

## Customer Base Analysis under Repeated Price Promotion in the Noncontractual Setting

김호, 김정기, 전덕빈

한국과학기술원 경영대학  
(137-722) 서울 동대문구 회기로 87  
Job75, gullung, dbjun@business.kaist.ac.kr,

### ABSTRACT

고객평생가치의 개념의 중요성이 증대함에 따라 오늘날의 경영자들은 고객의 미래 구매 행태를 예측하는 것에 큰 관심을 두고 있다. 특히 많은 기업이 데이터베이스의 발달로 고객의 구매 행태에 대한 다양한 정보는 축적하고 있으나 고객의 사멸에 대해서는 관측이 불가능한 현실에 직면해 있어, Pareto/NBD 모형, BG/NBD 모형과 같은 모형들의 유용성이 주목 받고 있다. 그러나 Fader et al., (2005)에서 언급하였듯이 상기의 모형을 현실에 적용함에 있어서 변화하는 마케팅환경을 반영하지 않으면 미래 고객의 구매행태를 예측하는데 오류를 범할 수 있다. 본 연구는 이러한 점에 착안하여 소매상이 반복적인 관측기간을 시행하는 경우에 고객의 구매패턴을 예측할 수 있는 모형을 개발한다. 그리고 이를 소매상의 실제 자료에 적용하여 소매상의 마케팅활동에 따라 고객의 구매패턴이 상이함을 확인한다. 특히 관측기간에는 고객의 구매율이 높은 집단과 낮은 집단으로 양분되는 현상을 발견하였는데 이는 관측기간에 낮은 구매율을 보이는 고객에 대해 타겟마케팅을 실시하여 구매빈도를 높일 수 있는 여지가 있음을 보여주고 있다. 또한 다양한 목적함수 하에서 관측의 효과를 극대화할 수 있는 관측기간의 최적 길이를 결정하는 문제에도 본 모형이 활용될 수 있음을 보인다.

### 1 Introduction

고객의 구매시점, 구매빈도, 그리고 구매금액에 대한 풍부한 자료가 구비되어감에 따라 고객의 과거 구매행태로부터 미래의 구매행태를 예측하고자 하는 연구가 자연스럽게 시도되었다. 특히 고객이 기업과 계약을 맺지 않고도 물건을 구매할 수 있는 상황 (noncontractual context)에서 고객의 미래 구매행태를 예측할 수 있는 유용한 연구들이 진행되어왔다. 예를 들어, Schmittlein et al. (1987)과 Fader et al. (2005)은 고객의 구매빈도 및 가장 최근 구매시점에 대한 자료만으로 특정 고객의 미래 일정기간의 구매빈도를 예측할 수 있는 모형을 개발하였고, Schmittlein and Peterson (1994)과 Colombo and Jiang (1999)은 구매빈도와 함께 구매금액을 예측할 수 있는 모형을 개발하였다. 이러한 류의 연구들은 마케팅에서 고객평생가치 (customer lifetime value: CLV) 개념이 강조되어 감에 따라 (예: Blattberg and Deighton, 1996; Gupta et al., 2004) 그 중요성이 지속적으로 증대되고 있다.

구매빈도와 관련된 연구의 시초는 Schmittlein, Morrison, and Colombo (1987) – 이하 SMC – 의

Pareto/NBD 모형으로, 이들은 고객의 사멸<sup>1</sup> (dropout, inactivity)이 관측되지 않는 상황에서 고객의 반복구매를 모델링하고자 하였다. 이를 위해 SMC는 고객이 일정기간 동안 일정한 비율 - 고객마다 다른 - 로 제품을 구매하고 일정한 비율 - 역시 고객마다 다른 - 로 사멸하는 과정을 거친다고 가정하였다. 구체적으로, 일정기간 동안 고객의 구매는 포아송 과정을 따르고 고객의 사멸시점은 지수분포를 따른다고 가정한 후 고객별로 상이한 구매율과 사멸율을 반영하기 위해 감마분포를 이용하였다. 이로부터 고객의 구매는 NBD 모형을 따르고 사멸은 Pareto (exponential-gamma mixture) 모형을 따르도록 유도하였다. 그러나 이 연구는 저자들도 지적하였듯이 모수 추정상의 어려움으로 인하여 실제 적용상의 한계에 직면한다.

한편, Fader et al. (2005)은 Pareto/NBD 모형의 모수 추정문제를 해결하기 위해 고객의 사멸과정에 대한 가정을 약간 수정하여 통상적으로 사용되는 최우추정법을 통해 쉽게 모수를 추정할 수 있도록 하였다. 즉, Pareto/NBD 모형은 고객의 사멸이 임의의 시점에서 발생할 수 있다고 가정하였으나 Fader et al. (2005)은 고객의 사멸이 제품 구매 직후에 일정한 확률 (사멸확률)로 발생한다고 가정하고 고객별로 상이한 사멸확률은 베타분포를 통해 반영하였다. 이로써 고객의 사멸과정에 대해 BG (beta-geometric) 모형을 유도하고 구매과정에 대해서는 NBD 모형을 적용하는 BG/NBD 모형을 개발하였는데, 이는 Microsoft Excel과 같은 통상적인 소프트웨어로도 모수를 추정할 수 있는 장점이 있다.

구매빈도와 함께 구매금액을 모델링한 대표적인 연구는 Schmittlein & Peterson (1994)과 Colombo & Jiang (1999)이 있다. Schmittlein & Peterson (1994)은 Pareto/NBD 모형을 이용하여 구매빈도를 모형화한 후 여기에 1회 구매당 구매금액을 추가적으로 모형화 하기 위해 정규분포 혼합 모형 (normal-normal mixture)을 이용한 반면, Colombo & Jiang (1999)은 구매금액을 모형화하기 위해 감마-감마 혼합 (gamma-gamma mixture) 모형을 이용하였다. 즉, 특정 고객의 1회 구매금액을 감마분포로 모형화하고 고객간의 구매금액의 이질성 (heterogeneity)을 다시 감마분포로 모형화 하였다.

지금까지 언급된 모형들이 고객의 구매 및 사멸과정에 대한 풍부한 행태적 논리를 바탕으로 전개되었으며 이 때문에 회귀분석 류의 모형(로지스틱 회귀분석, AID, CHAID, CART)과는 구분되는 장점이 존재한다 (Colombo & Jiang, 1999). 그러나 이 연구들은 실제 마케팅환경에서 보편적으로 발생하는 반복적인 가격할인기간의 효과를 반영하지 않은 한계가 존재한다<sup>2</sup>. 만약 가격할인기간 혹은 관측기간과 비관측기간 (정상영업기간)에서 고객의 구매행태 - 구매율, 사멸률 등 - 가 상이하다면 이를 반영하지 않은 모형으로 미래 구매행태를 예측한 결과는 편의를 유발시킬 수 있다. 특히 이러한 모형들의 주 활용분야 중 하나가 CLV 추정임을 고려할 때, 소매상의 관측에 따른 구매행태 변화를 반영하지 않는 상기의 모형들로 추정한 CLV에는 상당한 편의가 존재할 수 있을 것이다.

이에 본 연구는 가격할인과 정상영업이 반복되는 보편적인 상황에서 고객의 구매빈도 및

<sup>1</sup> 고객의 사멸 (dropout, becoming inactive)이란 고객이 특정 소매상으로부터 더 이상 그리고 영원히 제품을 구입하지 않는 상황을 말한다.

<sup>2</sup> 이러한 점은 Fader et al. (2005)이 BG/NBD 모형을 현실에 적용할 때 주의하여야 할 점으로 언급하기도 하였다.

구매금액을 분석하고 이 결과로부터 마케팅전략에 활용할 수 있는 시사점을 도출한다. 관측/비관측<sup>3</sup>의 상황에서 고객의 상이한 구매행태를 모형화하기 위해 BG/NBD 모형과 Colombo & Jiang의 구매금액분포모형 (monetary value distribution model, 이후로 CJ 모형이라고 명명)을 확장하였으며 이를 실제 자료에 적용하였다. 분석을 통해 우리는 다음과 같은 점을 발견할 수 있었다: 1) 관측기간과 비관측기간의 구매패턴 (구매빈도, 구매율, 사멸률, 및 구매금액)이 매우 상이하다. 2) 1)의 결과로 인해 제안 모형이 관측기간과 비관측기간의 주변 구매빈도 (marginal frequency of purchase) 및 조건부 구매빈도 (conditional frequency of purchase)를 훨씬 더 정확히 예측한다. 3) 관측기간과 비관측기간을 구분하지 않고 모형화할 경우 고객의 구매율의 이질성이 심각하게 과소 추정된다. 이는 향후 구매율의 이질성을 활용하여 개인수준의 마케팅 전략을 제시할 가능성을 사전에 차단하는 것이므로 문제가 된다. 4) 제품군에 따라 관측이 구매빈도 및 구매금액에 미치는 효과가 상이하였다. 예를 들어 스포츠웨어의 경우 관측기간에 구매빈도는 증가하고 1회 구매당 구매금액은 감소하였으나, 화장품의 경우 관측기간에 구매빈도가 오히려 감소하고 1회 구매당 구매금액은 증가하였다. 한편, 이와 같은 발견으로부터 본 연구의 결과가 개인 고객에 대한 타겟마케팅 (target marketing) 전략 및 제품군별 관측기간의 최적길이를 결정하는 문제에도 활용될 수 있음을 확인하였다.

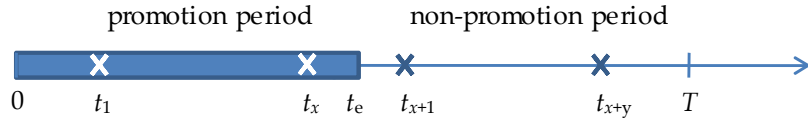
이를 위해 논문은 다음과 같이 구성되어 있다. 제 2절에서는 구매빈도 모형을 개발하고 제 3절에서는 구매금액 모형을 개발한다. 제 4절에서는 2절과 3절에서 개발된 모형을 실제 자료에 적용하는 실증분석을 실시한다 제 5절은 이 결과로부터 시사점을 도출한다. 제 6장은 본 연구의 결론이다.

## 2 Model Development for Purchase Frequency

### 2.1 Model Setup and Assumptions

본 연구에서는 1회의 관측기간과 1회의 비관측기간으로 구성된 기간을 기본 분석단위로 설정한다. 즉 총 관측기간은  $(0, T]$ 이며 이 중  $(0, t_e]$ 는 관측기간,  $(t_e, T]$ 는 비관측기간으로 정의하고 개인 고객의 관측기간 동안 구매 횟수를  $x$ , 비관측기간동안 구매횟수를  $y$ 로 표기한다. 또한 고객의  $j$ 번째 구매시점을  $t_j$ 로 표기한다 (<그림 1>).

<sup>3</sup> 본 연구에서는 다루는 관측이란 가격할인을 의미한다. 관측에는 가격할인 이외의 다양한 방법이 있으나 본 논문에서는 '가격할인'이라는 용어 대신 보편적인 '관측'이라는 용어를 이용한다.



<그림 1> 모형수립 상황

본 연구의 모형은 다음과 같은 가정에 기반해 있다.

(i) 직전 구매로부터 이번 구매까지의 시간 (inter-purchase time)은 구매율이  $\lambda$  인 지수분포를 따르며, 관측기간과 비관측기간의 구매율은 상이하다.

A.  $f(t_j | t_{j-1}, \lambda_1) = \lambda_1 e^{-\lambda_1(t_j - t_{j-1})}$ : 관측기간의 구매간격의 pdf

B.  $f(t_j | t_{j-1}, \lambda_0) = \lambda_0 e^{-\lambda_0(t_j - t_{j-1})}$ : 비관측기간의 구매간격의 pdf

(ii) 구매율  $\lambda_1, \lambda_0$  는 개인 고객별로 다르며 고객간 이질성 (heterogeneity)은 감마분포를 따른다.

$$f(\lambda_i | r_i, \alpha_i) = \frac{\alpha_i^{r_i} \lambda_i^{r_i-1} e^{-\alpha_i \lambda_i}}{\Gamma(r_i)}, \lambda_i > 0, i = 0, 1$$

(iii) 고객의 사멸은 오직 구매 직후에만  $p$ 의 확률로 발생한다. 따라서  $j$ 번 구매 후 사멸할 확률은 다음과 같다.

$$p_i (1 - p_i)^{j-1}, j = 1, 2, 3, \dots, i = 0, 1$$

또한 관측기간의 사멸확률( $p_1$ )과 비관측기간의 사멸확률( $p_0$ )은 다르다.

