)에서  $\alpha_1$ ,  $\alpha_2$ ,  $\alpha_3$  와  $p$ ,  $q$ ,  $m$ 과의 관계식에 따라 Bass 모형의 모수들을 다음과 같이 얻을 수 있다.

$$\begin{aligned} \hat{p} &= \frac{\hat{\alpha}_1}{\hat{m}} \\ \hat{q} &= -\hat{m}\hat{\alpha}_3 \\ \hat{m} &= -\hat{\alpha}_2 - \frac{\sqrt{\alpha_2^2 - 4\hat{\alpha}_1\hat{\alpha}_3}}{2\hat{\alpha}_3} \end{aligned}$$

### 2.1.2 새로운 OLS 추정방법

Satoh(2000)는 기존 OLS 방법의 단점을 보완한 새로운 OLS 추정 방법을 제안하였다. 그는 Bass 모형이 Riccati Equation의 한 종류라는 점에 착안하여, Hirota(1979)가 제안한 Riccati Equation의 Discrete Form을 이용하여 Bass 모형의 Discrete Equation을 식 (5)와 같이 유도하였다.

$$\begin{aligned} \frac{N_{n+\delta} - N_{n-\delta}}{2\delta} &= pm + \frac{q-p}{2}(N_{n+\delta} + N_{n-\delta}) \\ &\quad - \frac{q}{m}N_{n+\delta}N_{n-\delta} \end{aligned} \quad (5)$$

시간 간격  $\delta$ 를 1로 두고 정리하면 다음과 같이 OLS 추정을 위한 식을 얻는다.

$$\begin{aligned} \frac{S_n}{2} &= a + b(N_{n+1} + N_{n-1}) + cN_{n+1}N_{n-1} + \epsilon_n \\ \text{단 } S_n &= N_{n+1} - N_{n-1} \end{aligned} \quad (6)$$

식 (6)에서, 모수  $a$ ,  $b$ ,  $c$ 를 OLS 방법으로 추정한 후에, 식 (5)와 식 (6)의 모수들 사이의 관계식으로부터 다음과 같이 Bass 모형의 모수들을 얻는다.

$$\begin{aligned} \hat{p} &= -\hat{b} + \sqrt{\hat{b}^2 - \hat{a}\hat{c}} \\ \hat{q} &= \hat{b} + \sqrt{\hat{b}^2 - \hat{a}\hat{c}} \\ \hat{m} &= \frac{-\hat{b} - \sqrt{\hat{b}^2 - \hat{a}\hat{c}}}{\hat{c}} \end{aligned} \quad (7)$$

하지만 Satoh(2001)는 이렇게 추정된 확산 계수  $p$ ,  $q$ 에 대해서, Bass 모형의 이산화에 따른 오차를 줄이기 위해 논문에서 다음과 같은 보정을 추가로 제시한다.

$$\begin{aligned} p &= \tilde{k}\hat{p} \\ q &= \tilde{k}\hat{q} \\ \tilde{k} &= -\frac{1}{2(\hat{p} + \hat{q})} \ln \frac{1 - (\hat{p} + \hat{q})}{1 + (\hat{p} + \hat{q})} \end{aligned} \quad (8)$$

## 2.2 NLS 추정방법

NLS 방법은 추정된 모형과 실제 데이터의 오차 제곱합을 최소로 하는 다차원 비선형 모수 추정법이다. 다음과 같은 오차 제곱합을 목적함수로 두고 목적함수를 최소로 하기 위해 3차원 모수 탐색을 수행하게 된다(Srinivasan, 1986).

$$\min_{p, q, m_i} \sum_{i=0}^T [\hat{n}_i - n_i]^2 \quad (9)$$

여기서,  $n_i$ 는  $[i-1, i]$ 기간의 당기수요이고  $\hat{n}_i$ 는 예측수요이다. 예측수요는 Bass 모형으로부터 다음과 같이 구해진다.

$$\begin{aligned} \hat{n}_i &= m[F(i) - F(i-1)] \\ F(i) &= \frac{1 - e^{-(p+q)i}}{1 + \frac{q}{p}e^{-(p+q)i}} \end{aligned} \quad (10)$$

## 3. Bass 확산모형의 하이브리드 모수추정

### 3.1 Bass 확산 모형의 위험률 표현

식 (1)에서  $[m - N(t)]$ 로 양변을 나누면 다음과 같다.

$$\frac{dN(t)}{m - N(t)} = p + q \cdot \frac{N(t)}{m} \quad (11)$$

식 (11)의 좌변을 분모, 분자 각각  $m$ 으로 나누면,

$$\frac{d \frac{N(t)}{m}}{1 - \frac{N(t)}{m}} = p + q \cdot \frac{N(t)}{m},$$

이 되고,  $\frac{N(t)}{m} = F(t)$ 를 대입하면 다음의 식을 얻을 수 있다.

$$\frac{f(t)}{1-F(t)} = p + qF(t) \tag{12}$$

$\frac{f(t)}{1-F(t)}$ 를 위험률이라 하고,  $F(t)$ 는 침투율이라 불린다.

