

NBD모형의 구조변화 감지*

주영진**

〈요 약〉

본 연구에서는 안정적 NBD모형에 개별구매자들의 평균적 구매율 변화로 발생하는 임의의 불안정성이 개입하였는지를 체계적으로 검정하기 위한 방법을 개발하였다. 이를 위해 본 연구에서는 안정적 NBD모형에서 개별구매자들의 갖는 평균적 구매율이 임의의 불안정성의 영향으로 구조적 변화를 일으키는지에 대한 통계적 가설로부터 우도비를 도출하였다. 또한, 본 연구에서는 반복구매가 이루어지는 패널자료를 대상으로 개발된 방법의 실증적 적합성을 살펴보았다. 본 연구의 결과는 NBD모형이 지니는 안정성의 가정을 극복할 수 있는 수단을 제공한다는 점, 마케팅환경의 변화를 조기에 감지함으로써 관련 환경변화에 신속히 대처할 수 있는 마케팅전략의 운용을 가능케 할 것이란 점, 특정 마케팅믹스전략의 효과측정에도 활용될 수 있다는 점 등을 기대할 수 있다.

주제어: NBD(Negative Binomial Distribution)모형, 패널자료, 구매빈도, 구조변화

I. 서 론

패널자료를 이용한 계량마케팅모형에 대한 연구는 80년대 중반 이후 매우 왕성하게 이루어지고 있으며, 일반적으로 패널자료를 이용한 계량마케팅모형은 NBD(Negative Binomial Distribution) 모형(Ehrenberg, 1972; Ehrenberg, 2000; Goodhardt 등, 1984; Wright 등, 2002)과 같이 구매행위 결과를 설명하기 위한 모형과 이산선택모형(Discrete Choice Model)(Guadagni 등, 1983; Ben-Akiva and Lerman, 1985)과 같이 구매행위의 원인을 설명하기 위한 모형으로 구분될 수 있다(Lilien 등, 1992; Ehrenberg 등, 2000).

이 중 NBD모형은 일반적으로 빈번한 반복구매가 이루어지는 특정상품에 대한 단위 시간동안의 개인의 구매량이 포아송분포를 따르고 그 포아송분포의 평균이 개인들간의 이질성(Heterogeneity)을 고려하기 위하여 감마분포를 따른다고 가정할 때, 단위시간동안 해당상품에 대한 전체적인 구매빈도를 음이항분포(Negative Binomial Distribution)로 설명하기 위한 모형이다. 이와 같은 NBD모형은 이산선택모형에 비해 구체적 마케팅믹스의 효과를 측정함을 목적으로 이용되기보다, 개인의 구매행위 결과에 대해 방대한 규모의 패널자료를 비교적 단순한 방법으로 효과적으로 설명함을 목적으로 한다고 할 수 있다. 이에 따라 기본적인 NBD모형은 마케팅믹스를 포함한 다양한 구매원인에 대

* 이 논문은 2005학년도 충북대학교 학술연구지원사업의 연구비지원에 의하여 연구되었음. 논문에 이용된 자료의 사용을 허락한 David R. Bell 교수에 감사사를 표함.

** 충북대학교 경영학부 조교수

한 설명이 부족하다는 비판을 받아 왔다.

이러한 비판은 Lenk 등(1993)의 연구에서와 같이 이산선택모형이 지나는 장점인 구매원인에 대한 설명을 NBD모형에 내재시킴으로써 극복될 수 있다. 이는 기본적인 NBD모형이 갖는 안정성(stationarity)이 이들 설명변수와 관련된 마케팅 전략적 및 환경적 변화로 인하여 개인의 구매율 등이 설명변수의 변화 이전과 상이해짐에 따라 안정성이 유지되지 못함에 따라 나타나는 불안정성(non-stationarity)을 모형에 추가시키는 의미를 갖는다.

한편, 기본적 NBD모형의 장점이 다른 설명변수 없이 구매자패널이 보이는 특정제품에 대한 구매행위만을 이용하여 해당제품의 전체적인 구매행위 결과를 효과적으로 설명하고 있다는 점에 의미를 둔다면, 기본적 NBD모형에 설명변수를 내재시키는 대신 안정적 NBD모형이 불안정성을 나타내는지에 대한 판단이 필요하다. 이는 기업의 마케팅환경에서 예상치 못한 불안정성을 조기에 감지하기 위한 수단을 제공할 것이다.

이에 본 연구에서는 NBD모형을 기반으로 기준시점과 비교시점으로 구분된 두 시점 사이에서 기준시점의 구매행위 결과가 안정적으로 지속되어 비교시점의 구매행위 결과가 나타난 것인지 또는 기준시점에 없었던 마케팅 전략 및 환경의 변화에 의해 발생한 외부적 충격이 더해져 개별구매자들의 평균적 구매율이 변화된 형태로 비교시점의 구매행위 결과가 나타난 것인지를 감지하기 위한 방법론을 개발하였다. 즉, 본 연구는 NBD모형을 구분된 두 시점 사이에 적용할 수 있는 이변량(Bivariate) NBD모형

및 기준시점의 구매결과에 따른 비교시점에 대한 조건부(Conditional) NBD모형을 바탕으로 기준시점의 안정적 NBD모형이 비교시점에서 개별구매자들의 평균적 구매율이 임의의 외부적 충격을 받았는지를 검정하기 위한 절차를 개발하였다. 또한, 개발된 방법의 실증적 유용성을 확인하고자 빈번한 반복구매가 이루어지는 상품을 대상으로 한 패널자료에 적용하였다.