(iv) 개인 고객별  $p$ 는 상이하며 베타분포를 따른다. 즉,

$$f(p_i | a_i, b_i) = \frac{p_i^{a_i-1} (1-p_i)^{b_i-1}}{B(a_i, b_i)}, 0 \leq p_i \leq 1, i = 0, 1$$

(v) 구매율  $\lambda$  와 사멸확률  $p$ 는 고객끼리 서로 독립이다.

(vi) 분석을 시작하는 시점에서 데이터 상에 존재하는 고객은 사멸하지 않은 상태이다.

## 2.2 Individual Level Model

### 2.2.1 Likelihood function

관측기간의 구매수량  $x$ 와 비관측기간의 구매수량  $y$ 가 0인지 0보다 큰 값인지에 따라 [표 1]과 같은 4가지 경우가 가능하며 이에 대해 각각의 우도함수를 도출한다.

[표 1] x, y 값에 따른 경우의 분류

	x > 0	x = 0
y > 0	Case 1	Case 2
y = 0	Case 3	Case 4

Case 1) x>0, y>0

- $t_1$  에서 첫 구매를 할 우도:  $\lambda_1 e^{-\lambda_1 t_1}$
- $t_2$  에서 두번째 구매를 할 우도:  $(1-p_1)\lambda_1 e^{-\lambda_1(t_2-t_1)}$
- ...
- $t_x$  에서 x번째 구매를 할 우도:  $(1-p_1)\lambda_1 e^{-\lambda_1(t_x-t_{x-1})}$
- Non-promotion 기간 중  $t_{x+1}$  시점에서 구매할 우도:  $(1-p_1)e^{-\lambda_1(t_e-t_x)} \cdot \lambda_0 e^{-\lambda_0(t_{x+1}-t_e)}$
- Non-promotion 기간 중  $t_{x+1}$  시점에서 구매할 우도:  $(1-p_0)\lambda_0 e^{-\lambda_0(t_{x+2}-t_{x+1})}$
- ...
- $t_{x+y}$  시점에서 x+y 번째 구매를 할 우도:  $(1-p_0)\lambda_0 e^{-\lambda_0(t_{x+y}-t_{x+y-1})}$
- 마지막으로,  $(t_{x+y}, T]$  기간 동안 비구매 우도:  $p_0 + (1-p_0)e^{-\lambda_0(T-t_{x+y})}$

따라서 Case 1의 전체 우도는 위의 결과를 모두 곱하면

$$L(\lambda_1, \lambda_0, p_1, p_0 | t_1, t_2, \dots, t_{x+y}, t_e, T) = (1-p_1)^x (1-p_0)^{y-1} \lambda_1^x e^{-\lambda_1 t_e} \lambda_0^y e^{-\lambda_0(t_{x+y}-t_e)} \left( p_0 + (1-p_0)e^{-\lambda_0(T-t_{x+y})} \right) \quad (1)$$

유사한 방식으로 Case 2, Case 3, Case 4의 우도함수를 구하고 이들을 한 식으로 정리하면 다음과 같다. 각 경우별 도출과정은 부록 A.1을 참조하기 바란다.

$$L(\lambda_1, \lambda_0, p_1, p_0 | t_1, t_2, \dots, t_{x+y}, t_e, T) = e^{-\lambda_1(I_y t_e + (1-I_y)t_x + (1-I_x)(1-I_y)t_e)} \lambda_1^x (1-p_1)^{x-1+I_x I_y + (1-I_x)(1-I_y)} \times [I_y A + (1-I_y)B] \quad (2)$$

단,

$$A = (1-p_0)^{y-1} \lambda_0^y e^{-\lambda_0(t_{x+y}-t_e)} \left( p_0 + (1-p_0)e^{-\lambda_0(T-t_{x+y})} \right), \quad B = I_x \cdot (1-I_y)p_1 + \left[ (1-p_1)e^{-\lambda_1(t_e-t_x)} \right]^{I_x(1-I_y)} e^{-\lambda_0(T-t_e)}$$

$$I_x = \begin{cases} 1 & \text{if } x > 0 \\ 1 & \text{otherwise} \end{cases}, \quad I_y = \begin{cases} 1 & \text{if } y > 0 \\ 1 & \text{otherwise} \end{cases}$$

## 2.2.2 Derivation of $P(X(t_1) = x, Y(t_0) = y)$

관측기간을  $t_1$ , 비관측기간을  $t_0$ 로 하였을 때 1인의 고객이 관측기간 동안  $x$ 건을 구매하고 비관측기간 동안  $y$ 건을 구매할 확률  $P(X(t_1)=x, Y(t_0)=y)$  을 도출하면 다음과 같다. 앞의 우도함수를 도출할 때와 마찬가지로  $x, y$ 가 0인지 0보다 큰 값인지에 따라 4가지 경우로 나누어 분석한다. 이를 위해 관측/비관측에 상관없이 동질적인 한구간에서 구매건수에 대한 확률  $P(N(t)=n)$  을 먼저 도출한 후 이의 확장 형태로  $P(X(t_1)=x, Y(t_0)=y)$  를 도출하도록 한다.

먼저,  $t$  기간동안  $n$ 번의 구매가 발생하는 사상 (event)은  $n$ 번의 구매가 발생할 때까지 걸리는 시간  $T_n$  이  $t$ 보다 작다는 것과 수학적으로 동치임을 기억하자. 즉,  $N(t) \geq n \Leftrightarrow T_n \leq t$ . 그러면  $t$ 기간 동안  $n$ 번의 구매가 발생할 확률은 다음과 같다:

$$P(N(t)=n) = P(\text{active after } n\text{th purchase}) \cdot P(T_n \leq t, T_{n+1} > t) + I_n P(\text{become inactive after } n\text{th purchase}) \cdot P(T_n \leq t)$$

단,  $I_n$ 은  $n > 0$  이면 1,  $n = 0$  이면 0의 값을 갖는 변수이다.

구매간 시간간격 (inter-purchase time)이 지수분포를 따른다는 가정으로부터  $P(T_n \leq t, T_{n+1} > t)$ 는  $N(t)=n$  일 포아송 확률값이며  $P(T_n \leq t)$ 은 일랑- $n$  분포를 따른다. 따라서 위의 식은 다음과 같이 정리된다:

$$P(N(t)=n | \lambda, p) = (1-p)^n \frac{(\lambda t)^n e^{-\lambda t}}{n!} + I_n p (1-p)^{n-1} \cdot \left[ 1 - e^{-\lambda t} \sum_{j=0}^{n-1} \frac{(\lambda t)^j}{j!} \right] \quad (3)$$

이제 식 (3)이 도출된 과정을 엄두에 두고  $P(X(t_1)=x, Y(t_0)=y)$  를 각 경우별로 도출한 후 한 식으로 정리하면 다음과 같다. (Case 별 구체적인 도출과정은 부록 A.2 참조).

$$P(X(t_1)=x, Y(t_0)=y | \lambda_1, p_1, \lambda_0, p_0) = \left[ (1-p_1)^x \frac{e^{-\lambda_1 t_1} (\lambda_1 t_1)^x}{x!} (e^{-\lambda_0 t_0})^k + A_1 - k \cdot (A_1 - A_2) \right] \times B^{1-k} \quad (4)$$

단,  $k = I_x(1 - I_y)$  로 case 3를 의미,

$$A_1 = p_0(1-p_0)^{y-1} \cdot \left( 1 - e^{-\lambda_0 t_0} \sum_{j=0}^{y-1} \frac{(\lambda_0 t_0)^j}{j!} \right), A_2 = p_1(1-p_1)^{x-1} \cdot \left( 1 - e^{-\lambda_1 t_1} \sum_{j=0}^{x-1} \frac{(\lambda_1 t_1)^j}{j!} \right), B = \left[ (1-p_0)^y \frac{e^{-\lambda_0 t_0} (\lambda_0 t_0)^y}{y!} + I_y A_1 \right]$$

### 2.2.3 Derivation of $E[X(t_1)], E[Y(t_0)]$

관측기간의 평균구매건수  $E[X(t_1)]$  와 비관측기간의 평균구매건수  $E[Y(t_0)]$  는 다음과 같이 도출된다. 우선  $\tau$  를 사멸시점이라고 하면 관측기간 동안의 임의의 시점  $t$  에서 사멸하지 않을 확률은  $P(\tau > t) = P(\text{active at } t | \lambda_1, p_1) = \sum_{j=0}^{\infty} (1-p_1)^j \frac{(\lambda_1 t)^j e^{-\lambda_1 t}}{j!} = e^{-\lambda_1 p_1 t}$  이 되며, 따라서 임의의 시점  $t$  에서 사멸할 확률은  $P(\tau \leq t) = 1 - P(\tau > t) = 1 - e^{-\lambda_1 p_1 t}$  가 되고, 사멸시점에 대한 확률밀도함수는  $g(\tau | \lambda_1, p_1) = \lambda_1 p_1 e^{-\lambda_1 p_1 \tau}$  로 도출된다. 따라서 관측기간  $t_1$  동안 기대구매빈도  $E[X(t_1)]$  는 다음과 같다.

$$E[X(t_1) | \lambda_1, p_1] = \lambda_1 t_1 P(\tau > t_1) + \int_0^{t_1} \lambda_1 \tau g(\tau | \lambda_1, p_1) d\tau = \frac{1}{p_1} [1 - e^{-\lambda_1 p_1 t_1}] \quad (5)$$

한편 비관측기간  $t_0$  동안의 기대 구매빈도  $E[Y(t_0)]$  는 관측이 끝나는 시점  $t_e$  에서 이 고객이 사멸하였는지 아닌지에 따라 다르게 될 것이다. 만약  $t_e$  시점에서 사멸하지 않았다면 조건부 기대값  $E[Y(t_0) | \text{active at } t_e]$  는  $E[X(t_1)]$  과 유사하게 다음과 같다.

$$E[Y(t_0) | \lambda_0, p_0, \text{active at } t_e] = \frac{1}{p_0} [1 - e^{-\lambda_0 p_0 t_0}] \quad (6)$$

만약  $t_e$  시점에서 이 고객이 이미 사멸하였다면  $E[Y(t_0) | \lambda_0, p_0, \text{inactive at } t_e] = 0$  이 된다. 그런데  $t_e$  시점에서 사멸하지 않을 확률은  $P(\tau > t_e) = e^{-\lambda_1 p_1 t_e}$  , 사멸할 확률은  $P(\tau \leq t_e) = 1 - e^{-\lambda_1 p_1 t_e}$  이므로 비관측기간의 평균구매건수는 다음과 같다.

$$E[Y(t_0) | \lambda_0, p_0] = \frac{1}{p_0} [1 - e^{-\lambda_0 p_0 t_0}] \times e^{-\lambda_1 p_1 t_e} + 0 \times (1 - e^{-\lambda_1 p_1 t_e}) = \frac{1}{p_0} [1 - e^{-\lambda_0 p_0 t_0}] \times e^{-\lambda_1 p_1 t_e} \quad (7)$$

### 2.3 Model Development for a Randomly Chosen Individual

위의 개인 수준에서 도출된 확률과 기대 구매빈도는 특정 개인의 구매율  $(\lambda_1, \lambda_0)$  과 사멸확률  $(p_1, p_0)$  에 기반하여 도출되었다. 그러나 실제 자료로 분석할 때는 개인의 구매율과 사멸확률을

알 수 없으므로 이들에 대해 특정한 분포를 가정한 후 이 분포에 대해 위에서 도출된 확률과 기대 구매빈도의 기대값을 구한다. 앞선 가정에 의해  $\lambda_0, \lambda_1, p_0, p_1$ 는 다음과 같은 분포를 갖는다.

$$f(\lambda_0 | r_0, \alpha_0) = \frac{\alpha_0^{r_0}}{\Gamma(r_0)} e^{-\alpha_0 \lambda_0} \lambda_0^{r_0-1} \quad (8-1)$$

$$f(\lambda_1 | r_1, \alpha_1) = \frac{\alpha_1^{r_1}}{\Gamma(r_1)} e^{-\alpha_1 \lambda_1} \lambda_1^{r_1-1} \quad (8-2)$$

$$f(p_0 | a_0, b_0) = \frac{1}{B(a_0, b_0)} p_0^{a_0-1} (1-p_0)^{b_0-1} \quad (8-3)$$

$$f(p_1 | a_1, b_1) = \frac{1}{B(a_1, b_1)} p_1^{a_1-1} (1-p_1)^{b_1-1} \quad (8-4)$$

이 분포를 이용해 제 4장에서 도출된 우도함수, 확률 및 기대 구매빈도를 적분하면 다음과 같은 결과를 도출할 수 있다.