식 (12)에서 침투율을  $X_t$ 라 하고, 위험률을  $Y_t$ 라 하면 다음과 같은 선형 방정식을 얻을 수 있다.

$$Y_t = p + qX_t \tag{13}$$

따라서  $(X_t, Y_t)$  값이 주어질 경우,  $p, q$ 는 OLS 방법에 의해 추정이 가능하다.  $X_t = \frac{N(t)}{m}$  이고,  $Y_t$ 를 근사화시켜서 표현하면,  $\frac{[N(t)-N(t-1)]}{m-N(t)}$ , 즉,  $\frac{n(t)}{m-N(t)}$ 가 된다.

따라서  $m$ 이 주어질 경우 누적 수요  $N(t)$ 에 따라서 위험률과 침투율 값이 얻어진다. 따라서  $m$ 이 주어진 경우,  $p$ 와  $q$  값은 선형 회귀방정식의 모수추정 방법인 OLS를 활용하면 쉽게 구할 수 있다(Philips, 2007). 그런데 이렇게  $m$  값이 알려져 있다고 가정하고 침투율과 위험률을 토대로 그래프를 도시하고  $p$ 와  $q$ 를 OLS로 추정하는 과정은 논문에서 언급된 경우는 없고, F. Philip의 저서인 Philips(2001)에 간략히 나와 있다. 이 경우에,  $m$ 이 알려지지 않은 경우는 더 이상 다루고 있지 않다. 다음 절에서 우리는 이 과정을 다루고자 한다.

### 3.2 하이브리드 방법에 의한 모수추정

제 3.1절에서 기술한 바와 같이 Bass 모형의 모수  $m, p, q$  중  $p$ 와  $q$ 는  $m$ 의 값을 고정한 상태에서 식 (5)를 토대로 OLS 방법

으로 추정한다.  $m$  값의 추정은 NLS 방법에 사용된 식 (9)을 이용한다. NLS 방법에서는  $m, p$  그리고  $q$ 에 대해 삼차원 탐색을 실시하였다면, 이제는  $m$ 에 대해서 일차원 탐색을 실시하게 된다. 즉,  $m$ 의 하한과 상한을  $m_L$ 과  $m_U$ 로 설정한 경우  $m_L$ 부터 시작해서 일정하게  $m$  값을 증가시키면서 식 (9)의 최소값 탐색을 실시한다. 식 (9)에서  $m$  값이 주어지면, 나머지 두 개의 모수  $p$ 와  $q$ 는 식 (13)에서 OLS로 추정한 값이 사용된다. 이러한 하이브리드 과정을 설명하면 <Figure 1>과 같다. 여기서  $m_L$ 은 누적 수요가 잠재 수요보다 작아야 한다는 Bass 모형의 수리적 성격에 따라 통상적으로 누적 수요의 110%으로 하고,  $m_U$ 는 제품에 따라 전체 인구나 전체 가구수가 된다.

구체적인 하이브리드 모수 추정 절차 알고리즘은 다음과 같다.

- 1) 잠재수요  $m$ 의 하한  $m_L$ 과 상한  $m_U$ 를 설정하고 일정한 간격으로 나눈다.
- 2) 구간으로 나뉜  $m$  값 중  $k$ 번째 값인  $m_k$ 를 선택한다.
- 3)  $m_k$ 에 대하여  $Y_t = \frac{[N(t)-N(t-1)]}{m_k-N(t)}$ ,  $X_t = \frac{n(t)}{m_k-N(t)}$ 의 형태로 식 (13)에 적용한다.
- 4) OLS를 통해  $p_k, q_k$ 를 얻는다.
- 5) 구한  $m_k, p_k, q_k$ 에 대한 오차제곱합  $e_k$ 를 구한다.
- 6) 2)~5)를 반복하며 모든  $m_k$ 에 대하여  $m_k, p_k, q_k, e_k$ 를 구한다.
- 7) 구한  $m_k, p_k, q_k, e_k$  중 가장 작은 오차제곱합  $e_k$ 를 갖는  $m_k, p_k, q_k$ 를 모수 추정치  $m^*, p^*, q^*, e^*$ 로 사용한다.