이를 위해 본 연구에서는 II장에서 NBD모형에 대한 기존의 연구들을 고찰하고, 안정적 NBD모형, 이변량 NBD모형 및 조건부 NBD모형 등의 모형적 측면을 간략히 정리하였다. III장에서는 안정적 NBD모형에서 개별구매자들의 평균적 구매율을 변화시키는 임의의 불안정성을 감지하기 위한 체계적인 방법을 개발하였고, IV장에서 개발된 방법에 대한 실증분석을 수행하였다. 마지막으로 V장에서는 본 연구가 갖는 의미와 한계 등을 정리하였다.

II. 배 경

1. 문헌고찰

NBD모형은 음이항분포를 이용하여 효과적으로 설명될 수 있는 구매자 패널자료가 나타내는 특정 제품에 대한 전체적인 구매행위 결과를 추정하고, 안정적 조건하에서 추정결과가 미래시점의 전체적인 구매행위 결과를 예측함에 있어서도 효과적인 모형이다. NBD모형은 특히 여러 마케팅믹스 변수 등 다른 설명변수의 도움 없이 구매행위자

체에 대한 패널자료만으로 구성된 모형인 탓에 그 적용이 비교적 용이한 장점을 지닌다. 이러한 NBD모형에 대한 체계적인 이해는 Ehrenberg(1972, 2000)를 참고할 수 있다. Ehrenberg(1972, 2000)는 구매자 패널자료가 나타내는 특정 제품에 대한 전체적인 구매행위 결과를 효과적으로 설명함에 적합한 NBD모형 및 LSD(Logarithmic Series Distribution)모형에 대한 기본 개념 및 적용과정을 상세하게 다루고 있다. 기본적인 NBD모형은 Goodhardt 등(1984)의 연구에서와 같이 해당 제품의 개별 브랜드에 대한 개인의 선택이 선호확률이 Dirichlet 분포를 통해 이질적으로 표현되는 상황에서 다항분포 (Multinomial Distribution)를 따른다는 가정을 결합하여 NBD-Dirichlet모형으로 확장되어 해당 제품의 브랜드별 경쟁상황까지를 효과적으로 설명할 수 있게 되었다.

한편, Lenk 등(1993)의 연구에서는 Tide와 분말세제를 대상으로 기본적인 NBD모형이 지니는 안정성의 문제를 극복하기 위하여 광고, 쿠폰, 가격 등 마케팅믹스변수들의 효과를 안정적 NBD모형에 내재시킬 수 있는 모형을 개발하여 모형의 설명력을 높이고 여러 설명변수의 효과를 측정할 수 있는 수단을 제공하였다.

또한 최근에는 웹사이트에 대한 인터넷 사용자들의 방문기록이 로그파일 등을 통해 매우 효과적으로 제공될 수 있는 환경에서, 특정 웹사이트(들)에 대한 인터넷 소비자의 패널자료에 NBD모형을 적용한 연구들이 있는데 이석규(2003), 김소영 등(2002)의 연구 등이 있다. 이석규(2003)의 연구에서는 NBD모형이 인터넷 소비자들이 특정 웹사이트에 방문하는 행위를 효과적으로 설명하

기 위한 수단이 될 수 있음과 함께 NBD모형에 설명변수를 내재시켜 인터넷 소비자들의 일반적인 인터넷 사용패턴과 인구 통계적인 변수들이 반복된 이용 빈도에 미치는 영향을 조사하였다. 김소영 등(2002)의 연구에서는 웹사이트 방문행동을 구성하는 사이트밀착도, 페이지뷰, 체류시간과 방문빈도 등을 이용하여 커뮤니티사이트에 대한 eCRM 전략을 도출하기 위한 과정에서 방문빈도를 설명하기 위한 모형으로 NBD모형을 사용하였다.

2. NBD 모형

NBD모형에 의하면 특정제품에 대한 단위시간동안의 특정개인의 구매빈도가 포아송분포(Poisson distribution)를 따르고, 개인간의 이질성(heterogeneity)을 고려하기 위하여 각 개인의 구매빈도를 설명하는 포아송분포의 평균이 감마분포(Gamma distribution)를 따를 때, 단위시간동안 해당제품에 대한 전체적인 구매빈도가 음이항분포(Negative Binomial distribution)로 설명될 수 있다 (Ehrenberg, 1972 (p. 132)).

이에 따라 NBD모형은 T 시간동안 특정제품에 대한 전체적인 구매빈도가 r 일 확률을 식 (1)과 같이 표현할 수 있다.

$$\Pr(r) = \left(1 + \frac{m}{k} T\right)^{-k} \frac{\Gamma(k+r)}{\Gamma(r+1)\Gamma(k)} \left(\frac{mT}{k+mT}\right)^r \quad (1)$$

식 (1)에서 m 과 k 는 각각 감마이질성을 나타내기 위해 도입된 감마분포의 평균과 지수이다.

NBD모형은 해당제품에 대한 개인의 구매행위가 안정적으로 지속되는 경우, 복수시점에서의 구매빈도에 대한 다변량 NBD 모형으로 확장될 수 있다. 이 때, 복수시점 $\{T_1, T_2, \dots, T_t\}$ 동안의 특정제품에 대한 전체적인 구매빈도가 $\{r_1, r_2, \dots, r_t\}$ 일 결합확률은 식 (2)의 확률발생함수를 이용하여 구할 수 있다(Ehrenberg, 1972 (p. 136)).

$$(1 + a \sum_{i=1}^t T_i (1 - u_i))^{-k} \quad (2)$$

식 (2)에서 a 는 m/k 로 정의되며, u_i 는 확률발생함수의 정의를 위해 사용된 더미변수이다. 식 (2)의 확률발생함수로부터 복수시점 $\{T_1, T_2, \dots, T_t\}$ 동안의 특정제품에 대한 전체적 구매빈도가 $\{r_1, r_2, \dots, r_t\}$ 일 결합확률은 식 (2)를 전개하여 구분될 수 있는 $(u_1)^{r_1} (u_2)^{r_2} \dots (u_t)^{r_t}$ 항의 계수로 구해지게 된다.