● Likelihood function  $L(r_0, r_1, \alpha_0, \alpha_1, a_0, b_0, a_1, b_1 | x, y, t_x, t_e, t_{x+y}, T)$

Case 1)  $x > 0, y > 0$

$$L(r_0, r_1, \alpha_0, \alpha_1, a_0, b_0, a_1, b_1 | x, y, t_x, t_e, t_{x+y}, T) = \frac{\alpha_1^{r_1} \alpha_0^{r_0}}{\Gamma(r_1) \Gamma(r_0)} \frac{\Gamma(x+r_1) \Gamma(y+r_0)}{B(a_1, b_1) B(a_0, b_0)} \frac{B(a_1, x+b_1)}{(t_e + \alpha_1)^{x+r_1}} \left[ \frac{B(a_0+1, y+b_0-1)}{(t_{x+y} - t_e + \alpha_0)^{y+r_0}} + \frac{B(a_0, y+b_0)}{(T - t_e + \alpha_0)^{y+r_0}} \right] \quad (2-1)$$

Case 2)  $x=0, y > 0$

$$L(r_0, r_1, \alpha_0, \alpha_1, a_0, b_0, a_1, b_1 | x, y, t_x, t_e, t_{x+y}, T) = \frac{\alpha_1^{r_1} \alpha_0^{r_0} \Gamma(y+r_0)}{\Gamma(r_0) B(a_0, b_0) (t_e + \alpha_1)^{r_1}} \left[ \frac{B(a_0+1, y+b_0-1)}{(t_{x+y} - t_e + \alpha_0)^{y+r_0}} + \frac{B(a_0, y+b_0)}{(T - t_e + \alpha_0)^{y+r_0}} \right] \quad (2-2)$$

Case 3)  $x > 0, y=0$

$$L(r_0, r_1, \alpha_0, \alpha_1, a_0, b_0, a_1, b_1 | x, y, t_x, t_e, t_{x+y}, T) = \frac{\alpha_1^{r_1} \alpha_0^{r_0}}{\Gamma(r_1)} \frac{\Gamma(x+r_1)}{B(a_1, b_1)} \left[ \frac{B(a_1+1, x+b_1-1)}{\alpha_0^{r_0} (t_x + \alpha_1)^{x+r_1}} + \frac{B(a_1, x+b_1)}{(T - t_e + \alpha_0)^{r_0} (t_e + \alpha_1)^{x+r_1}} \right] \quad (2-3)$$

Case 4)  $x = y = 0$

$$L(r_0, r_1, \alpha_0, \alpha_1, a_0, b_0, a_1, b_1 | x, y, t_x, t_e, t_{x+y}, T) = \left( \frac{\alpha_1}{t_e + \alpha_1} \right)^{r_1} \left( \frac{\alpha_0}{T - t_e + \alpha_0} \right)^{r_0} \quad (2-4)$$

●  $P(X(t_1) = x, Y(t_0) = y | \alpha_0, r_0, \alpha_1, r_1, a_0, b_0, a_1, b_1)$

Case 1, 2, 4)



$$\begin{aligned}
 & P(X(t_1)=x, Y(t_0)=y | \alpha_0, r_0, \alpha_1, r_1, a_0, b_0, a_1, b_1) \\
 &= \left[ \frac{B(a_0, b_0 + y) \Gamma(r_0 + y)}{B(a_0, b_0) \Gamma(r_0) y!} \left( \frac{\alpha_0}{\alpha_0 + t_0} \right)^{r_0} \left( \frac{t_0}{\alpha_0 + t_0} \right)^y + \delta_{y>0} \frac{B(a_0 + 1, b_0 + y - 1)}{B(a_0, b_0)} \left\{ 1 - \left( \frac{\alpha_0}{\alpha_0 + t_0} \right)^{r_0} \sum_{j=0}^{y-1} \frac{\Gamma(r_0 + j)}{\Gamma(r_0) j!} \left( \frac{t_0}{\alpha_0 + t_0} \right)^j \right\} \right] \quad (4-1)
 \end{aligned}$$

Case 3)

$$\begin{aligned}
 & P(X(t_1)=x, Y(t_0)=0 | \alpha_0, r_0, \alpha_1, r_1, a_0, b_0, a_1, b_1) \\
 &= \frac{B(a_1, b_1 + x) \Gamma(x + r_1)}{B(a_1, b_1) \Gamma(r_1) x!} \left( \frac{\alpha_1}{\alpha_1 + t_1} \right)^{r_1} \left( \frac{t_1}{\alpha_1 + t_1} \right)^x \left( \frac{\alpha_0}{\alpha_0 + t_0} \right)^{r_0} + \frac{B(a_1 + 1, b_1 + x - 1)}{B(a_1, b_1)} \left[ 1 - \left( \frac{\alpha_1}{\alpha_1 + t_1} \right)^{r_1} \sum_{j=1}^{x-1} \frac{\Gamma(r_1 + j)}{\Gamma(r_1) j!} \left( \frac{t_1}{\alpha_1 + t_1} \right)^j \right] \quad (4-2)
 \end{aligned}$$

- $E[X(t_1) | \alpha_0, r_0, \alpha_1, r_1, a_0, b_0, a_1, b_1], E[Y(t_0) | \alpha_0, r_0, \alpha_1, r_1, a_0, b_0, a_1, b_1]$

관측기간  $t_1$  동안의 평균구매빈도는 다음과 같다.

$$E[X(t_1) | r_1, \alpha_1, a_1, b_1] = \frac{B(a_1 - 1, b_1)}{B(a_1, b_1)} \left[ 1 - \left( \frac{\alpha_1}{\alpha_1 + t_1} \right)^{r_1} {}_2F_1 \left( r_1, b_1; a_1 + b_1 - 1; \frac{t_1}{\alpha_1 + t_1} \right) \right] \quad (5-1)$$

비관측기간  $t_0$  동안의 평균구매빈도는 다음과 같다.

$$\begin{aligned}
 & E[Y(t_0) | r_1, \alpha_1, a_1, b_1, r_0, \alpha_0, a_0, b_0, t_1] \\
 &= \frac{B(a_0 - 1, b_0)}{B(a_0, b_0)} \left[ 1 - \left( \frac{\alpha_0}{\alpha_0 + t_0} \right)^{r_0} {}_2F_1 \left( r_0, b_0; a_0 + b_0 - 1; \frac{t_0}{\alpha_0 + t_0} \right) \right] \times \frac{1}{B(a_1, b_1)} \left( \frac{\alpha_1}{\alpha_1 + t_1} \right)^{r_1} {}_2F_1 \left( r_1, b_1; a_1 + b_1; \frac{t_1}{\alpha_1 + t_1} \right) \quad (7-1)
 \end{aligned}$$

단,  ${}_2F_1(a, b; c; z) = \frac{1}{B(b, c-b)} \int_0^1 t^{b-1} (1-t)^{c-b-1} (1-zt)^{-a} dt, (c > b)$  로써 가우스 초기하함수(Gaussian hypergeometric function)을 의미한다.

### 3 Model Development for Purchase Amount

먼저 개인고객 수준에서 1회 구매건 당 구매금액을 모델링하기 위해 구매건 당 구매금액이 감마분포를 따르는 것으로 가정한다. 한편, 관측여부에 따라 고객의 1회 구매 건 당 구매금액 분포에 차이가 있을 수 있으므로 이를 반영하기 위해 감마분포의 scale 모수를 관측여부에 따라 조정할 수 있도록 감마분포를 식 (9)와 같이 수정한다.

$$f(z | u, \theta) = \frac{(\theta \exp(x\beta))^u}{\Gamma(u)} z^{u-1} e^{-(\theta \exp(x\beta))z} \quad x = 1 \text{ in promotion, } 0 \text{ in nonpromotion} \quad (9)$$

여기서  $\beta$  는 관측효과에 따라 구매금액의 scale이 변화되는 정도를 측정한다. 그러나 식 (9)의 감마분포는 고객에 따라 상이해야 한다. 즉, 어떤 고객은 1회 구매 건당 구매금액의 분포가 전반적으로 다른 고객에 비해 높은 수준에서 형성되거나 그 반대의 경우가 발생할 것이다. 따라서 이러한 이질성이 반영되어야 하는데 이를 위해 식 (9)의 scale 모수  $\theta$ 가 고객별로 상이할 것으로 가정하고  $\theta$ 가 다음과 같은 감마분포를 따르는 것으로 가정한다.

$$f(\theta | v, \phi) = \frac{\phi^v}{\Gamma(v)} \theta^{v-1} e^{-\phi\theta} \quad (10)$$

따라서, 식 (9)를 식 (10)에 대해 적분하면 임의로 추출된 개인고객의 1회 구매건 당 구매금액의 분포를 도출할 수 있으며 그 결과는 다음과 같다.

$$f(z | x, u, v, \phi, \beta) = \frac{\Gamma(u+v)}{\Gamma(u)\Gamma(v)} \left( \frac{z \cdot \exp(x\beta)}{\phi + z \cdot \exp(x\beta)} \right)^u \left( \frac{\phi}{\phi + z \cdot \exp(x\beta)} \right)^v \frac{1}{z} \quad (11)$$

또, 위의 식 (11)로부터 1회 구매건당 기대구매금액은 다음과 같이 도출된다.

$$E(z | x, u, v, \phi, \beta) = \frac{u \cdot \phi}{(v-1) \cdot e^{x\beta}} \quad (12)$$

#### 4 Empirical Analysis

실증분석을 위해 국내 모 백화점의 판매 자료를 이용하였다. 판매일수는 총 82일이며 이 중 첫 10일은 정기바겐세일 기간으로 관측기간에 해당하고 나머지 72일은 정상적인 영업일, 즉 비관측기간에 해당한다. 분석에 사용된 제품군은 스포츠웨어, 시티케주얼, 그리고 화장품이다. 자료는 동 기간 동안 적어도 1회 이상 구매한 고객들의 구매건수, 구매시점, 매 구매당 구매금액으로 구성되어 있는데, 스포츠웨어는 16627명의 고객이 25584건을 구매하였고, 시티케주얼은 14377명의 고객이 21623건의 구매를 하였다. 또 화장품은 총 22367명이 36276건을 구매하였다. [표 2] ~ [표 4]는 관측효과를 반영하지 않은 기본모형과 본 연구에서 제안하는 모형으로 각 제품군의 구매빈도와 구매금액을 추정한 결과이다. 시티케주얼의  $b_1$ 을 제외한 모든 추정치가 1% 유의수준에서 유의한 것으로 분석되었다.

[표 2] 스포츠웨어 추정 결과

Purchase Frequency (N=16627 Customers)			Purchase Amount (N=25584 Transactions)		
Parameters	Base*	Proposed	Parameters	Base*	Proposed
$r$	1042.0		$u$	3.3135	3.4549
$\alpha$	40090.7		$v$	3.2746	3.3612
$a$	3.6930		$\phi$	49602.67	54024.88
$b$	7.119		$\beta$		0.3610
$r_0$		52.3368			
$\alpha_0$		1458.26			
$a_0$		2.4811			
$b_0$		1.6785			
$r_1$		0.5075			
$\alpha_1$		10.5066			
$a_1$		2.1242			
$b_1$		1.6884			
LL	-270074	-119534.5	LL***	-241406	-241014

\* 구매빈도에 대한 Base Model은 BG/NBD 모형이며 구매금액에 대한 Base Model은 Colombo & Jiang (1999)의 monetary value distribution model을 이용하였다.