### 3.3 확장된 Bass 모형에 대한 하이브리드 방법의 적용

NLS와 OLS를 결합하여 Bass 모형의 모수를 추정하는 하이

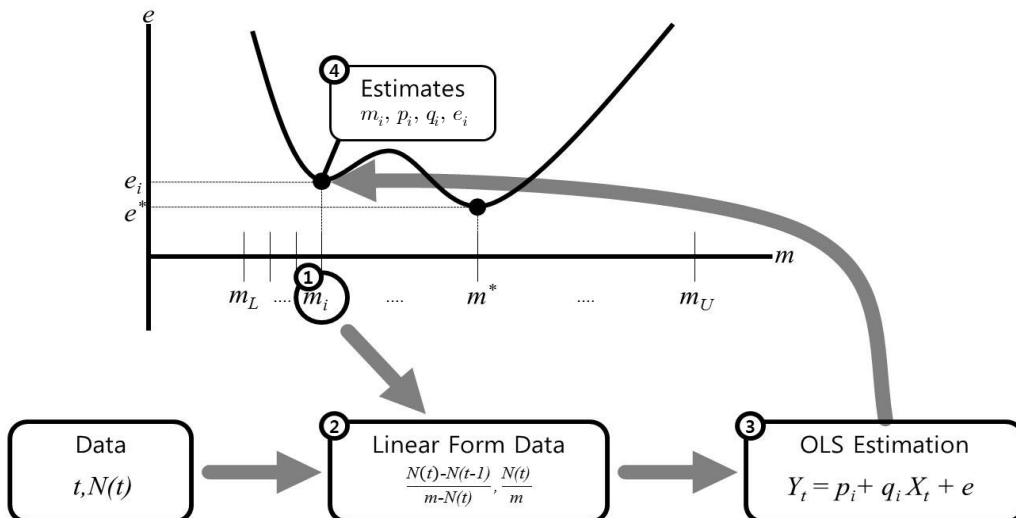


Figure 1. An Estimation Procedure of The Hybrid Method

브리드 방법은 Bass 모형을 확장한 모형의 모수 추정에도 사용될 수 있다. 먼저 Bass 확장 모형 중 가장 널리 알려진 NUI (non-Uniform Influence) 모형을 고려하자(Easingwood, 1983). 이 모형의 위험률 표현식은 다음과 같다.

$$\frac{f(t)}{1-F(t)} = p + q \cdot F^\delta(t) \quad (14)$$

식 (14)을 누적 수요를 사용하여 다시 표현하면 다음과 같다.

$$\frac{N(t) - N(t-1)}{m - N(t)} = p + q \left[ \frac{N(t)}{m} \right]^\delta \quad (15)$$

식 (15)을 보면  $m$ 과  $\delta$  값이 주어질 경우, 누적 수요를 바탕으로  $p$ 와  $q$ 의 OLS 추정이 가능함을 알 수 있다. 따라서 NUI 모형에서 NLS 방법으로 모수 추정을 할 경우에는 4차원 탐색을 실시하게 되나, OLS와 결합한 하이브리드 방법을 활용하면 2차원 탐색으로 모든 모수의 추정이 가능하다.

다음으로 최근 제시된 Bass 모형의 이분 확장 모형(Hong, 2009)을 고려하자. 이 모형은 시점  $t_c$ 를 기준으로 Bass 모형의 외부 영향 계수  $p$ 와 내부 영향 계수  $q$ 가 서로 다른 값을 갖도록,  $p$ ,  $q$ 의 시간 불변 가정(time-invariant assumption)을 완화한 모형이다. 따라서 이 모형은 NLS 방법으로 모수를 추정할 때,  $m$ 을 Bass 모형을 토대로 먼저 추정된 후, 다음 수식을 통해,  $p_1$ ,  $p_2$ ,  $q_1$ ,  $q_2$  그리고  $t_c$ 를 추정한다.

$$\begin{aligned} & \text{Min} \sum_{i=1}^{t_c-1} (\widehat{n}_{k,1} - n_k)^2 + \sum_{k=t_c}^j (\widehat{n}_{k,2} - n_k)^2 \\ & \widehat{n}_{k,1} = m(F_1(k) - F_1(k-1)) \\ & \widehat{n}_{k,2} = m(F_2(k) - F_2(k-1)) \\ & F_1(t) = \frac{1 - e^{-(p_1+q_1)t}}{1 + \frac{q_1}{p_1} e^{-(p_1+q_1)t}} \\ & F_2(t) = \frac{1 - e^{-(p_2+q_2)t}}{1 + \frac{q_2}{p_2} e^{-(p_2+q_2)t}} \end{aligned} \quad (16)$$

이 모형의 장점은 모수 수가 늘어나지만,  $t_c$ 를 기점으로 각 구간에서 추정할 모수의 수는 같다는 것이다. 즉,  $(p_1, q_1)$ 과  $(p_2, q_2)$ 는 별 개의 데이터를 토대로 독립적으로 추정된다. 그런데, 이 경우에도  $m$ 과  $t_c$ 가 정해지면, 각 구간에서는 통상적인 Bass 모형이 되므로, 제 3.2절에서 논한 바와 같이  $(p_1, q_1)$ ,  $(p_2, q_2)$ 를 OLS 방법으로 추정할 수 있는 것이다. 이 경우  $t_c$  값에 따라 각 구간에서 추정에 사용되는 데이터 수가 적어지므로, 추정의 안정성이 매우 중요한 문제가 된다. 실제로 Van den Bulte(1997)에 의하면 일반적인 NLS 모수추정 방법은 추정에 사용되는 데이터 수가 적을 때, 안정성이 부족한 단점을 지니

고 있다. 이제 시뮬레이션과 실제 데이터 분석을 통해 하이브리드 모수 추정 방법의 성능을 살펴보자.

