

브랜드 선택확률 분석을 위한 구조방정식 모형

이상호 · 이혜선 · 김윤대 · 전치혁[†]

포항공과대학교 산업경영공학과

Estimation of a Structural Equation Model Including Brand Choice Probabilities

Sang-Ho Lee · Hyeseon Lee · Yun-Dae Kim · Chi-Hyuck Jun

Department of Industrial and Management Engineering POSTECH

The partial least squares (PLS) method is popularly used for estimating the structural equation model, but the existing algorithm may not be directly implemented when probabilities are involved in some constructs or manifest variables. We propose a structural equation model including the brand choice as one construct having brand choice probabilities as its manifest variables. Then, we develop a PLS-based algorithm for the structural equation model by utilizing the multinomial logit model. A case is introduced as an application and simulation studies are performed to validate the proposed algorithm.

Keywords: Customer Equity, Multinomial Logit, Partial Least Squares, Simulation, Structural Equation Model

1. 서론

구조방정식 모형(structural equation model; SEM)은 요인분석(factor analysis)과 회귀분석이 결합된 형태로써 교육학, 심리학, 경영학, 경제학, 마케팅 등 주로 사회과학분야에서 활용되고 있는데, 어떤 조직이나 사회현상에서 여러 추상적인 요인간의 인과관계를 분석하기 위하여 사용된다. 여기서 추상적인 요인을 잠재변수(latent variable 또는 construct)라 하며 이는 직접 관측되지 않으나 간접적으로 측정되는 여러 변수가 연결되어 있다고 가정한다. 구조방정식 모형은 특히, Bagozzi(1980), Fornell and Larcker(1981) 등이 마케팅분야에 적용한 이후 활발하게 도입되고 있으며 ACSI(American Customer Satisfaction Index)(Fornell *et al.*, 1996), ECSI(European Customer Satisfaction Index)(Grigoroudis and Siskos, 2003; Martensen *et al.*, 2000) 등의 소비자 만족지수를 산출하는데 적용되고 있다. 구조방정식 모형의 추정을 위한 접근방법으로 LISREL 형태의 방법(Joreskog and Sorbom, 1989)과 부분 최소자승법(partial least squares; PLS)이 널리 사용되고 있으며 각각 장단점을 갖고 있다. 이 중 PLS의 주요 장점은 오차항에 대한 분포가정이 필요 없으며, 계산

시간이 빠르고, 관측치가 적은 경우에도 비교적 안정적인 결과를 도출한다는 데 있다(Fornell and Bookstein, 1982; Hulland, 1999).

최근에는 고객자산(customer equity)에 대한 관심이 증대되어 이를 모형에 반영시키고자 하는 노력이 있다. 고객자산이란 고객의 평생가치에 바탕을 두고 있으며 신규고객을 확보하고 유지하는 활동을 최적화에 활용될 수 있다(Blattberg and Deighton, 1996). 고객자산을 평가하기 위해서는 현재 및 잠재 고객의 브랜드 선택을 분석할 필요가 있으므로 우선적으로 브랜드 선택을 모형에 어떻게 반영시킬 것인지를 결정하여야 한다. Berger and Nasr(1998)은 고객의 생애가치(customer lifetime value : CLV)를 평가하고 이의 현가를 고객자산으로 고려하는 것을 제안하였다. Rust *et al.*(2004)은 고객의 브랜드 선택을 분석하기 위하여 CLV를 바탕으로 한 로짓모형(logit model)을 이용하였다.

본 연구에서는 구조방정식 모형에서 여러 기업 활동 및 고객만족 등이 브랜드 선택에 영향을 준다고 하고 표본 고객들에 대한 브랜드별 선택확률이 입력될 때 모형 추정방안을 모색하고자 한다. 관련된 연구로 구조방정식 모형으로 브랜드

[†] 연락저자 : 전치혁 교수, 790-784 경북 포항시 남구 효자동 산31 포항공과대학교 산업경영공학과, Tel : 054-279-2197,

E-mail : chjun@postech.ac.kr

2010년 2월 9일 접수; 2010년 5월 10일 수정본 접수; 2010년 5월 13일 게재 확정.

선호도를 분석한 Hellier *et al.*(2003)의 연구가 있으나, 잠재변수 또는 측정변수에 확률이 개입된 구조방정식 모형을 PLS로 추정한 연구는 아직 없었다. 본 연구에서는 잠재변수 간 관계를 다항 로짓모형으로 설정하여 PLS로 추정하는 방안을 사용코자 한다.