[표 3] 시티케주얼 추정결과

Purchase Frequency (N=14377 Customers)			Purchase Amount (N=21623 Transactions)		
Parameters	Base*	Proposed	Parameters	Base*	Proposed
$r$	1262.82		$u$	4.6348	4.6265
$\alpha$	52010.45		$v$	3.5226	3.5332
$a$	3467.54		$\phi$	73798.97	73394.19
$b$	6966.03		$\beta$		-0.0621
$r_0$		73.6851			
$\alpha_0$		2266.52			
$a_0$		2.0049			
$b_0$		1.7512			
$r_1$		0.2984			
$\alpha_1$		9.5134			
$a_1$		1.4289			
$b_1$		1.0000			

LL	-231603.4	-103736.1	LL***	-215133.3	-215123.1
----	-----------	-----------	-------	-----------	-----------

\* 구매빈도에 대한 Base Model은 BG/NBD 모형이며 구매금액에 대한 Base Model은 Colombo & Jiang (1999)의 monetary value distribution model을 이용하였다.

[표 4] 화장품 추정결과

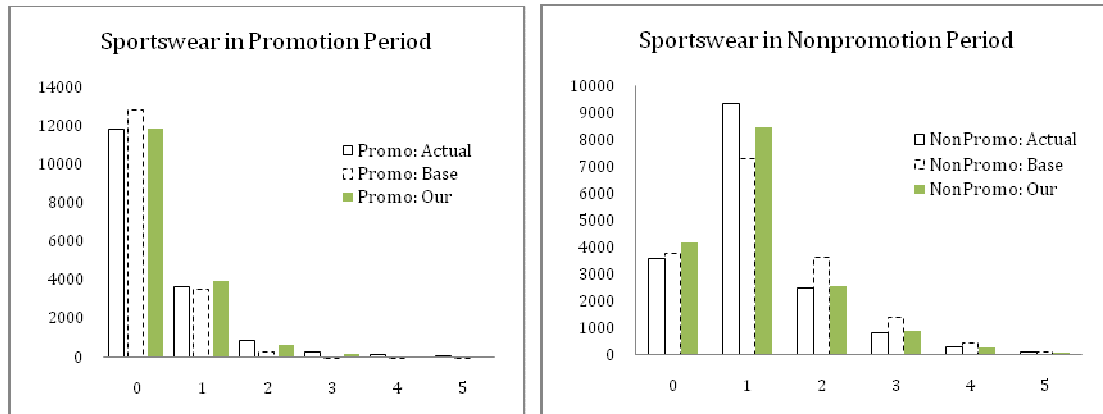
Purchase Frequency (N=22367 Customers)			Purchase Amount (N=36276 Transactions)		
Parameters	Base*	Proposed	Parameters	Base*	Proposed
$r$	1513.23		$u$	2.2678	2.2655
$\alpha$	58098.78		$v$	3.6517	3.6594
$a$	3.6932		$\phi$	112296.7	111870.8
$b$	7.6599		$\beta$		-0.0458
$r_0$		75.2346			
$\alpha_0$		2124.90			
$a_0$		1.5488			
$b_0$		1.4130			
$r_1$		0.6663			
$\alpha_1$		27.8107			
$a_1$		0.9964			
$b_1$		1.1145			
LL	-381905.3	-172643.7	LL***	-372006	-371997

\* 구매빈도에 대한 Base Model은 BG/NBD 모형이며 구매금액에 대한 Base Model은 Colombo & Jiang (1999)의 monetary value distribution model을 이용하였다.

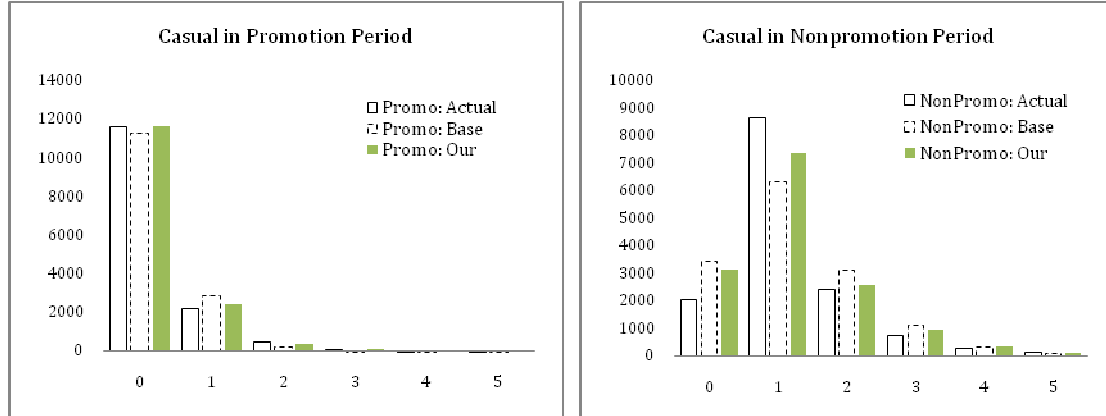
위의 추정결과로부터 다음을 확인할 수 있다: 1) 모든 제품군에서 제안모형이 구매빈도와 구매금액 데이터를 더 잘 적합시키고 있다. 이는 관측기간과 비관측기간의 고객의 구매행태가 상이함을 의미한다. 2) 구매빈도에 대한 모수 중 shape 모수인  $r, r_0, r_1$ 은 구매비율  $\lambda$ 의 동질성 (homogeneity)의 정도를 나타내는데 관측/비관측기간을 구분하지 않고 추정한 기존 모형에서는 구매비율의 동질성을 크게 과대 추정하는 것으로 나타났다. 또한 제안모형에서 관측기간과 비관측기간의 구매비율의 동질성을 비교할 때 비관측기간의 동질성이 관측기간의 동질성에 비해 매우 다는 것을 알 수 있다. 즉, 비관측기간에 구매비율 관점에서 상대적으로 동질적인 고객도 관측의 영향을 받아 구매비율이 매우 달라질 수 있음을 보여준다. 3)  $\beta$ 는 관측/비관측기간의 1회 구매당 기대 구매금액의 상대적인 크기를 나타내는데,  $\beta > 0$ 이면 관측기간의 1회 구매당 기대 구매금액이 비관측기간의 그것보다 더 작음을 의미한다. 스포츠웨어의 1회 구매당 기대 구매금액은 관측기간에 감소하는 반면, 시티케주얼과 화장품의 기대 구매금액은 오히려 증가하고 있다. 4) 스포츠웨어의 기대 구매비율은 관측기간에 증가하는 반면, 시티케주얼과 화장품의 기대 구매비율은 관측기간에 오히려 감소하고 있다. 3)의 결과와 함께 비교하면 스포츠웨어는

관측기간에 더 자주 구매하는 반면 1회 구매당 구매금액은 감소한다. 반면, 시티캐주얼과 화장품은 관측기간에 1회 구매당 구매금액이 증가하나 구매 횟수는 줄어들어서 제품군별로 고객이 서로 다른 생태를 보이고 있음을 알 수 있다.

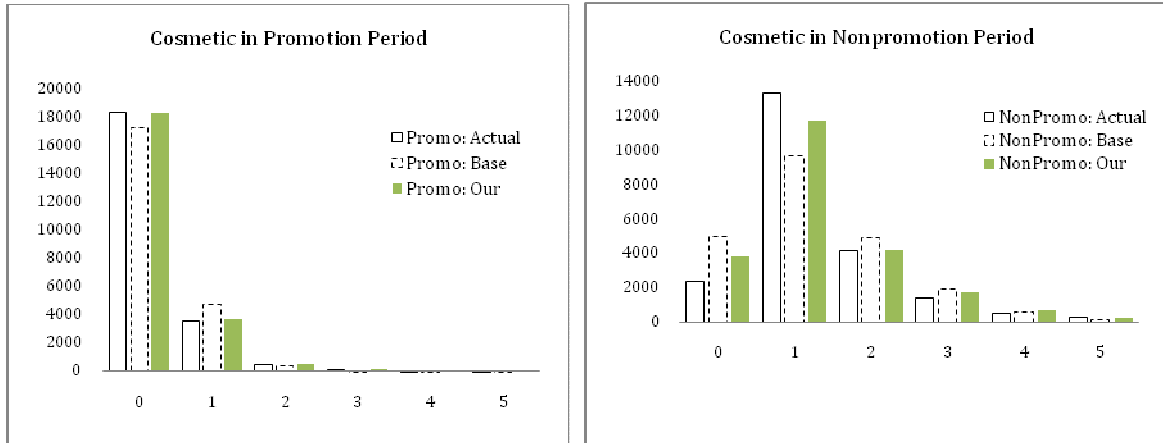
<그림 2> ~ <그림 4>는 식 (9-1)과 (7-1)에 위의 추정된 모수값을 적용하여 세 제품군의 관측기간과 비관측기간의 구매빈도를 예측한 것으로, 세 제품군 모두 제안모형이 BG/NBD 모형에 비해 예측력이 우수함을 보여준다.