## 4. 기존 방법과의 비교 분석

제 3장에서 언급된 하이브리드 모수 추정 방법의 추정 성능을 분석하기 위해 제 2장에서 소개한 Bass 모형의 다양한 추정 방법들을 비교 대상으로 삼아 시뮬레이션과 실제 데이터를 이용하여 각각의 방법으로 추정하고 결과를 비교하고자 한다.

### 4.1 시뮬레이션에 의한 비교 분석

통계적으로 유의한 수준의 많은 추정 결과를 얻기 위해서는 상당히 많은 수요 시계열 데이터가 필요하다. 그러나 수백 종의 상품과 서비스로부터 각각의 수요 데이터를 얻는 것은 현실적으로 한계가 있다. 이를 극복하기 위하여 현실의 수요 데이터와 비슷한 성질을 갖는 시뮬레이션 데이터를 다수 생성하여 충분한 수의 추정 결과를 토대로 추정 방법들의 성능을 비교하고자 한다.

#### 4.1.1 시뮬레이션 데이터

현실의 수요 데이터들은 Bass 모형과 정확히 일치하지 않는다. 현실과 비슷한 수요 데이터를 만들기 위해서 Bass 모형을 이용하여 임의의 수요 데이터를 생성한 후, 그 데이터에 정규분포를 따르는 백색오류를 첨가하는 과정을 거쳐 시뮬레이션 데이터가 생성된다.

수요 데이터의 기본을 이루는 Bass 모형은  $m = 100$ ,  $p = 0.02$ ,  $q = 1$ 의 설정이며 평균이 0이고 표준편차( $\sigma$ )가 5%, 10%, 15%인 정규분포 오류를 첨가하여 실험한다. 사용하는 데이터의 수( $K$ )는 5개( $t = 3 \sim 7$ ), 7개( $t = 1 \sim 7$ ), 11개( $t = 1 \sim 11$ )로 설정한다.

#### 4.1.2 시뮬레이션 데이터 추정 결과

다음은 시뮬레이션으로 발생시킨 수요 데이터 중  $t = 1 \sim 11$ 의 11개 데이터를 사용하여 추정한 결과이다. 백색오류의 수준에 따라 5%, 10%, 15%, 20%, 30%의 5가지 실험 결과를 표로 정리하여 나타내었다. 실험 결과는 같은 조건으로 100번 발생시킨 시뮬레이션 데이터 각각에 대한 Bass 모형의 모수  $m$ ,  $p$ ,  $q$ 의 추정 결과의 평균값과 추정된 모수들의 표준편차, 그리고 추정된 모수를 통해 얻은 Bass 모형과 추정에 쓰인 시뮬레이션 데이터 간의 오차제곱합과 그 표준편차로 이루어져있다.

11개의 데이터를 사용하는 경우, OLS와 Satoh 모형의 추정 성능이 크게 다르지 않은 반면 NLS와 하이브리드 방법의 경우에는 상대적으로 뛰어난 추정 결과를 보여주었다. 오차제곱합에 있어서도 NLS와 하이브리드 방법은 다른 2가지 추정 방법들보다 월등히 나은 결과를 보여주었으며, 추정된 모수의 표준편차 값 역시 상대적으로 낮게 나타났다.

**Table 1.** Simulation Result( $\sigma = 5\%$ ,  $K = 11$ )

5%	$m$	$std(m)$	$p$	$std(p)$	$q$	$std(q)$	$ln(SE)$	$std(ln(SE))$
OLS	107.307	6.636	-0.0055	0.0078	0.8334	0.1288	12.0618	2.892
Satoh	100.440	3.954	0.0021	0.0048	0.9926	0.1082	7.8454	3.214
NLS	100.203	3.761	0.0021	0.0007	0.9977	0.0770	3.8680	0.873
Hybrid	101.972	5.160	0.0028	0.0010	0.9472	0.1210	4.2742	0.977

**Table 2.** Simulation Result( $\sigma = 10\%$ ,  $K = 11$ )

10%	$m$	$std(m)$	$p$	$std(p)$	$q$	$std(q)$	$ln(SE)$	$std(ln(SE))$
OLS	189.352	450.527	0.0076	0.0182	0.5762	0.2667	9.3082	3.329
Satoh	102.968	8.697	0.0029	0.0092	0.9523	0.2330	8.9616	2.971
NLS	101.621	7.556	0.0027	0.0019	0.9891	0.2047	5.3943	0.847
Hybrid	107.184	10.524	0.0039	0.0022	0.8745	0.1985	5.7585	0.907

**Table 3.** Simulation Result( $\sigma = 15\%$ ,  $K = 11$ )

15%	$m$	$std(m)$	$p$	$std(p)$	$q$	$std(q)$	$ln(SE)$	$std(ln(SE))$
OLS	141.112	180.937	0.0334	0.0667	0.3574	0.3431	9.0873	2.202
Satoh	105.664	13.665	-0.0014	0.0167	1.0077	0.3876	9.3997	2.988
NLS	105.017	14.218	0.0026	0.0025	1.0159	0.2614	6.1544	0.794
Hybrid	114.759	15.494	0.0050	0.0027	0.7988	0.2061	6.3844	0.822