이변량 NBD모형은 식 (2)의 다변량 NBD모형을 위한 확률발생함수의 t 를 2로 뒀으로써 구할 수 있으며, 이에 따라 두 시점 $\{T_1, T_2\}$ 동안의 특정제품에 대한 전체적인 구매빈도가 $\{r_1, r_2\}$ 일 결합확률은 식 (3)과 같이 구해진다.

$$\Pr(r_1, r_2) = \frac{(1 + a(T_1 + T_2))^{-k} \Gamma(k + r_1 + r_2)}{\Gamma(r_1 + 1) \Gamma(r_2 + 1) \Gamma(k)} \left(\frac{a}{1 + a(T_1 + T_2)} \right)^{r_1 + r_2} T_1^{r_1} T_2^{r_2} \quad (3)$$

또한, 식 (1)과 식 (3)으로부터 두 시점 $\{T_1, T_2\}$ 사이에서 T_1 동안의 전체적인 구매

빈도가 r_1 인 경우, T_2 동안의 전체적인 구매빈도가 r_2 일 조건부확률(P_{r_1, r_2})은 다시 NBD모형을 따르며 식 (4)와 같이 구해질 수 있다.

$$P_{r_1, r_2} = \Pr(r_2 | r_1) = \frac{\Pr(r_1, r_2)}{\Pr(r_1)} = (1 + a' T_2)^{-k} \frac{\Gamma(k' + r_2)}{\Gamma(r_2 + 1) \Gamma(k')} \left(\frac{a' T_2}{1 + a' T_2} \right)^{r_2} \quad (4)$$

식 (4)에서 k' 과 a' 은 각각 $k + r_1$, $a/(1 + a T_1)$ 으로 정의된다.

III. 모 형

1. 문제제기 및 모형설정

이제 본 연구에서는 두 시점 $\{T_1, T_2\}$ 동안의 특정제품에 대한 전체적인 구매빈도가 NBD모형을 따르지만, 기준시점인 T_1 동안의 구매행위에 비해 비교시점인 T_2 동안의 구매행위가 구조적 변화를 지니는 경우를 가정한다. 비교시점에서의 구매행위가 기준시점에 비해 구조적 변화를 지닌다는 것은 기본적인 NBD모형에 대한 안정성(stationarity)의 가정을 완화하여 불안정성(non-stationarity)을 고려한 NBD모형의 한 가지 유형을 제공하게 된다. 이때 일반적으로 고려될 수 있는 불안정성의 근원으로는 계절요인, 외부적인 마케팅환경의 변화, 내부적인 마케팅믹스전략의 변화 등 다양한 것을 꼽을 수 있다.

안정적 NBD모형은 조건부추세분석(Conditional

Trend Analysis (CTA))을 통해 기준시점의 구매행위가 안정적으로 지속될 경우 도출될 수 있는 비교시점의 구매빈도에 대한 이론적 분포를 비교시점에서 실제 관찰할 수 있는 구매빈도와 비교함으로써, 기준시점과 비교시점간에 안정적 구매행위가 지속되고 있는지에 대한 판단의 수단을 제공하고 있다(Ehrenberg, 1972 (p. 141)). 이와 관련하여 Schmittlein과 Morrison(1983)의 연구에서는 안정적인 NBD모형 및 CNBD(Condensed NBD)모형의 CTA를 통한 개별 구매자의 구매행위에 대한 예측력을 비교하고 있다. 한편, CTA를 통한 기준시점과 비교시점간의 안정적 구매행위의 지속여부에 대한 판단은 체계적인 통계적 판단에 의한 것이기 보다 단순히 이론적 구매빈도와 관측된 구매빈도간의 절대적 차이를 임의로 판단하는 것이다. 불안정적 NBD모형에 대한 후속연구들은 CTA에 대한 통계적 판단의 체계화보다는 Lenk 등(1993)에서와 같이 불안정성의 근원인 광고, 쿠폰, 가격 등의 마케팅믹스변수들의 효과를 안정적 NBD모형에 내재시켜 모형의 설명력을 높이는 방향이 주를 이루고 있다.

마케팅믹스 등을 고려한 불안정적 NBD 모형의 결과는 마케팅믹스변수들의 효과를 측정할 수 있는 수단을 제공함으로써 효율적인 마케팅의사결정수립에 활용할 수 있는 장점을 제공할 것이나, 안정적 NBD모형이 다른 설명변수의 도움 없이 구매결과에 대한 자료만으로 구매행위를 효과적으로 설명하고 있는 장점을 유지하기 위해서는 안정적 NBD모형에서의 CTA에 대한 체계적인 통계적 판단방법을 개발하는 것이 요구된

다. 이는 특정제품에 대한 구매자 패널의 구매 자료만을 관찰하여 전체적인 구매행위가 유의하게 변화하였음을 조기에 감지하기 위한 유용한 수단으로 활용될 수 있다.

이에 따라 본 연구에서는 안정적 NBD모형에서 개별구매자들의 평균적 구매율을 변화시키는 임의의 불안정성에 기인하여 기준시점에 비해 비교시점에서의 전체적인 구매행위가 유의하게 차이가 발생하는지의 여부를 판단하기 위한 통계적 방법을 개발하고자 한다.