<그림 2> 스포츠웨어의 관측기간과 비관측기간의 구매빈도 예측결과

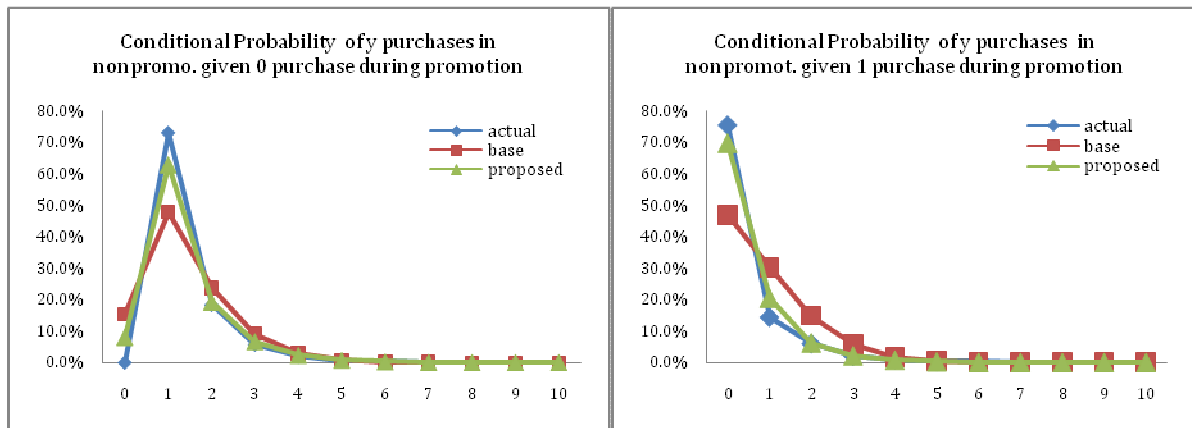


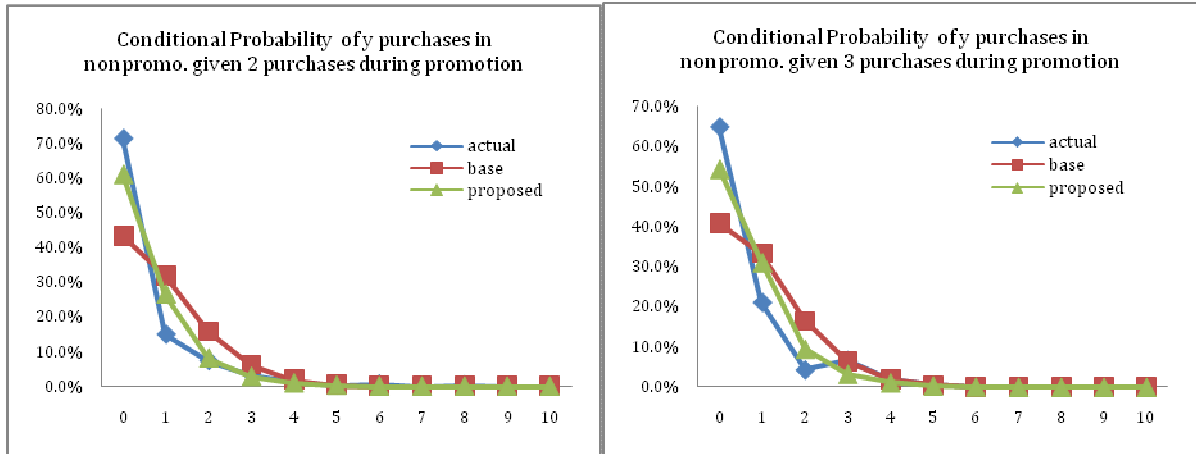
<그림 3> 시티캐주얼의 관측기간과 비관측기간의 구매빈도 예측결과



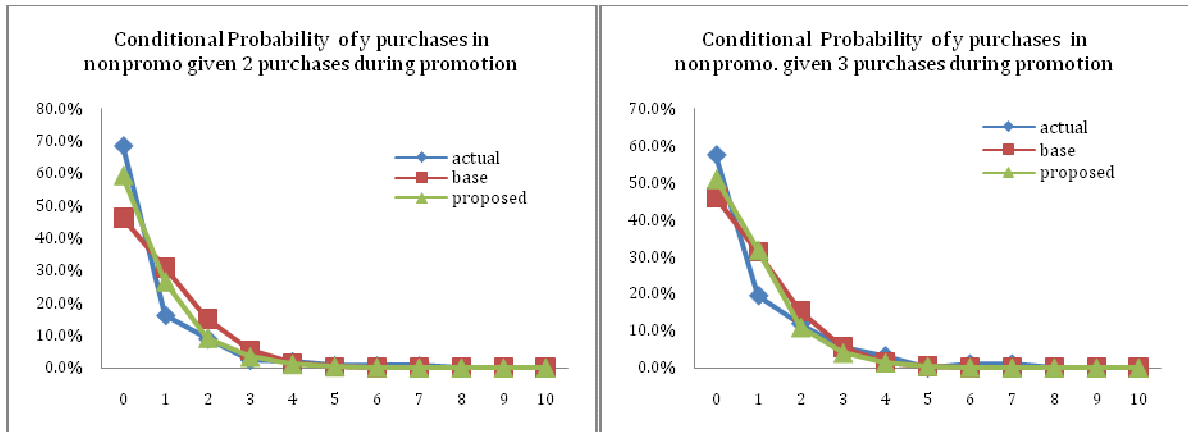
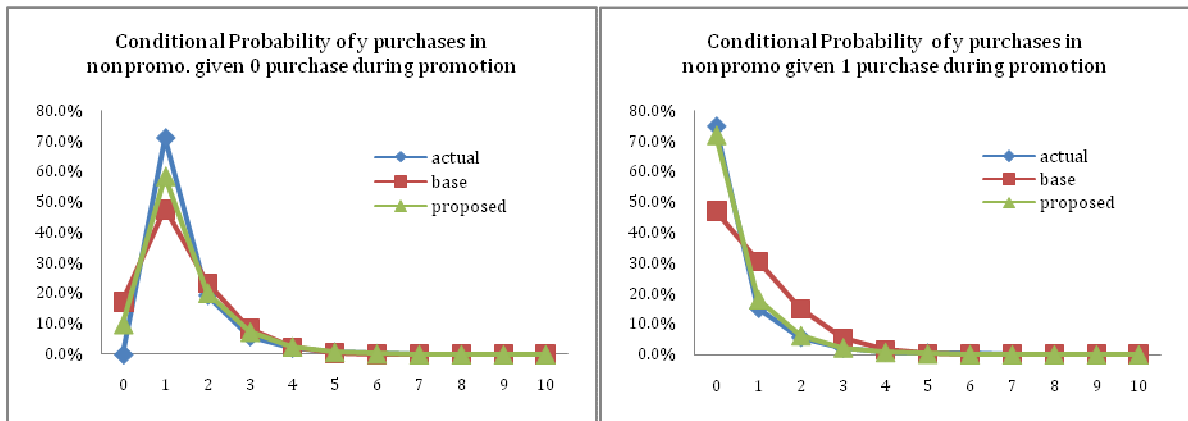
<그림 4> 화장품의 관측기간과 비관측기간의 구매빈도 예측결과

또한 <그림 5> ~ <그림 7>은 각 제품군의 비관측기간에서의 조건부확률분포로서, 이는 식 (9-1)과 (7-1)의 결합확률분포로부터 관측기간의 구매건수가  $x$ 건이라는 조건하에 도출하였으며, 조건부확률도 제안모형이 기존모형보다 더 우수함을 보여주고 있다. 특히, 관측기간에서 구매횟수의 대부분을 차지하는 0회 구매와 1회 구매에서 (<그림 2> ~ <그림 4> 참조) 제안모형이 기존 모형보다 월등함을 보여주고 있다. 이 결과는 특정 고객의 관측기간 동안의 구매건수를 알고 있을 때 이 고객의 비관측기간 동안의 구매건수의 확률 및 기대구매건수를 도출하는데 활용할 수 있다. 만약, 비관측기간이 관측기간에 우선한 자료를 분석하게 된다면 특정고객의 비관측기간에 구매건수에 따른 이 고객의 관측기간의 기대구매건수를 계산하는데 이 결과를 활용할 수 있다.

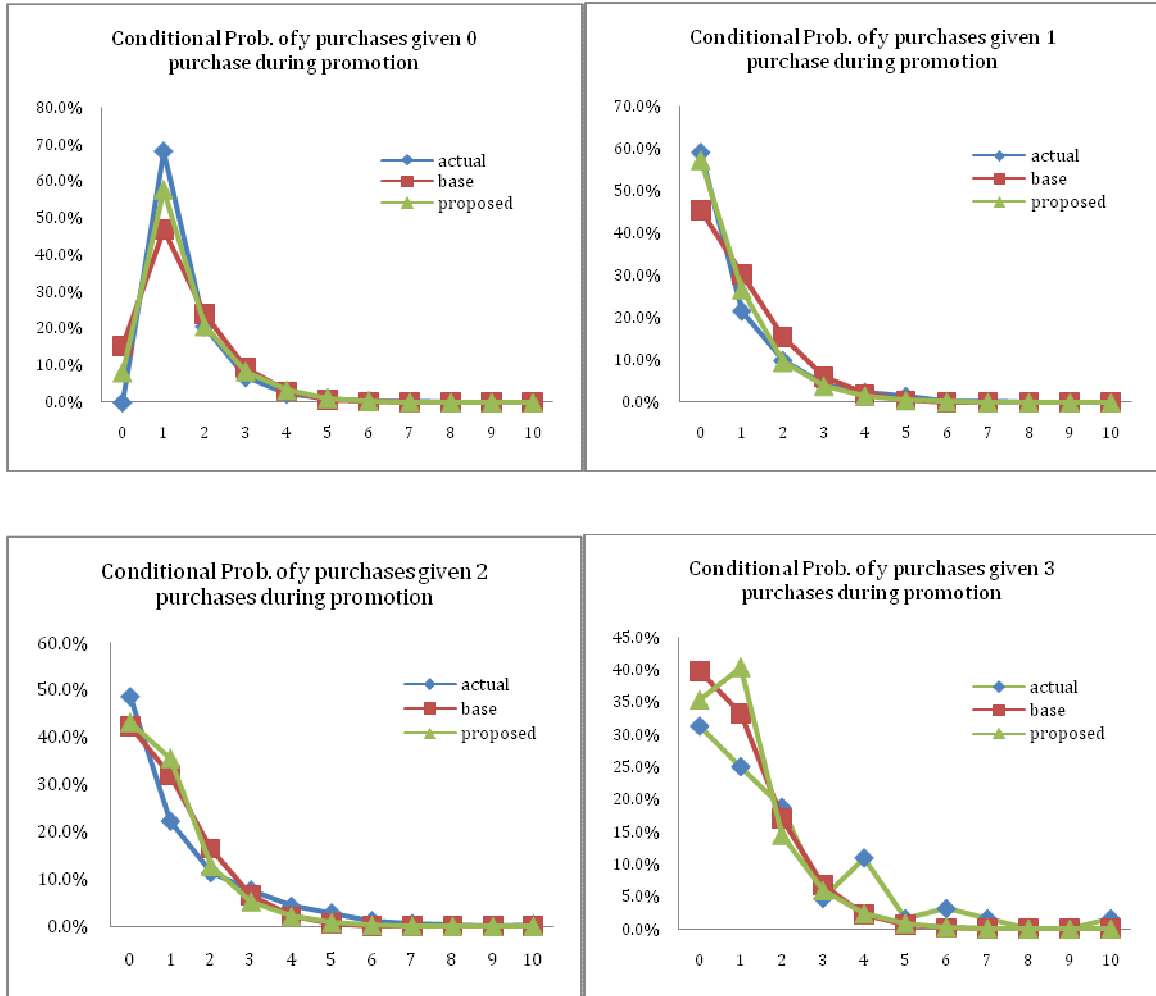




<그림 5> 관측기간 구매빈도에 기반한 비관측기간의 스포츠웨어 구매빈도 조건부 확률

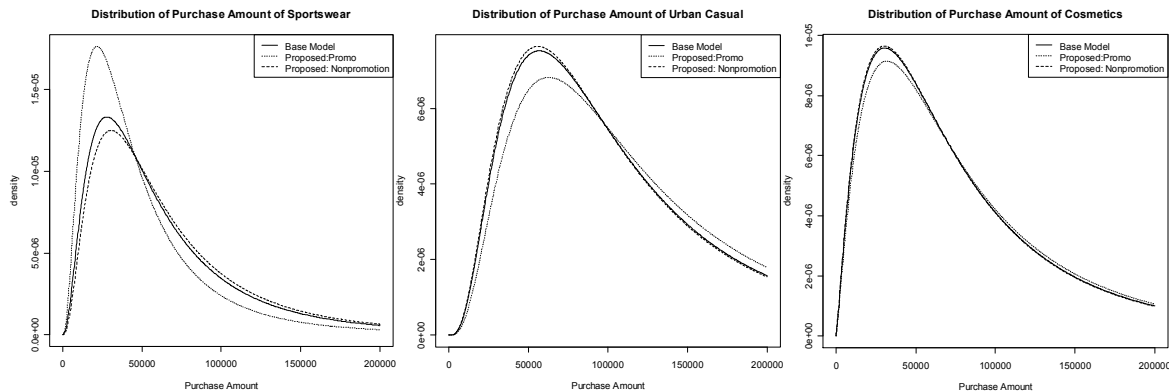


<그림 6> 관측기간 구매빈도에 기반한 비관측기간의 시티캐주얼 구매빈도 조건부 확률



<그림 7> 판촉기간 구매빈도에 기반한 비판촉기간의 화장품 구매빈도 조건부 확률

<그림 8>은 세 제품군의 판촉기간과 비판촉기간의 1회 구매당 구매금액의 분포를 보여준다. 스포츠웨어는 판촉기간 동안의 1회 구매금액이 비판촉기간 동안의 1회 구매금액보다 평균적으로 적은 것을 알 수 있으며, 시티캐주얼과 화장품은 반대로 판촉기간의 1회 구매금액이 비판촉기간의 1회 구매금액 보다 큰 것을 알 수 있다.