**Table 4.** Simulation Result( $\sigma = 20\%$ ,  $K = 11$ )

20%	$m$	$std(m)$	$p$	$std(p)$	$q$	$std(q)$	$ln(SE)$	$std(ln(SE))$
OLS	188.592	915.278	0.1840	0.2741	0.1468	0.3212	9.4368	1.648
Satoh	108.150	35.271	-0.0066	0.0236	1.0365	0.4838	10.1196	3.143
NLS	105.514	35.948	0.0028	0.0036	1.1155	0.4223	6.7258	0.857
Hybrid	116.166	22.349	0.0064	0.0060	0.7761	0.2907	7.0413	0.923

**Table 5.** Simulation Result( $\sigma = 30\%$ ,  $K = 11$ )

5%	$m$	$std(m)$	$p$	$std(p)$	$q$	$std(q)$	$ln(SE)$	$std(ln(SE))$
OLS	65.747	136.603	0.3529	0.3531	0.0192	0.2293	10.4939	2.521
Satoh	117.644	87.524	-0.0112	0.0368	0.9892	0.6378	10.2742	2.399
NLS	113.985	51.097	0.0032	0.0040	1.1084	0.5364	7.4804	0.701
Hybrid	131.208	29.365	0.0069	0.0070	0.7301	0.3065	7.8294	0.727

하이브리드와 NLS의 비교에 있어서는 두 방법이 거의 같은 성능을 보여주었으나 근소한 차이로 NLS의 추정 성능이 더 나은 것으로 나타났다. 그러나 백색오류의 수준에 따라서는 추정된 모수  $m$ 이나  $q$ 의 표준편차가 하이브리드 방법이 더 낮게 나타나는 경우도 볼 수 있다.

데이터 수와 추정 성능의 관계를 살펴보기 위하여 백색오류 수준을 20%로 고정하고 사용하는 데이터의 수를 달리하여 실험한 결과를 다음의 표에 나타내었다. <Table 6>의 실험 결과

는  $t = 1\sim 7$ 의 시뮬레이션 데이터를 사용한 결과이고( $K = 7$ ), <Table 7>의 결과는  $t = 3\sim 7$ 의 시뮬레이션 데이터를 사용한 결과( $K = 5$ )이다.

데이터 수가 줄어들수록 가장 뚜렷한 변화를 보이는 것은 OLS 방법이다. 모수의 추정 평균은 물론 모수의 표준편차가 지나치게 높은 값을 나타내고 있어 추정 성능이 매우 나쁘다는 것을 알 수 있다.

Satoh 방법은 상대적으로 모수의 표준편차가 낮은 안정된

Table 6. Simulation Result( $\sigma = 20\%$ ,  $K = 7$ )

5%	$m$	$std(m)$	$p$	$std(p)$	$q$	$std(q)$	$ln(SE)$	$std(ln(SE))$
OLS	560.798	2111.97	-0.0067	0.0325	0.8386	0.3497	10.2777	2.427
Satoh	178.051	242.981	-0.0021	0.0105	1.1122	0.4104	8.2939	3.482
NLS	161.143	243.647	0.0014	0.0012	1.2189	0.5377	2.4156	1.452
Hybrid	98.010	53.259	0.0011	0.0045	1.1637	0.1887	3.7043	1.442

Table 7. Simulation Result( $\sigma = 20\%$ ,  $K = 5$ )

5%	$m$	$std(m)$	$p$	$std(p)$	$q$	$std(q)$	$ln(SE)$	$std(ln(SE))$
OLS	495.592	2259.88	0.0234	0.2127	0.8200	0.4947	9.7788	3.176
Satoh	143.058	177.716	0.0030	0.0335	1.1254	0.5016	8.9607	1.634
NLS	138.450	180.350	0.0013	0.0012	1.2914	0.4886	2.5635	1.545
Hybrid	87.461	31.289	0.0014	0.0008	1.2232	0.2064	3.8605	1.263

결과를 보여주고 있으나 평균적 오차제곱합을 고려할 때 추정 성능은 역시 좋지 않다고 볼 수 있다.

NLS 방법은 여전히 모든 방법들 중 가장 낮은 오차제곱합을 보여주고 있으나 추정된 모수의 표준편차가 하이브리드 방법보다 높게 나타나고 데이터 수가 적은 경우에는 Satoh 방법보다 높게 나타나는 등 데이터 수의 감소에 따라 여러 한계를 보이고 있다.

하이브리드 방법은 오차제곱합 자체는 NLS에 비해 다소 높지만 추정된 모수의 표준편차에 있어서 상당히 큰 차이로 가장 뛰어난 성능을 보여주고 있다. 이는 데이터가 부족한 상황에서도 상대적으로 안정적인 추정이 가능하다는 의미를 가진다. 이러한 추정치의 안정성은 실제문제에의 활용가능성 측면에서 가치가 크다고 할 수 있다. 그 이유는 실제 확산 모형을 사용하여 향후 수요를 예측하고자 할 때, 확산 모형이 주로 혁신적인 신제품의 확산 현상을 다루므로, 사용가능한 데이터 수가 많지 않은 경우가 대부분이기 때문이다.