이러한 개별구매자들의 평균적 구매율을 변화시키는 임의의 불안정성에 기인한 비교시점에서의 전체적인 구매행위의 변화를 모형화 하기 위해 본 연구에서는 안정적 NBD모형의 도출과정에서 각 개인의 구매빈도의 감마이질성을 위해 도입된 감마분포의 평균이 임의의 불안정성(Δ)에 의해 비교시점에서 변화됨을 가정한다. 즉, 임의의 불안정성(Δ)이 통계적으로 유의하지 않다면 전체적인 구매행위는 기준시점과 비교시점에서 모두 식 (1)을 따를 것이지만, 임의의 불안정성(Δ)이 통계적으로 유의하다면 식 (1)에서 감마분포의 평균 m 이 $m + \Delta$ 로 변화하여 전체적인 구매행위는 기준시점에서는 식 (1)을 따르지만 비교시점에서는 식 (5)를 따르게 된다.

$$\Pr(r) = \left(1 + \frac{m + \Delta}{k} T\right)^{-k} \frac{\Gamma(k+r)}{\Gamma(r+1)\Gamma(k)} \left(\frac{(m + \Delta)T}{k + (m + \Delta)T}\right)^r \quad (5)$$

2. 모형검정

식 (5)에서 임의의 불안정성(Δ)이 통계적으로 유의한지 여부를 검정하기 위하여, 전체 n 명의 구매자패널을 대상으로 해당제품에 대한 기준시점과 비교시점의 전체적인 구매빈도 $\{\{n_{00}, n_{01}, n_{02}, \dots\}, \{n_{10}, n_{11}, n_{12}, \dots\}, \{n_{20}, n_{21}, n_{22}, \dots\}, \dots\}$ 를 분석하고자한다. 이때 n_{ij} 는 기준시점과 비교시점에 해당제품을 각각 i 개와 j 개를 구매한 구매자패널의 수를 의미하며, $\{n_{i0}, n_{i1}, n_{i2}, \dots\}$ 는 기준시점의 전체적인 구매빈도가 i 개인 경우에 대한 비교시점의 전체적인 구매빈도로서 이러한 조건부구매행위는 식 (4)에서와 같이 다시 NBD모형을 따른다. 단 이후의 수식에서는 식 (4)의 T_1, T_2 를 각각 1과 T 로 두어 T 를 기준시점의 시간 대비 비교시점의 시간으로 새롭게 정의한다.

이제 특정제품에 대한 전체적인 구매행위가 기준시점에 비해 비교시점에서 임의의 불안정성에 영향을 받았는지에 대한 통계적 판단은 식 (5)의 Δ 에 대해 식 (6)의 가설을 검정함으로써 이루어지게 된다.

$$\begin{aligned} H_0 : \Delta = 0 \\ H_a : \Delta \neq 0 \end{aligned} \tag{6}$$

식 (6)의 가설에 대한 검정을 위해 수집 자료 $\{\{n_{00}, n_{01}, n_{02}, \dots\}, \{n_{10}, n_{11}, n_{12}, \dots\}, \{n_{20}, n_{21}, n_{22}, \dots\}, \dots\}$ 에 대한 H_0, H_a 하의 우도비(Likelihood Ratio (Λ))를 이용할 수 있다. 식 (7), 식 (8)은 각각 H_0, H_a 하의 우도(Likelihood) $l_0(\cdot), l_a(\cdot)$ 를 나타낸다.

$$\begin{aligned} l_0(\cdot; \{n_{00}, n_{01}, n_{02}, \dots\}, \{n_{10}, n_{11}, n_{12}, \dots\}, \dots) \\ = \Pr(n_{00}, n_{01}, \dots | n_{0\cdot}) \Pr(n_{10}, n_{11}, \dots | n_{1\cdot}) \\ \dots \Pr(n_{0\cdot}, n_{1\cdot}, \dots) \\ = \{\Pi_i \Pi_j p_{ij0}^{n_{ij}}\} \Pr(n_{0\cdot}, n_{1\cdot}, \dots) \end{aligned} \tag{7}$$

단, $n_{i\cdot} = n_{i0} + n_{i1} + n_{i2} + \dots$

$$\begin{aligned} p_{ij0} &= \left(1 + \frac{m'}{k'} T\right)^{-k'} \\ &\frac{\Gamma(k'+j)}{\Gamma(j+1)\Gamma(k')} \left(\frac{m' T}{k' + m' T}\right)^j \\ k' &= k+i, \quad m' = \frac{m(k+i)}{k+m} \end{aligned}$$

$$\begin{aligned} l_a(\cdot; \{n_{00}, n_{01}, n_{02}, \dots\}, \{n_{10}, n_{11}, n_{12}, \dots\}, \dots, \Delta) \\ = \Pr(n_{00}, n_{01}, \dots | n_{0\cdot}, \Delta) \Pr(n_{10}, n_{11}, \dots | n_{1\cdot}, \Delta) \dots \Pr(n_{0\cdot}, n_{1\cdot}, \dots) \\ = \{\Pi_i \Pi_j p_{ija}^{n_{ij}}\} \Pr(n_{0\cdot}, n_{1\cdot}, \dots) \end{aligned} \tag{8}$$

$$\begin{aligned} \text{단, } p_{ija} &= \left(1 + \frac{m''}{k'} T\right)^{-k'} \\ &\frac{\Gamma(k'+j)}{\Gamma(j+1)\Gamma(k')} \left(\frac{m'' T}{k' + m'' T}\right)^j \\ m'' &= \frac{(m+\Delta)(k+i)}{k+m+\Delta} \end{aligned}$$

식 (8)에서 $n_{i\cdot}$ 및 k' 에 대한 정의는 식 (7)에서와 같으며, 식 (7)과 식 (8)로부터 H_0, H_a 하의 우도비(Λ)는 식 (9)와 같이 도출된다.

$$\begin{aligned} \Lambda &= \Pi_i \Pi_j \left(\frac{p_{ij0}}{p_{ija}}\right)^{n_{ij}} \\ &= \Pi_i \Pi_j \left\{ \left(\frac{k' + m' T}{k' + m'' T}\right)^{-(k'+j)n_{ij}} \left(\frac{m'}{m''}\right)^{jn_{ij}} \right\} \end{aligned} \tag{9}$$

식 (9)에서 k', m' 및 m'' 에 대한 정의는 식 (7) 및 식 (8)에서와 같다.