<그림 8> 판촉기간과 비판촉기간의 1회 구매당 구매금액의 분포



[표 5]는 식 (5-1)과 식 (12)으로 계산한 각 제품군의 관측기간 및 비관측기간의 기대 구매빈도 및 기대구매금액을 보여주고 있다. 이 때 관측기간과 비관측기간의 길이는 모두 10일로 설정하였다<sup>4</sup>. 이 표에 따르면 스포츠웨어의 경우 기대구매건수는 관측기간이 비관측기간에 비해 큰 것으로 나타났으나 1회 구매건 당 구매액은 비관측기간이 더 큰 것으로 나타났다. 반면, 시티캐주얼과 화장품의 경우 관측기간의 기대구매건수가 비관측기간의 기대구매건수에 비해 작으나 1회 구매당 구매액은 관측기간이 더 큰 것으로 나타났다. 즉, 소매상의 관측행위가 제품군에 따라 다른 구매행태를 유발시키는 것으로 보인다: 스포츠웨어의 관측은 기대구매건수를 증가시키나 1회 구매당 기대구매액은 감소시키는 작용을 한다. 반면 시티캐주얼과 화장품의 관측은 1회 구매당 구매액은 증가시키나 구매건수를 감소시키는 작용을 한다.

[표 5] 관측/비관측기간(t=10)의 고객 1인의 기대 매출

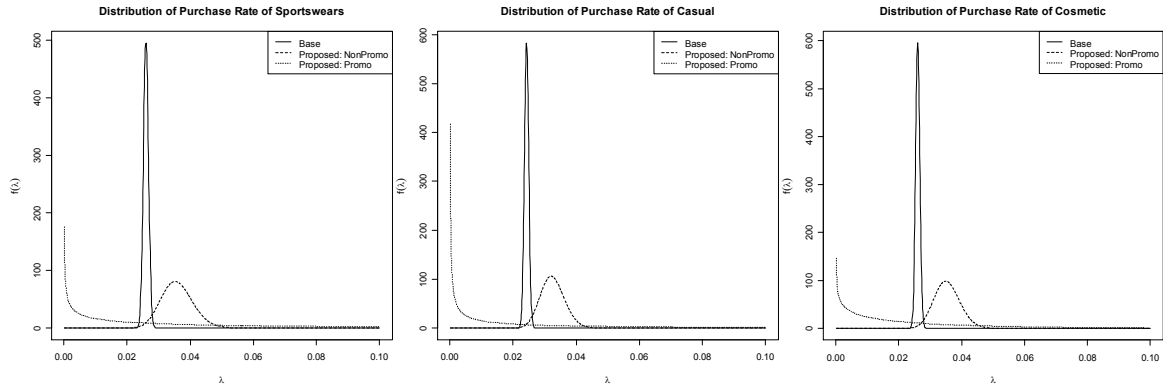
		스포츠웨어	시티캐주얼	화장품
관측기간 (t=10)	기대 구매 건수 ( $f_1$ )	0.3534	0.2315	0.2113
	1회 기대구매액 ( $A_1$ )	55,095	142,631	99,767
	기대 매출 ( $A_1 * f_1$ )	19,471	33,019	21,081
비관측기간 (t=10)	기대 구매 건수 ( $f_0$ )	0.3228	0.2984	0.3233
	1회 기대 구매액 ( $A_0$ )	79,049	134,043	95,300
	기대 매출 ( $A_0 * f_0$ )	25,517	39,998	30,811

## 5 Implications

### 5.1 Marketing Strategies for Individual Customers

고객의 이질성을 이해하는 것은 보다 정확한 예측을 가능하게 할 뿐 아니라 창의적 마케팅 전략을 위한 소중한 아이디어의 원천이 된다 (DeSarbo & Ramaswamy, 1994). 이러한 이유로 인해 마케팅분야에서는 관측되지 않은 고객의 이질성에 대한 연구가 활발히 진행되고 있으며 (예: Kamakura & Russell, 1989; Allenby et al., 1998; Allenby & Rossi, 1999) 이미 고객의 이질성을 통해 개별 고객에 대한 마케팅전략적 시사점을 얻고자 하는 것은 이미 보편적인 연구주제로 자리잡아 가고 있다. 본 연구에서는 구매율, 사멸률, 그리고 구매금액에 대해 고객간에 관측되지 않은 이질성을 모형화 하였다 (식 (8-1) ~ (8-4), (10)). [표 2] ~ [표 4]는 이 추정 결과로서 이를 식 (8-1) ~ (4-4)에 적용하면 각 제품군의 관측기간과 비관측기간 동안의 고객의 구매비율 및 사멸확률에 대한 확률분포를 <그림 9>와 같이 도출할 수 있다.

<sup>4</sup> 1회 구매당 기대 구매액은 t에 관계없이 일정하나 구매건수는 t에 따라 가변적이다. 이로부터 매출(혹은 수익이나 구매빈도)를 극대화하는 관측기간의 길이 t를 결정할 수 있다. 이는 다음 절에서 다룬다.



<그림 9> 제품군별 구매고객의 구매율  $\lambda$  분포

우리는 구매율의 분포로부터 다음과 같은 시사점을 발견할 수 있다. 첫째, 관측기간과 비관측기간의 구매율의 분포가 매우 상이하다. 즉, 비관측기간의 구매율은 평균을 중심으로 몰려있는 정규분포와 유사한 형태를 보이나 관측기간의 구매율은 구매율이 매우 낮은 소비자 그룹과 매우 높은 소비자 그룹의 양 극단으로 나뉘는 것을 확인할 수 있다. 이러한 현상은 관측기간의 타겟마케팅을 위한 전략에 이용될 수 있다. 즉 구매율이 낮은 것으로 예상되는 고객들을 선별하고 이들에게 직접 마케팅 (direct marketing)을 실시하면 관측기간의 구매율을 현재보다 증가시킬 수 있을 것으로 생각된다. 둘째, 가격할인 등의 관측을 반복적으로 시행하는 소매상의 경우 관측기간과 비관측기간을 구분하지 않은 기존 모형을 이용하여 고객의 미래 구매빈도를 예측할 경우 이는 심각한 오류를 유발시킬 수 있을 것으로 예상된다. 이는 관측기간과 비관측기간 중 한 기간만의 자료로부터 모형을 추정하거나 두 기간의 자료를 모두 이용하되 관측/비관측 효과를 구분하지 않은 기존 모형을 이용할 때나 모두 발생할 수 있는 문제이다.

## 5.2 Determining Promotion Duration

제안모형은 관측기간과 비관측기간을 구분하여 기대구매빈도와 기대구매금액을 도출함으로써 인해 특정 목표를 달성하기 위한 관측기간의 적정 길이를 도출하는데 활용될 수 있다. 예를 들어 수익 (혹은 매출이나 구매빈도)를 극대화 해주는 관측기간의 길이를 다음과 같이 결정할 수 있다. 이를 위해 다음과 같은 부호를 정의하자.

$\pi$ : 수익;  $TR$ : 총 매출;  $TC$ : 총 비용

$A$ : 구매건 당 기대구매금액;  $f$ : 기대구매건수

$F$ : 고정비용;  $c$ : 고객의 1회 구매건수에 대해 소매상에게 발생하는 비용 (한계비용)

$t$ : 기간의 길이

하첨자: 1 = 관측기간, 0 = 비관측기간

$$\Theta = \{r_0, \alpha_0, a_0, b_0, r_1, \alpha_1, a_1, b_1, u, v, \phi, \beta\}: \text{모수}$$

그러면 수익을 극대화하기 위한 판촉기간의 적정길이 산출은 다음과 같은 최적화 문제를 해결하여 도출할 수 있다.

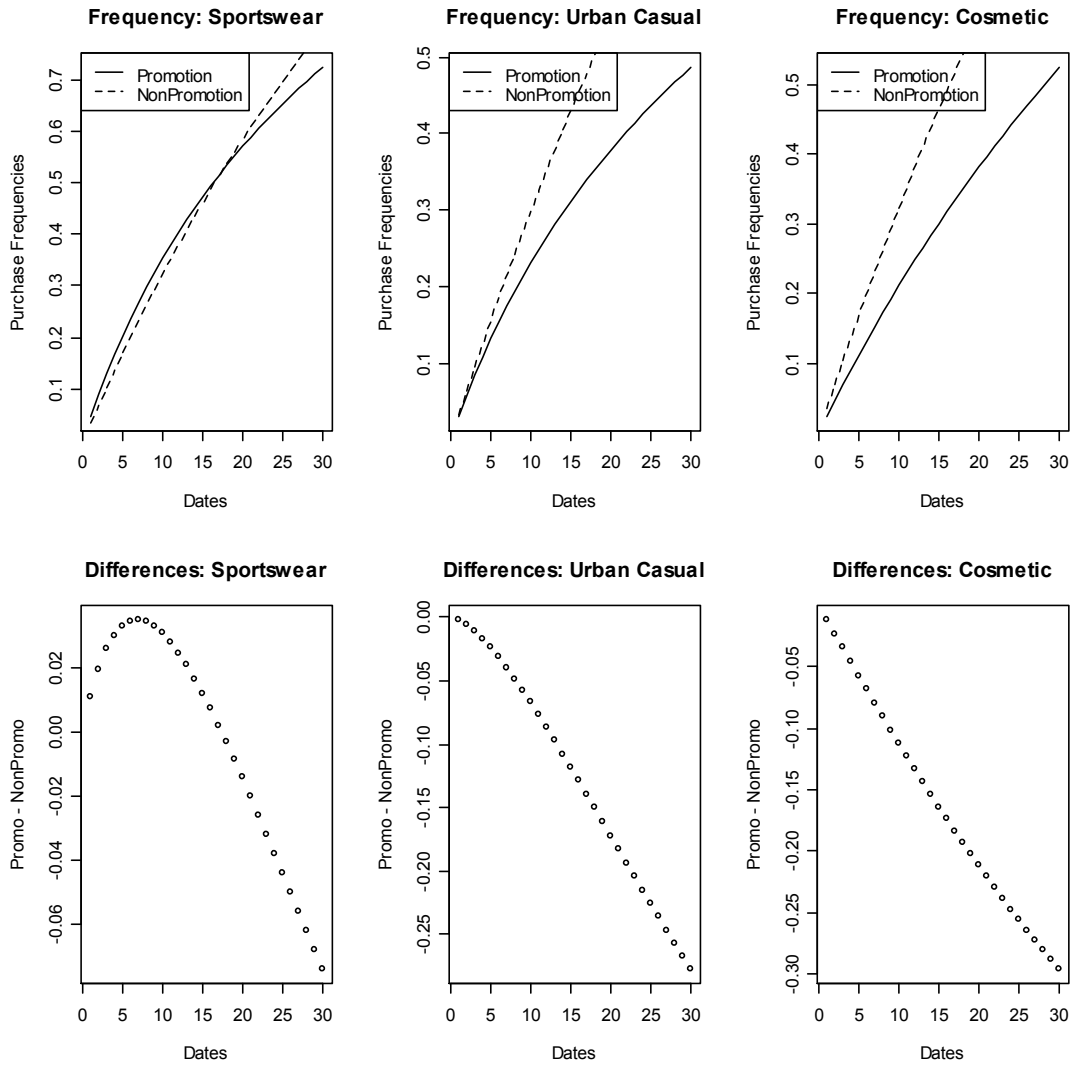
$$\begin{aligned} & \underset{t>0}{\text{Max}} \quad \pi_1(t|\Theta) - \pi_0(t|\Theta) \\ & \text{s.t.} \\ & \quad \pi_1(t|\Theta) - \pi_0(t|\Theta) > 0 \end{aligned} \tag{13}$$

이때 위의 조건을 만족하는 기간  $t$ 가 존재하지 않는 제품군에 대해서는 판촉을 실시하는 것이 단기수익 측면에서 유리하지 않음을 의미한다. 또 모든  $t$ 에 대해  $\pi_1(t) - \pi_0(t) > 0$  라면 이 제품군은 상설할인 형태로 취급하는 것이 더 효과적일 수 있다. 또한 구매빈도 혹은 매출을 극대화하기 위한 판촉기간의 적정길이는 위의 목적함수에서 수익 대신 구매빈도  $f_i(t)$  혹은 매출  $TR_i = A_i \cdot f_i(t)$  을 대입함으로써 산출할 수 있다. 한편, 판촉기간과 비판촉기간에서 수익은 다음과 같이 정의된다.