## 4.2 실제 데이터를 이용한 비교 분석

다양한 종류의 실제 제품들의 수요 데이터를 통해 추정 방법들의 성능을 여러 가지 상황에서 비교해 보기 위하여 실제 데이터를 이용한 비교 분석을 실시하였다. MRI 장비, CT 스캐너, 초음파 영상기의 경우 2001년~2006년의 데이터는 국민건강보험공단 전산처리 기준을 참고하였으며, 2000년 이전 자료는 OECD Health Data 2006을 참고하였다. CATV는 한국케이블 TV방송협회, 초고속 인터넷의 자료는 한국인터넷진흥원의 자료를 사용하였으며 자동차의 경우는 생산연감데이터를 참고하였다.

### 4.2.1 실제 데이터의 추정 결과

다음의 <Table 8>, <Table 9>는 실제 데이터를 여러 가지 방법으로 추정한 결과이다. 각 제품의 이름과 데이터 수, 그리고

추정 결과를 모아 정리하였다.

실제 제품의 추정 결과 역시 시뮬레이션 데이터를 이용한 실험 결과와 마찬가지로 OLS와 Satoh 모형의 성능은 비슷하게 나타났다. 시뮬레이션 데이터를 이용한 실험과 달리 OLS가 Satoh 모형보다 근소하게 나은 결과를 보여주고 있다.

실제 제품의 데이터를 이용한 실험 결과의 가장 주목할 만한 점은 NLS의 추정 결과이다. MRI 장비, CATV, 초고속 인터넷의 경우 NLS는 추정을 실패하였다. 이 세 제품의 데이터 수는 각각 13개, 13개, 6개로 상대적으로 적은 수이다. 3차원 탐색을 수행하는 NLS의 방법의 특성 상 데이터 수가 부족한 경우 이와 같이 추정에 실패할 가능성이 존재한다는 한계가 드러나는 결과이다.

반면 하이브리드방법의 경우는 모든 제품에 대해서 안정적으로 추정 결과를 얻을 수 있었으며 NLS 방법이 추정에 실패한 제품에 대해서도 추정에 성공한 다른 방법들 (OLS, Satoh)보다 낮은 오차제곱합을 보여주었다.

## 5. 결론

Bass 모형을 이산화하여 모수추정에 활용하는 기존 모수추정 방법인 OLS 모형은 추정 성능에 있어서 한계를 가지고 있다. 이를 개선하기 위해 다양한 추정방법이 개발되었으며 그 중에서도 3차원 탐색을 하는 NLS 방법이 가장 뛰어난 성능을 보이고 있다. 최근에는 Satoh의 연구를 통해 기존의 OLS를 개선하는 새로운 이산화 추정 방법이 제시되었지만 역시 추정 성능에서 NLS를 따라가지 못하고 있다.

우리는 Bass 모형의 위험률(hazard rate) 표현 방식을 이용한 새로운 선형 추정 방법을 제안하였다. 이 방법은 잠재수요가 주어진 상태에서, 위험률과 침투율의 선형관계를 이용하여  $p$ 와  $q$ 를 OLS 방법으로 구하고, 잠재수요 값의 결정은 NLS 방법을 사용하는 것이다. NLS 방법의 경우 3차원 탐색에 의해 최적

**Table 8.** Estimation Results with Real Data Set 1

		MRI 장비	CT 스캐너	CATV	초음파 영상기	초고속 인터넷	자동차
데이터 수		13	14	13	8	6	43
OLS	<i>m</i>	8.83	76.65	67.04	98.00	102.48	109.07
	<i>p</i>	0.0075	0.6134	0.3163	0.3627	-0.0577	0.0001
	<i>q</i>	-0.0075	-0.6134	-0.3163	0.5515	1.9221	0.2111
	<i>ln(SE)</i>	12.286	12.167	13.576	8.627	9.604	8.861
Sato	<i>m</i>	-33.27	88.76	65.86	99.16	98.40	105.38
	<i>p</i>	-0.0172	0.2208	0.3539	1.1822	0.0230	0.0004
	<i>q</i>	0.0172	-0.2208	-0.3539	-1.1822	2.0596	0.2225
	<i>ln(SE)</i>	12.004	9.499	12.947	14.244	9.493	8.170
NLS	<i>m</i>	-	109.50	-	106.98	-	102.74
	<i>p</i>	-	0.0186	-	0.0062	-	0.0002
	<i>q</i>	-	0.2022	-	0.2768	-	0.2324
	<i>ln(SE)</i>	-	4.659	-	2.087	-	4.783
Hybrid	<i>m</i>	149.04	101.25	122.96	81.79	92.84	107.19
	<i>p</i>	0.0017	0.0087	0.0060	2.4373	0.8518	0.0003
	<i>q</i>	0.2763	0.3196	0.2734	-2.4278	-1.1440	0.2168
	<i>ln(SE)</i>	6.792	5.504	7.097	7.961	8.980	4.965