마지막으로 식 (9)의 Λ 를 이용하여 $-2\log\Lambda$ 가 χ^2 분포(자유도=1)를 따르는 것

을 이용하여 식 (6)의 가설에 대한 검정을 수행할 수 있다. 이때 식 (8)에서 요구되는 임의의 불안정성(Δ)은 그 추정치로서 비교시점과 기준시점의 전체적인 구매빈도에 대한 평균의 차이를 이용할 수 있다.

IV. 사례연구

1. 개요

본 연구에서는 앞서 개발하여 제시한 NBD모형에서 개별구매자들의 평균적 구매율을 변화하는 임의의 불안정성에 대한 검정을 Bell과 Lattin(1998)의 연구에서 사용되었던 Stanford Basket Data로 알려진 IRI 패널자료에 적용하여 그 유용성을 확인하였다. 사용한 패널자료는 도시지역에 거주하는 494명의 패널리스트들이 104주 동안에 베이컨을 구매한 자료이다.

기준시점과 비교시점은 단위시간의 길이를 고려하여 연간 구매행위 및 분기별 구매행위에 대한 분석으로 구분하여 수행하였다. 연간 구매행위에 대한 분석에서는 최초 52주 동안의 구매를 기준으로 이후의 52주 동안의 구매를 비교하였고, 분기별 구매행위에 대한 분석에서는 104주를 총 8구간으로 균등하게 나누어 최초 13주 동안의 구매를 기준으로 이어지는 13주씩의 7개 구간에서의 구매를 비교하였다.

이제 NBD모형에서 개별구매자들의 평균적 구매율의 변화를 유발하는 임의의 불안정성에 대한 검정을 본 연구에서는 다음과

같은 절차를 통해 수행하였다.

① 기준시점에서의 구매자패널의 전체적인 구매빈도를 통해 NBD모형의 모수 m , k 를 추정한다. 본 연구에서는 m 에 대한 추정치로는 관찰된 구매빈도의 평균을 이용하였으며, k 에 대한 추정치는 관찰된 구매빈도의 분산을 이론적 분산인 $m(1+m/k)$ 과 일치시켜 추정하는 모멘트법을 이용하였다.

② 기준시점과 비교시점의 교차구매빈도(모든 i, j 들에 대한 식 (9)의 n_{ij})를 측정한다.

③ 기준시점과 비교시점간의 임의의 불안정성(Δ)를 추정한다. 본 연구에서는 앞에서 제시한 바와 같이 Δ 의 추정치로 비교시점과 기준시점의 전체적인 구매빈도 평균의 차이를 이용하였다.

④ ①~③의 과정을 통해 구한 m , k , Δ 에 대한 추정치와 모든 i, j 들에 대한 $\{n_{ij}\}$ 값들을 이용하여 식 (9)의 우도비(Λ)를 구한 다음 $-2\log\Lambda$ 에 대해 χ^2 검정(자유도=1)을 수행한다.

2. 모형적용

이제 앞에서 제시된 절차에 의해 연간 및 분기별 구매행위에 개별구매자들의 평균적 구매율의 변화를 유발하는 임의의 불안정성이 개입되었는지를 검정하였다.

먼저, 연간 및 분기별 기준시점인 최초 52주 및 최초 13주에 대한 NBD모형의 모수 $\{m, k\}$ 에 대한 추정치는 각각 $\{5.719, 0.569\}$ 와 $\{1.283, 0.613\}$ 으로 구해졌다. 이에 따라 <그림 1>과 <그림 2>는 각각 연간 및 분기별 기준시점의 구매횟수별 빈도의

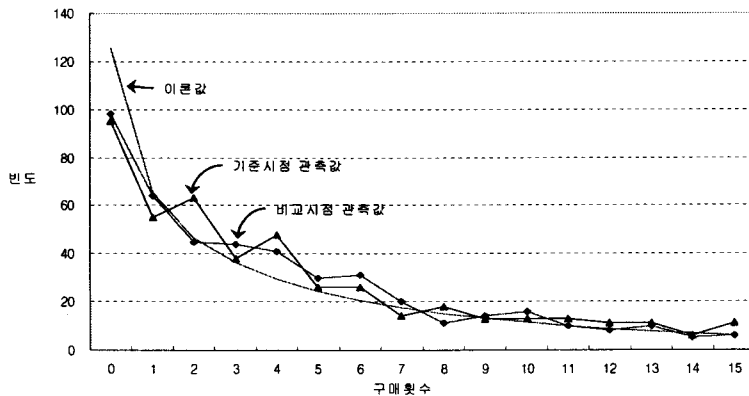
관측값을 추정된 NBD모형에 의한 이론적인 값 및 비교시점의 구매횟수별 빈도의 관측값 등과 비교하여 나타내고 있다. 이때 비교시점은 <그림 1>의 연간 구매에서는 둘째년도인 53주부터 104주까지를, <그림 2>의 분기별 구매에서는 제3분기인 27주부터 39주까지를 사용하였다.