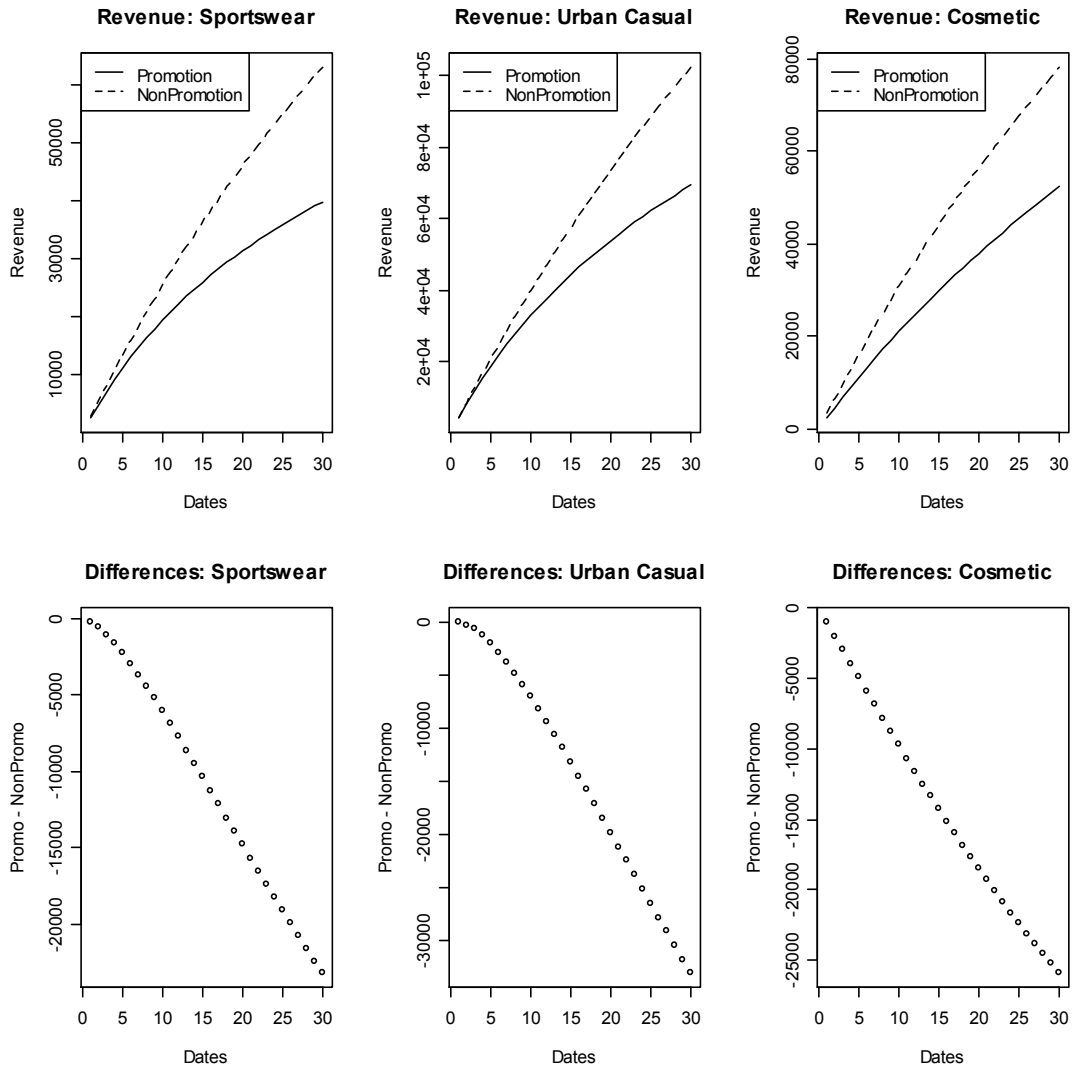
$$\pi_1(t|\Theta) = TR_1(t|\Theta) - TC_1(t|\Theta) = A_1(\Theta)f_1(t|\Theta) - (F + c_1f_1(t|\Theta)) \tag{14-1}$$

$$\pi_0(t|\Theta) = TR_0(t|\Theta) - TC_0(t|\Theta) = A_0(\Theta)f_0(t|\Theta) - (F + c_0f_0(t|\Theta)) \tag{14-2}$$

위의 식에서  $A_i(\Theta)$  와  $f_i(t|\Theta)$  는 각각 식 (12) 및 (5-1)을 통해 도출된다. 아래 <그림 11>은 세 제품군에 대해 기간  $t$ 에 따른 판촉기간과 비판촉기간의 구매빈도 및 그 차이를 보여주고 있다. 이 결과에 따르면 스포츠웨어의 경우 판촉기간을 7일로 설정할 때 구매빈도 관점에서 최대의 효과를 누릴 수 있으며 판촉기간이 약 20일을 초과하면 오히려 구매건수 측면에서 판촉을 하지 않음만 못함을 알 수 있다. 시티케주얼과 화장품류의 경우 구매빈도 측면에서는 판촉을 전혀 실시하지 않는 것이 더 유리하다. 유사하게 <그림 12>는 세 제품군에 대해 기간  $t$ 에 따른 판촉기간과 비판촉기간의 매출  $TR_i(t|\Theta) = A_i(\Theta)f_i(t|\Theta)$  및 그 차이  $A_1(\Theta)f_1(t|\Theta) - A_0(\Theta)f_0(t|\Theta)$  를 보여주고 있다. 이 결과에 따르면 매출측면에서는 어떠한 제품도 가격할인을 실시하지 않는 것이 더 유리하다.

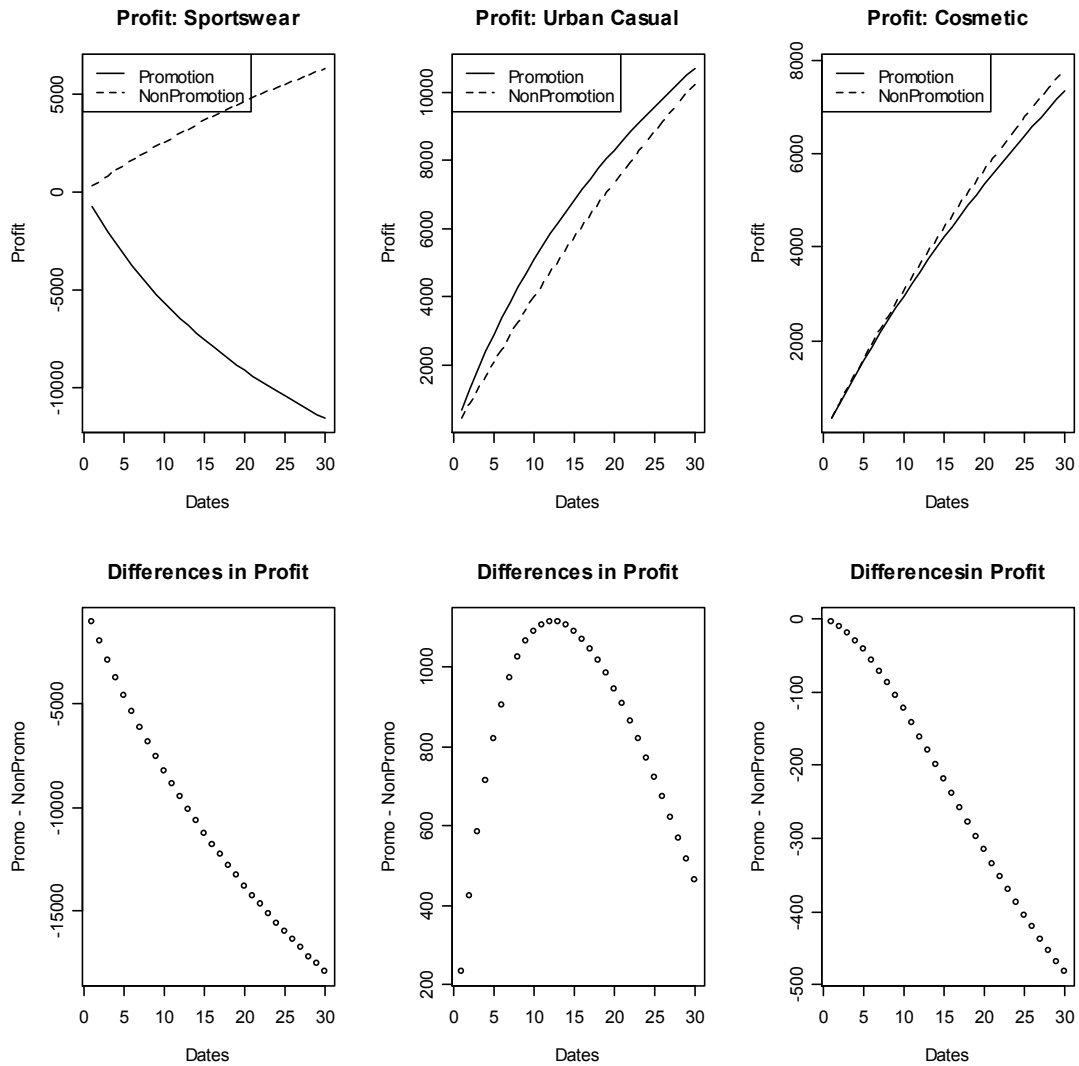


<그림 10> 판촉기간과 비판촉기간 길이에 따른 구매빈도 및 그 차이



<그림 11> 관촉기간과 비관촉기간 길이에 따른 매출 및 그 차이

마지막으로 관촉기간과 비관촉기간 길이에 따른 수익 및 그 차이를 산정하기 위해서 비관촉기간동안 고객의 1회 구매당 소매상의 수익율로  $\delta$  를 가정하자 ( $0 \leq \delta < 1$ ). 즉, 고객의 구매 1건에 대해 소매상에 발생하는 비용을  $c = (1 - \delta)A_0$  로 가정한 후 앞과 동일한 분석을 실시하도록 한다. <그림 13>은 소매상 수익율을 10%로 가정했을 때의 분석결과이다. 이 결과에 따르면 시티케주얼 제품군은 관촉기간을 12일로 설정하였을 때 최대의 수익을 유발시키나 스포츠웨어나 화장품은 모든  $t$ 에서 음의 수익이 발생하여 수익측면에서는 관촉활동을 실시하지 않는 것이 더 유리한 것으로 나타났다.



<그림 12> 관측기간과 비관측기간 길이에 따른 수익 및 그 차이

## 6 Conclusion

SMC의 Pareto/NBD 모형 이래로 비계약 상황하에서 고객의 구매행태에 대한 연구는 지속적인 관심의 대상이 되어왔다. 특히 온라인 쇼핑몰의 등장으로 고객의 구매 이력에 대한 자료가 축적되나 고객이 사멸은 직접적으로 관찰할 수 없는 상황에서 이를 모형화 하기 위한 Pareto/NBD 모형, BG/NBD 모형 등은 경영자에게 매우 유용한 결과들을 제시해 줄 수 있다는 점에서 매우 의미 있는 연구들이다. 그러나 이러한 류의 모형은 과거와 미래의 마케팅 활동상에 변화가 없고 따라서 고객의 구매패턴 상에 변화가 없을 것이라는 가정에 기반해 있다는 점에서 현실적으로 이 모형을 적용할 때는 많은 주의가 요구된다 (Fader et al., 2005). 특히, 소매상의 마케팅 활동 - 가격 할인 등 - 에 따라 고객의 구매패턴이 변화할 수 있을 때 이를 고려하지 않고 기존의 모형들을 그대로 활용하는 것은 경영자에게 잘못된 결과를 제시해 줄 수 있다.

본 연구의 시발점은 이와 같이 소매상의 마케팅 활동에 따라 고객의 구매패턴이 변할 수도

있다는 가정에서 출발하였고 실제로 마케팅 활동에 따라 고객의 구매빈도와 구매금액이 영향을 받음을 실증분석으로 확인하였다. 특히 가격할인을 하는 기간과 그렇지 않은 기간의 고객의 구매율 분포에 현저한 차이가 있으며, 가격할인 기간에는 구매율의 이질성이 눈에 띄게 증가하는 것을 확인하였다. 이는 가격할인 기간에 구매율이 낮은 특정 고객군을 상대로 타겟 마케팅을 실시함으로써 구매빈도를 현재 수준보다 증가시킬 수 있음을 의미한다. 또한, 관측기간과 비관측기간을 구분하여 분석함으로써 최적 관측기간의 길이를 결정하는 문제에 활용될 수 있음을 보여주었다. 이러한 문제를 확장하면 고객의 CLV를 극대화하기 위한 일련의 관측계획을 수립할 수도 있을 것이다.