**Table 9.** Estimation Results with Real Data Set 2

		이동전화	미국 잔디깎기	미국 에어컨	미국 식기세척기	미국 냉장고
데이터 수		23	31	44	41	44
OLS	<i>m</i>	102.10	174.21	120.04	150.39	132.43
	<i>p</i>	0.0052	0.0047	0.0039	0.0016	0.0061
	<i>q</i>	0.4280	0.1072	0.1168	0.1259	0.0780
	<i>ln(SE)</i>	9.775	4.039	7.448	6.602	5.951
Sato	<i>m</i>	96.19	168.11	112.75	137.44	130.47
	<i>p</i>	0.0071	0.0054	0.0042	0.0017	0.0068
	<i>q</i>	0.4816	0.1097	0.1267	0.1371	0.0779
	<i>ln(SE)</i>	10.137	5.736	8.085	7.769	4.924
NLS	<i>m</i>	92.04	159.28	101.79	121.34	142.99
	<i>p</i>	0.0002	0.0048	0.0020	0.0010	0.0068
	<i>q</i>	0.5542	0.1169	0.1538	0.1558	0.0691
	<i>ln(SE)</i>	6.016	3.405	4.866	3.807	4.808
Hybrid	<i>m</i>	101.94	150.26	113.34	134.99	128.77
	<i>p</i>	0.0005	0.0046	0.0023	0.0011	0.0069
	<i>q</i>	0.4579	0.1264	0.1348	0.1433	0.0790
	<i>ln(SE)</i>	6.376	3.635	5.447	4.205	4.932

해가 초기해에 민감하나 본 논문에서 제시된 방법은 잠재수요 *m*만 탐색하므로, 이러한 단점을 보완하는 장점이 있다.

새로운 추정방법을 다른 추정방법들과 함께 비교하기 위하여 시뮬레이션 데이터와 실제 제품 데이터를 이용한 실험을 수행한 결과에서 OLS와 Sato 방법은 상대적으로 NLS와 하이

브리드 방법에 비해 높은 오차제곱을 보이는 등 추정 성능이 떨어지는 것으로 나타났다. 특히 시뮬레이션 데이터를 이용한 실험에서 데이터 수가 적어짐에 따라 OLS의 성능이 크게 저하되어 신뢰할만한 추정 결과를 보여주지 못하였다.

NLS는 시뮬레이션 데이터 상에서는 가장 뛰어난 성능을 보



여주었으나 실제 제품의 데이터를 이용한 추정에서는 상대적으로 데이터 수가 적은 몇몇 제품들의 추정에 실패하였다. 이는 NLS가 갖는 3차원 탐색 방법의 특성에 따른 한계라고 볼 수 있다.

하이브리드 방법은 전체 실험의 모든 데이터에 대해서 오차제곱합이 NLS에 근접하는 뛰어난 추정 성능을 보여주었으며, 특히 데이터 수가 적은 시뮬레이션 데이터 실험에서는 추정 모수의 표준편차는 NLS와 같거나 낮은 수준을 나타내어 추정 결과가 안정적임을 증명하였다. 이러한 추정치의 안정성은 실제문제에의 활용가능성 측면에서 가치가 크다고 할 수 있다. 그 이유는 실제 확산 모형을 사용하여 향후 수요를 예측하고자 할 때, 확산 모형이 주로 혁신적인 신제품의 확산 현상을 다루므로, 사용가능한 데이터 수가 많지 않은 경우가 대부분이기 때문이다. 무엇보다 데이터 수가 부족한 실제 제품의 추정에 NLS 방법이 실패한 것과는 달리 하이브리드 방법은 1차원 탐색을 이용하기 때문에 모든 경우에 있어서 추정 결과를 얻을 수 있다는 장점을 보이고 있다. 또한, NUI 모형이나 이분 조각 Bass 모형같이 Bass의 확장모형의 경우에도 본 논문에서 제시된 하이브리드 추정방법이 활용될 수 있음을 입증하였다.