<그림 1>의 연간 구매에서는 비교시점의 구매횟수 빈도가 기준시점 및 기준시점에 대한 이론값과 유사한 안정적인 모습을 보

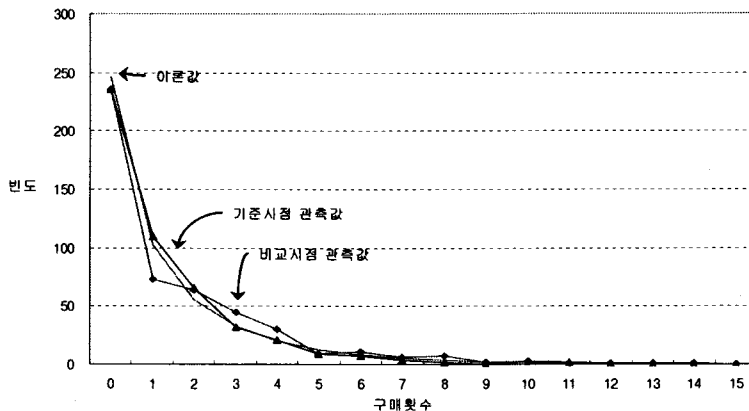
이고 있으나, <그림 2>의 분기별 구매에서는 비교시점의 구매횟수 빈도가 기준시점 및 기준시점에 대한 이론값에 비해 낮은 구매횟수에 대한 빈도가 줄어든 반면 높은 구매횟수에 대한 빈도가 늘어난 것으로 나타나 기준시점에 비해 비교시점에서의 전반적인 구매율이 높아졌을 것으로 예상할 수 있다.

<표 1>은 연간 구매행위에 대한 기준시점과 비교시점의 교차구매빈도를 예시하고

<그림 1> 연간 구매빈도에 대한 NBD모형의 추정결과 및 비교



<그림 2> 분기별 구매빈도에 대한 NBD모형의 추정결과 및 비교



<표 1> 기준시점과 비교시점의 연간 교차구매빈도

구분		비교시점 구매빈도						기준시점 구매빈도합
		0	1	2	3	...	68	
기준시점 구매빈도	0	57	18	5	5	...	0	95
	1	13	13	7	5	...	0	55
	2	8	15	8	9	...	0	63
	3	10	3	6	5	...	0	38
	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
	94	0	0	0	0	...	0	1
비교시점 구매빈도합		98	64	45	44	...	1	494

있다. <표 1>에 의하면 총 494명의 패널 중 57명은 기준시점과 비교시점 모두 전혀 구매를 하지 않은 것으로 나타났으며, 358명은 기준시점과 비교시점에 모두 1회 이상의 구매를 한 반복구매자로, 38명은 비교시점에서의 신규구매자로, 나머지 41명은 기준시점의 구매자가 비교시점에서 구매를 중단한 것으로 나타났다. 또한, <표 1>에서 '기준시점 구매빈도합'은 연간 구매행위의 기준시점에 대한 NBD모형을 추정하기 위하여 사용된 자료로 <그림 1>에서의 관측값에 해당되며, <표 1>의 각 행별 '비교시점 구매빈도'는 식 (4)의 조건부 구매확률에 지배받는 NBD모형을 따른다. 즉, 식 (9)에 제시된 임의의 불안정성 검정을 위한 우도비(Λ)는 <표 1>의 교차빈도가 종합적으로 임의의 불안정성이 개입되지 않은 경우와 개입된 경우 중 어느 쪽에 가까운가를 판단하기 위한 통계적 척도라고 할 수 있다.

이제 <표 2>에서는 본 연구에서 제안된 방법에 의하여 수행된 연간 및 분기별 구매행위에 대해 비교시점에서 개별구매자들의 평균적 구매율의 변화를 유발하는 임의의 불안정성이 개입하였는가에 대한 가설검정

의 결과를 비교모형에 의한 가설검정 결과와 함께 나타내고 있다. 이 때, 비교모형에 의한 가설검정은 비교시점에서의 구매횟수에 대한 빈도만을 대상으로 기준시점에서 추정된 NBD모형의 모수가 비교시점에서도 지속적으로 같다는 귀무가설과 다르다는 대립가설간의 우도비검정을 이용하였다.

<표 2>의 연간 구매에 대한 임의의 불안정성 검정은 기준시점인 최초 52주 동안에 비해 이후의 52주 동안에 임의의 불안정성이 개입하였는지에 대한 검정을 수행하고 있으며, 분기별 구매에 대한 임의의 불안정성 검정은 기준시점인 제1분기(최초 13주 동안)에 비해 이후의 이어지는 7분기(13주 동안의 7개 구간)에서 임의의 불안정성이 개입하였는지에 대한 검정을 수행하고 있다. <표 2>에서 알 수 있는 바와 같이 연간 구매에 있어서는 비교시점에서 임의의 불안정성이 개입하였다고 할 수 없으며, 분기별 구매에 있어서는 제2분기, 제4분기, 제5분기, 제6분기 및 제8분기에서는 기준시점에 비해 임의의 불안정성이 개입하였다고 할 수 없지만, 제3분기와 제7분기(각각 27주~39주와 79주~91주)에서는 $-2\log\Lambda$ 값

<표 2> 연간 및 분기별 비교시점에서의 임의의 불안정성에 대한 검정과 비교

구분	연간 구매 ^a	분기별 구매 ^a						
		14주 ~26주	27주 ~39주	40주 ~52주	53주 ~65주	66주 ~78주	79주 ~91주	92주 ~104주
비교시점	53주 ~104주							
\hat{m}	5.719	1.283						
\hat{k}	0.569	0.613						
$\hat{\Delta}f$	0.134	0.079	0.415	0.091	0.182	0.219	0.340	-0.022
$-2\log\Lambda b$	0.059 (0.8076)	0.322 (0.5704)	7.478 (0.0062)	0.426 (0.5141)	1.620 (0.2031)	2.289 (0.1303)	5.207 (0.0225)	0.027 (0.8691)
\hat{m}_1cf	5.852	1.362	1.698	1.374	1.466	1.502	1.623	1.261
\hat{k}_1cf	0.546	0.425	0.494	0.417	0.408	0.375	0.599	0.549
$-2\log\Lambda_1d$	-3.493e (1.0000)	-0.596e (1.0000)	20.336 (0.0000)	3.866 (0.1447)	-0.645e (1.0000)	5.730 (0.0570)	12.307 (0.0021)	2.123 (0.3460)