그러나 이러한 점에도 불구하고 다음과 같은 점이 개선된다면 더 좋은 연구로 발전할 수 있을 것으로 기대한다. 첫째, 일반적으로 고객의 구매빈도와 구매금액간에는 상관관계가 존재할 것으로 생각할 수 있다. 즉, 조금씩 자주 구매하는 고객이 있는 반면 한번 구매 시 대량으로 구매하는 고객도 존재한다. 그리고 이러한 상관관계의 강도는 제품군 별로 상이하게 나타날 수 있다. 따라서 고객의 구매빈도와 구매금액간의 상관관계를 고려한 모형으로 향후 발전시킬 필요가 있다. 둘째, 분석 대상을 특정 제품군에 한정 짓지 않고 장바구니 (shopping basket) 수준으로 확대시키는 것도 좋은 연구 주제인 것으로 생각된다. 장바구니 수준의 기존 연구 (예: Manchanda et al., 1999)가 주로 마케팅 활동으로 인한 소비자의 단기 반응을 살펴보았다면 본 연구의 모형 방법론은 마케팅 활동이 고객의 장바구니에 장기적으로 어떠한 영향을 미치는지를 확인하기 위해 활용될 수 있으며, 이는 보다 정확한 고객평생가치 계산에 도움을 줄 것이다. 셋째, 최적 관측기간을 결정하는데 본 분석결과를 활용하였으나 관측기간과 고객의 구매패턴 간에는 내생성 (endogeneity)문제가 존재하는 것을 인정하지 않을 수 없다. 따라서 이러한 내생성 문제를 모델링할 수 있다면 더 정확한 관측기간의 산정이 가능할 것이다.

## Appendix

### A.1 Derivation of the likelihood function at individual customer level

Case 1)  $x > 0, y > 0$

- $t_1$  에서 첫 구매를 할 우도:  $\lambda_1 e^{-\lambda_1 t_1}$
- $t_2$  에서 두번째 구매를 할 우도:  $(1-p_1)\lambda_1 e^{-\lambda_1(t_2-t_1)}$
- ...
- $t_x$  에서 x번째 구매를 할 우도:  $(1-p_1)\lambda_1 e^{-\lambda_1(t_x-t_{x-1})}$
  
- Non-promotion 기간 중  $t_{x+1}$  시점에서 구매할 우도:  $(1-p_1)e^{-\lambda_1(t_e-t_x)} \cdot \lambda_0 e^{-\lambda_0(t_{x+1}-t_e)}$
- Non-promotion 기간 중  $t_{x+1}$  시점에서 구매할 우도:  $(1-p_0)\lambda_0 e^{-\lambda_0(t_{x+2}-t_{x+1})}$

- ...
- $t_{x+y}$  시점에서  $x+y$  번째 구매를 할 우도:  $(1-p_0)\lambda_0 e^{-\lambda_0(t_{x+y}-t_{x+y-1})}$
- 마지막으로,  $(t_{x+y}, T]$  기간 동안 비구매 우도:  $p_0 + (1-p_0)e^{-\lambda_0(T-t_{x+y})}$

따라서 Case 1의 전체 우도는 위의 결과를 모두 곱하면

$$L(\lambda_1, \lambda_0, p_1, p_0 | t_1, t_2, \dots, t_{x+y}, t_e, T) = (1-p_1)^x (1-p_0)^{y-1} \lambda_1^x e^{-\lambda_1 t_e} \lambda_0^y e^{-\lambda_0(t_{x+y}-t_e)} \left( p_0 + (1-p_0)e^{-\lambda_0(T-t_{x+y})} \right) \quad (\text{A.1})$$

Case 2)  $x=0, y>0$

- 관측기간  $(0, t_e]$  동안 한번도 구매를 하지 않을 우도:  $e^{-\lambda_1 t_e}$
- 비관측기간  $(t_e, T]$  동안  $y$ 번 구매를 할 우도:  $\lambda_0^y (1-p_0)^{y-1} e^{-\lambda_0(t_{x+y}-t_e)} \times \left( p_0 + (1-p_0)e^{-\lambda_0(T-t_{x+y})} \right)$

따라서 Case 2의 전체 우도는 위의 결과를 모두 곱하면

$$L(\lambda_0, \lambda_1, p_1, p_0 | X=0, Y>0, t_e, t_y, T) = e^{-\lambda_1 t_e} \times \lambda_0^y (1-p_0)^{y-1} e^{-\lambda_0(t_{x+y}-t_e)} \times \left( p_0 + (1-p_0)e^{-\lambda_0(T-t_{x+y})} \right) \quad (\text{A.2})$$

Case 3)  $x>0, y=0$

- 관측기간  $(0, t_e]$  동안  $x$ 번 구매 후 사멸하지 않을 우도:  $(1-p_1)^x \lambda_1^x e^{-\lambda_1 t_x} e^{-\lambda_1(t_e-t_x)}$
- 관측기간  $(0, t_e]$  동안  $x$ 번 구매 후 사멸할 우도:  $p_1 (1-p_1)^{x-1} \lambda_1^x e^{-\lambda_1 t_x}$
- 비관측기간  $(t_e, T]$  동안 한번도 구매하지 않을 우도:
  - 관측기간동안  $x$ 번 구매 후 사멸하지 않았으면  $e^{-\lambda_0(T-t_e)}$
  - 관측기간동안  $x$ 번 구매 후 사멸했으면 1

따라서 Case 3의 전체 우도는

$$L(\lambda_0, \lambda_1, p_1, p_0 | X>0, Y=0, t_x, t_e, T) = \lambda_1^x (1-p_1)^{x-1} e^{-\lambda_1 t_x} \times \left( p_1 + (1-p_1)e^{-\lambda_1(t_e-t_x)} e^{-\lambda_0(T-t_e)} \right) \quad (\text{A.3})$$

Case 4)  $x=0, y=0$

- 관측기간  $(0, t_e]$  동안 한번도 구매하지 않을 우도:  $e^{-\lambda_1 t_e}$
- 비관측기간  $(t_e, T]$  동안 한번도 구매하지 않을 우도:  $e^{-\lambda_0(T-t_e)}$

따라서 Case 4의 전체 우도는



$$L(\lambda_0, \lambda_1, p_1, p_0 | X = Y = 0, t_e, T) = e^{-\lambda_1 t_e} e^{-\lambda_0 (T - t_e)} \quad (\text{A.4})$$

**A2. Derivation of**  $P(X(t_1) = x, Y(t_0) = y | \lambda_1, p_1, \lambda_0, p_0)$

Case 1)  $x > 0, y > 0$

$$\begin{aligned} & P(X(t_1) = x, Y(t_0) = y | \lambda_1, p_1, \lambda_0, p_0) \\ &= P(\text{active after } x^{\text{th}} \text{ purchase}) \cdot P(T_x \leq t_1, T_{x+1} > t_1) \\ &\times [P(\text{active after } y^{\text{th}} \text{ purchase}) \cdot P(T_y \leq t_0, T_{y+1} > t_0) + \delta_{y>0} P(\text{become inactive after } y^{\text{th}} \text{ purchase}) \cdot P(T_y \leq t_0)] \\ &= (1-p_1)^x \frac{e^{-\lambda_1 t_1} (\lambda_1 t_1)^x}{x!} \times \left[ (1-p_0)^y \frac{e^{-\lambda_0 t_0} (\lambda_0 t_0)^y}{y!} + p_0 (1-p_0)^{y-1} \cdot \left( 1 - e^{-\lambda_0 t_0} \sum_{j=0}^{y-1} \frac{(\lambda_0 t_0)^j}{j!} \right) \right] \end{aligned} \quad (\text{A.1})$$

Case 2)  $x=0, y > 0$

$$\begin{aligned} & P(X(t_1) = 0, Y(t_0) = y | \lambda_1, p_1, \lambda_0, p_0) = P(X(t_1) = 0)P(Y(t_0) = y | \text{alive})P(\text{alive}) \\ &= e^{-\lambda_1 t_1} \left[ (1-p_0)^y \frac{e^{-\lambda_0 t_0} (\lambda_0 t_0)^y}{y!} + p_0 (1-p_0)^{y-1} \cdot \left( 1 - e^{-\lambda_0 t_0} \sum_{j=0}^{y-1} \frac{(\lambda_0 t_0)^j}{j!} \right) \right] \end{aligned} \quad (\text{A.2})$$

Case 3)  $x > 0, y=0$

$$\begin{aligned} & P(X(t_1) = x, Y(t_0) = 0 | \lambda_1, p_1, \lambda_0, p_0) \\ &= P(\text{active after } x^{\text{th}} \text{ purchase}) \cdot P(T_x \leq t_1, T_{x+1} > t_1) \times P(Y(t_0) = 0) + P(\text{become inactive after } x^{\text{th}} \text{ purchase})P(T_x \leq t_1) \\ &= (1-p_1)^x \frac{e^{-\lambda_1 t_1} (\lambda_1 t_1)^x}{x!} e^{-\lambda_0 t_0} + p_1 (1-p_1)^{x-1} \left( 1 - e^{-\lambda_1 t_1} \sum_{j=0}^{x-1} \frac{(\lambda_1 t_1)^j}{j!} \right) \end{aligned} \quad (\text{A.3})$$

Case 4)  $x=0, y=0$

$$\begin{aligned} & P(X(t_1) = 0, Y(t_0) = 0 | \lambda_1, p_1, \lambda_0, p_0) = P(X(t_1) = 0)P(Y(t_0) = 0) \\ &= e^{-\lambda_1 t_1} e^{-\lambda_0 t_0} \end{aligned} \quad (\text{A.4})$$

**References**

Allenby, Greg M., Neeraj Arora, and James L. Ginter (1998), "On the Heterogeneity of Demand," *Journal of Marketing Research*, 35 (3), 384-89.

Allenby, Greg M. and Peter E. Rossi (1999), "Marketing Models of Consumer Heterogeneity," *Journal of Econometrics*, 89, 57-58.

Blattberg, Robert C., and John Deighton (1996), "Manage Marketing by the Customer Equity Test," *Harvard Business Review*, 74 (4), 136-44.

- Colombo, Richard, and Weina Jiang (1999), "A stochastic RFM model," *Journal of Interactive Marketing*, 13(3), 2-12.
- DeSarbo, W, and V. Ramaswamy (1994), "CRISP: Customer Response Based Iterative Segmentation Procedures for Response Modeling in Direct Marketing," *Journal of Direct Marketing*, 8 (3), 7-20.
- Fader, Peter S., Bruce G. S. Hardie, and Ka Lok Lee (2005), "Counting Your Customers the Easy Way: An Alternative to the Pareto/NBD Model," *Marketing Science*, 24 (2), 275 - 84.
- Gupta, Sunil, Donald R. Lehmann, and Jennifer Ames Stuart (2004), "Valuing Customers," *Journal of Marketing Research*, 41 (1), 7-18.
- Kamakura, Wagner A. and Gary J. Russell (1989), "A Probabilistic Choice Model for Market Segmentation and Elasticity Structure," *Journal of Marketing Research*, 26 (4), 379-90.
- Manchanda, Puneet, Asim Ansari, and Sunil Gupta (1999), "The 'Shopping Basket': A Model for Multicategory Purchase Incidence Decisions," *Marketing Science*, 18 (2), 95-114.
- Schmittlein, David C., Donald G. Morrison, and Richard Colombo (1987), "Counting Your Customers: Who Are They and What Will They Do Next?," *Management Science*, 33 (1), 1-24.
- Schmittlein, David C. and Robert A. Peterson (1994), "Customer base analysis: An industrial purchase process application," *Marketing Science*, 13(1) 41-67.