## 참고문헌

- Bass, F. M. (1969), A New Product Growth Model for Consumer Durables, *Management Science*, **15**, 215-227.
- Bayus, B. (1987), Forecasting Sales of New Contingent Products : An Application to the Compact Disc Market., *Journal of Product Innovation Management*, **4**, 243-255.
- Boswijk, H. P. and Franses, P. H. (2002), The Econometrics of the Bass Diffusion Model, *Erasmus Research Institute of Management*.
- Chandraekaran, D. and Tellis, G. J. (2004), A Critical Review of Marketing Research on Diffusion of New Products, *Review of Marketing Research*, **3**, 39-80.
- Delre, S. A., Jager, W., and Janssen, M. A. (2006), Diffusion dynamics in small-world networks with heterogeneous consumers, *Comput. Math. Organ. Theory*, **13**(2), 185-202.
- Easingwood, C. J., Mahajan, V., and Muller, E. (1983), A Nonuniform Influence Innovation Diffusion Model of New Product Acceptance, *Marketing Science*, **2**(3), 273-295.
- Goldenberg, J., Han, S., Lehmann D. R., and Hong J. (2009), The Role of Hubs in the Adoption Process, *Journal of Marketing*, **73**, 1-13.
- Goldenberg, J., Libai, B., and Muller, E. (2001), Talk of the Network : A Complex Systems Look at the Underlying Process of Word-of-Mouth, *Marketing Letters*, **12**(3), 211-223.
- Goldenberg, J., Libai, B., and Muller, E. (2006), Renana Peres, Blazing Saddles: The Early And Mainstream Markets in the High-Tech Product Life Cycle, *Israel Economic Review*, **4**(2), 85-108.
- Guseo, R. and Guidolin, M. (2009), Modelling a dynamic market potential: A class of automata networks for diffusion of innovations, *Technological Forecasting and Social Change*, **76**, 806-820.
- Hirota, R. (1979), Nonlinear Partial Difference Equations. V. Nonlinear Equations Reducible to Linear Equations, *Journal of the Physical Society of Japan*, **46**, 312-319.
- Hong, J. and Uhm, S. (2009), Two pieces extension of the Bass Diffusion Model, *Journal of Korean Management Science*, **34**(4), 15-26.
- Hong, J., Kim, T., and Lim, D. (2008), A Study on a Long-term Demand Forecasting and Characterization of Diffusion Process for Medical Equipments, *Proceedings of Spring Conference, Korean Institute of Industrial Engineering*.
- Horsky, D. (1990), A diffusion model incorporating product benefits, price, income and information, *Marketing Science*, **9**, 342-365.
- Jun, D. B. and Park, Y. S. (1999), A choice-based diffusion model for multiple generations of products, *Technological Forecasting and Social Change*, **61**, 45-58.
- Kalish, S. (1985), A New Product Adoption Model with Pricing, Advertising and Uncertainty., *Management Science*, **31**, 1569-1585.
- Ko, Y., Hong, S., and Jun, C. (2008), On Parameter Estimation of Growth Curves for Technological Forecasting by Using Non-linear Least Squares, *International Journal of Management Science*, **14**(2), 89-104.
- Mahajan, V., Muller, E., and Bass, F. M. (1990), New product diffusion models in marketing : a review and directions for research, *Journal of marketing*, **54**, 1-26.
- Ozkaya, E. (2008), Demand management in global supply chains, Ph. D. Thesis, Georgia Institute of Technology.
- Park, S. J. and Seetharaman, P. B. (2005), Which Stage of the Consumer's New Product Adoption Process follows the Bass Model? An Empirical Exploration, Free Press, Houston.
- Phillips, F. (2001), Market-oriented technology management: innovating for profit in Entrepreneurial Times.
- Phillips, F. (2007), On S-curves and tipping points, *Technological Forecasting and Social Change*, Springer, **74**(6), 715-730.
- Rahmandad, H. and Sterman, J. (2007), Heterogeneity and Network Structure in the Dynamics of Diffusion: Comparing Agent-Based and Differential Equation Models, *Management Science*, **54**(5), 998-1014.
- Robinson, B. and Lakhani, C. (1975), Dynamic Price Models for New Product Planning, *Management Science*, **21**, 1113-1122.
- Satoh, D. (2001), A discrete bass model and its parameter estimations, *Journal of the Operations Research Society of Japan*, **44**(1).
- Sawhney, M. S. and Eliashberg, J. (1996), A Parsimonious Model for Forecasting Gross Box-Office Revenues of Motion Pictures, *Marketing Science*, **15**(2), 113-131.
- Schmittlein, D. and Mahajan, V. (1982), Maximum Likelihood Estimation for an Innovation Diffusion Model of New Product Acceptance, *Marketing Science*, **1**(1), 57-78.
- Srinivasan, V. and Mason, C. H. (1986), Nonlinear Least Squares Estimation of New Product Diffusion Models, *Marketing Science*, **5**(2), 169-178.
- Vakratsas, D. and Kolarici, C. (2008), A dual-market diffusion model for a new prescription pharmaceutical, *Intern. J. of Research in Marketing*, **25**, 282-293.
- Van den Bulte, C. and Joshi, Y. V. (2007), New Product Diffusion with Influentials and Imitators, *Marketing Science*, **26**(3), 400-421.
- Van den Bulte, C. and Lilien, G. L. (1997), Bias and Systematic Change in the Parameter Estimates of Macro-Level Diffusion Models, *Marketing Science*, **16**(4), 338-353.
- Venkatesan, R., Krishnan, T. V., and Kumar, V. (2004), Evolutionary Estimation of Macro-Level Diffusion Models Using Genetic Algorithms: An Alternative to Nonlinear Least Squares, *Marketing Science*, **23**(3), 451-464.