- a: 기준시점은 연간 구매의 경우 1주~52주, 분기별 구매의 경우 1주~13주임
- b: 제안모형에 의한 우도비통계량($\chi^2_{df=1}$) 및 p-값
- c: 비교시점에 대한 독립적인 모수추정치(모멘트법을 이용함)
- d: 비교모형에 의한 우도비통계량($\chi^2_{df=2}$) 및 p-값; $\Lambda_1 = \Pi \left\{ \frac{l(\cdot | \hat{m}, \hat{k})}{l(\cdot | \hat{m}_1, \hat{k}_1)} \right\}$
- e: MLE에 근거한 우도비통계량이 아님으로 인하여 검정통계량이 음수로 나타남에 따라 p값을 1로 정함
- f: \hat{m}, \hat{k} 의 추정과 구분되어 추가로 추정됨

과 p값이 각각 7.478, 0.0062 및 5.207, 0.0225로 나타나 임의의 불안정성이 개입하였다고 할 수 있다.

이러한 검정결과는 <표 2>에서 보이는 바와 같이 기준시점과 비교시점에서 독립적인 NBD모형을 적용한 결과에 따른 우도비검정에서 제6분기에서 비교적 높은 유의수준 하에서 임의의 불안정성이 개입하였다고 볼 수 있는 것을 제외하고는 유사한 결과를 나타내고 있으며, 본 연구에서 제안된 모형에 의한 검정방법이 다소 극단적인 검정결과를 보이고 있는 비교모형에 의한 검정방법에 비하여 통계적으로 좀더 강건한 결과를 보이고 있다고 할 수 있다.

한편, 기준시점에 비해 비교시점에서 임의의 불안정성이 개입하였다고 판단될 경우

대상 제품의 구매상황에 대한 추가적인 분석을 통해 그와 같은 임의의 불안정성의 근원이 무엇인지를 밝히기 위한 추가적 노력이 요구된다. 이와 관련하여 본 연구에서 <표 2>에서와 같이 연간 구매행위는 뚜렷한 임의의 불안정성이 개입되지 않았으나 분기별 구매행위에 있어서는 대체로 겨울철에 해당되는 제3분기와 제7분기에 있어 개별구매자들의 평균적 구매율을 변화하는 임의의 불안정성이 개입되었다고 할 수 있는 것은 적용 대상 제품인 베이컨의 특성상 계절주기의 영향을 받아 겨울철에 전반적인 구매횟수가 증가하였기 때문으로 유추해 볼 수 있다.

V. 결 론

본 연구에서는 안정적 NBD모형이 개별 구매자들의 평균적 구매율의 변화를 유발하는 임의의 불안정성에 영향을 받아 기준시점과 비교시점에서 변화된 구매행위 결과를 보이는지에 대한 검정모형을 개발하여 패널자료에 적용하여 실증적 유용성을 확인하였다.

이를 위해 본 연구에서는 임의의 불안정성에 의해 개별구매자들의 감마이질성을 나타내는 감마분포의 평균이 변화하는 상황을 가정하고, 기준시점의 감마분포의 평균이 비교시점에서 임의의 불안정성 Δ 의 영향을 받았는지의 여부에 대한 가설을 검정하기 위한 우도비를 도출하였으며, 가설검정을 위해서는 도출된 우도비를 이용한 χ^2 검정이 이용되었다. 또한 본 연구는 개발된 임의의 불안정성에 대한 검정방법을 Stanford Basket Data 중 도시지역 패널리스트의 베이컨 구매에 적용하여 해당 제품이 년별 자료에서는 안정성이 유지되나 분기별 자료에서 일부 불안정성이 존재함을 확인하였으며, 분기별 자료에 존재하는 불안정성은 계절요인에 의해 발생하는 것으로 해석할 수 있다.

본 연구의 결과는 기존의 NBD모형에 계절성 및 마케팅믹스 등에 기인한 외부적 충격을 감지하기 위한 방법을 결합함으로써 NBD모형이 지니는 안정성의 가정을 극복하기 위한 수단을 제공하였다. 이에 따라 본 연구에서 개발된 방법을 패널자료에 적용함으로써 마케팅환경의 변화를 조기에 효과적으로 감지하여 그 원인에 대한 파악을

위한 추가적인 노력을 결합하여 마케팅환경의 변화에 적시적으로 대처할 수 있는 마케팅전략의 운용이 가능할 것이다. 또한, 본 연구에서 개발된 방법을 이용하여 사전적으로 알 수 있는 마케팅믹스 전략을 구사한 시기를 전후로 구분된 기준시점과 비교시점에서 전략을 구사한 이후의 패널자료 구매 행위 결과가 임의의 불안정성을 보이는가를 검정함으로써 관련 마케팅믹스 전략의 효과 측정에 활용할 수 있을 것이다.

그러나 본 연구의 결과는 임의의 불안정성이 어떤 원천에서 발생된 것인가에 대한 해답을 분명히 제시할 수 없는 한계를 지니고 있어, 이산선택모형이나 설명변수를 포함한 NBD모형에서와 같이 사전에 정의된 설명변수의 유의성을 바탕으로 정의된 인과모형에 비해 구체적인 전략적 활용을 위해서는 여전히 제한적이라 할 수 있다. 이에 따라 본 연구의 결과는 패널자료에서 사전에 정의될 수 없는 보다 포괄적인 불안정성을 검정하기 위한 목적으로 다른 인과모형과 상호 보완적으로 사용될 필요가 있다.

한편, 본 연구는 NBD모형에 개입된 임의의 불안정성을 감마분포의 평균을 통해서 반영하였는데, 이는 기준시점에서 모든 수준의 구매빈도를 보인 패널리스트들이 일관되게 일정한 수준의 구매빈도 변화를 나타냄을 의미하는 것이다. 그러나 임의의 불안정성에 의한 구매빈도의 변화는 기준시점의 구매빈도 수준에 따라 다르게 나타날 수 있다. 즉, 기준시점에서 낮은 수준의 구매빈도를 보이는 패널리스트와 높은 수준의 구매빈도를 보이는 패널리스트가 임의의 불안정성에 반응하는 결과는 다르게 나타날 수 있

으며, 이를 위해 본 연구에서 개발된 방법은 NBD모형에서 임의의 불안정성이 감마 분포의 평균과 분산 모두에 영향을 줄 수 있는 형태로 일반화 될 필요가 있다. 아울러 본 연구에서 NBD모형의 틀에서 개발된 임의의 불안정성 검정방법은 Chatfield와 Goodhardt(1973)에서 언급한 바와 같이 보다 정규적인 구매패턴을 따르는 CNBD모형의 틀 및 특정 제품의 브랜드별 경쟁까지를 설명하기 위한 NBD-Dirichlet모형의 틀로 확장되어 개별 브랜드차원의 구조적 변화를 함께 검정할 수 있는 형태로 일반화 될 필요가 있다.

또한 본 연구에서 제시된 임의의 불안정성에 대한 검정방법이 <표 2>에서 비교되고 있는 비교모형을 포함한 다른 검정방법들에 비해 효과적이지를 확인하기 위한 추가적인 연구가 요구된다.

이상의 NBD모형에서 확인된 임의의 불안정성의 원천을 탐색하기 위한 방법론의 보완, NBD모형에서의 임의의 불안정성 검정방법의 일반화, 제안된 검정방법에 대한 추가적 검증 등은 본 연구의 결과를 확장하여 진행될 필요가 있는 주요 연구주제들이다.

(논문접수일: 2005년 9월 14일)

(게재확정일: 2006년 1월16일)

참고문헌

김소영, 한상만, 박세환 (2002), “웹사이트 평가요인과 고객 방문행동의 관계에 관한 연구: STIVI 매트릭스를 이용한

eCRM전략을 중심으로,” 경영학연구, 제 31권 제2호, 485-507.

이석규 (2003), “인터넷 이용자들의 웹사이트 재방문 빈도에 관한 실증적 연구,” 마케팅과학연구, 제11집.

Bell, David R. and James M. Lattin (1998), “Shopping Behavior and Consumer Preference for Store Price Format: Why Large Basket Shoppers Prefer EDLP,” *Marketing Science*, Vol. 17, No. 1, 66-88.

Ben-Akiva and Lerman (1985), *Discrete Choice Analysis: Theory and Application to Travel Demand*, Cambridge: MIT Press.

Chatfield, C. and G. J. Goodhardt (1973), “A Consumer Purchasing Model with Erlang Inter-Purchase Times,” *Journal of the American Statistical Association*, Vol. 68, No. 344, 828-835.

Ehrenberg, Andrew S. C. (1972), *Repeat-Buying: Theory and Applications*, Amsterdam: North-Holland Publishing Company.

Ehrenberg, Andrew S. C. (2000), “Repeat-Buying,” *Journal of Empirical Generalisations in Marketing Science*, Vol. 5, 392-770.

Ehrenberg, Andrew S. C., Neil R. Barnard and Byron Sharp (2000), “Decision Models or Descriptive Models?,” *International Journal of Research in Marketing*, 17, 147-158.

Goodhardt, Ehrenberg and Chatfield (1984), “The Dirichlet: A Comprehensive

- Model of Buying Behaviour," *Journal of Royal Statistical Society*, 147, 621-655.
- Guadagni, Peter M. and John D. C. Little (1983), "A Logit Model of Brand Choice Calibrated on Scanner Data," *Marketing Science*, 2(3), 203-238.
- Jeuland, Abel P., Frank Bass and Gordon Wright (1980), "A Multibrand Stochastic Model Compounding Heterogeneous Erlang Timing and Multinomial Choice Processes," *Operations Research*, Vol. 28, No. 2, 255-277.
- Lenk, Peter J., Ambar G. Rao and Vikas Tibrewala (August 1993), "Nonstationary Conditional Trend Analysis: an Application to Scanner Panel Data," *Journal of Marketing Research*, Vol. 30, 288-304.
- Lilien, Kotler and Moorthy (1992), *Marketing Models*, Prentice-Hall.
- Schmittlein, David C. and Donald G. Morrison (1983), "Prediction of Future Random Events with the Condensed Negative Binomial Distribution," *Journal of the American Statistical Association*, Vol. 78, No. 382, 449-456.
- Wright, Malcolm, Anne Sharp and Byron Sharp (2002), "Market Statistics for the Dirichlet Model: Using the Juster Scale to Replace Panel Data," *International Journal of Research in Marketing*, 19, 81-90.

〈Abstract〉

Detecting Structural Change in NBD Model*

Joo, Young Jin *

In this research, we develop a procedure for detecting a random non-stationarity to the individual's purchasing rate in a stationary NBD model. On this purpose, we derive the likelihood ratio statistic for a testing null and alternative hypotheses defined as whether there is no significant structural change in a stationary NBD model or any. Where the structural change comes from a random non-stationarity (marketing mix activities or seasonality, for example) to the individual's purchasing rate. We also apply the developed method to a panel data for a frequently purchased good. This research could be a solution to include the non-stationarity in a stationary NBD model. We also expect that the developed model could give a signal for an early detection of significant changes in marketing environment, and a mean for a measurement of the effects of marketing mix activities.

Key Words: NBD(Negative Binomial Distribution) Model, Panel Data, Purchase Frequency, Structural Change

* This study was supported by the research grant of the Chungbuk National University in 2005. The author wishes to thank Prof. David R. Bell for the permission to use the data applied in this study.

** Chungbuk National University, Assistant Professor