본 논문은 다음과 같이 구성되어 있다. 제 2장에서 일반적인 구조방정식 모형을 소개하며 특히 브랜드 선택확률이 포함되는 경우의 모형을 제시한다. 제 3장에서는 모형의 추정을 위하여 제안하는 PLS에 기반한 알고리즘을 설명한다. 제 4장에서 하나의 응용 사례를 소개하고 제안하는 방법의 타당성 검증을 위해 제 5장에서 시뮬레이션을 수행하고 결과를 보고한다.

2. 브랜드 선택확률을 포함한 구조방정식 모형

서론에서 언급한 바와 같이 구조방정식 모형은 잠재변수간의 인과관계를 분석한다. 잠재변수는 다시 외생 잠재변수(exogenous latent variable)와 내생 잠재변수(endogenous latent variable)로 구분하는데, 외생 잠재변수란 다른 잠재변수에 영향을 주나 받지 않는 잠재변수를 말하며, 내생 잠재변수란 다른 잠재변수에 영향을 주기도하고 받기도 하는 것을 일컫는다.

브랜드 선택을 포함하는 구조방정식 모형은 아래와 같다고 가정한다. k 개의 외생잠재변수 벡터를 $\xi = (\xi_1, \dots, \xi_k)^T$ 라 하고 이에 대응하는 q 개의 측정변수 $x = (x_1, \dots, x_q)^T$ 가 있다고 하자. 이 때, 외생잠재 변수에 대한 모형은 아래와 같다.

$$x = A^x \xi + \delta \quad (1)$$

여기서 A^x ($q \times k$)는 추정하여야 할 계수 행렬이며 δ 는 평균이 0인 오차항 벡터를 나타낸다. 일반적으로 측정변수를 형성적(formative)인 것과 반영적(reflective)인 것으로 구분하는데 본 연구에서는 편의상 모두 반영적인 것으로 가정한다. 이 때, 각각의 측정변수는 하나의 잠재변수로부터 영향을 받는다고 가정한다.

또한, 브랜드 선택을 제외한 m 개의 내생잠재변수 벡터를 $\eta = (\eta_1, \dots, \eta_m)^T$ 라 하고 이에 대응하는 p 개의 측정변수 $y = (y_1, \dots, y_p)^T$ 가 있다고 할 때, 내생잠재변수에 대한 모형은 다음과 같다.

$$y = A^y \eta + \epsilon \quad (2)$$

여기서 A^y ($p \times m$)는 추정하여야 할 계수 행렬이며 ϵ 는 평균이 0인 오차항을 나타낸다. 이와 별도로 브랜드 선택을 나타내는 잠재변수를 브랜드 잠재변수라 칭하며 η^B 라 하자. 브랜드 종류는 다수이므로 이 변수는 실제 벡터이다. 브랜드 잠재변수에 대응하는 측정변수는 브랜드 선택확률들이다. 브랜드 수를 b 로 나타내고 브랜드 i 의 선택 확률을 P_i 라 할 때, 브랜드

i 에 대응하는 브랜드 잠재변수를 아래와 같이 브랜드 b 를 기준으로 하는 기준범주로짓(baseline-category logit)으로 나타내자.

$$\eta_i^B = \log(P_i/P_b), \quad i = 1, \dots, b-1 \quad (3)$$

만약, 선택 확률 P_i 가 직접 측정되지 않고 대신 브랜드 중 하나를 선택한 형태의 데이터가 주어진다면 P_i 를 추정하여야 한다. 본 연구에서는 P_i 가 직접 측정되는 경우만을 다룬다. 참고로 η_i^B 가 산출된 후 P_i 는 다음과 같이 추정할 수 있다.

$$\hat{P}_i = \frac{\exp(\eta_i^B)}{1 + \sum_{i=1}^{b-1} \exp(\eta_i^B)}, \quad i = 1, \dots, b-1 \quad (4a)$$

$$\hat{P}_b = \frac{1}{1 + \sum_{i=1}^{b-1} \exp(\eta_i^B)} \quad (4b)$$

한편, 외생 잠재변수와 내생 잠재변수(브랜드 잠재변수를 제외) 간의 구조모형으로 다음과 같은 선형 모형을 가정한다.

$$\eta = B\eta + \Gamma\xi + \zeta \quad (5)$$

위에서 B 와 Γ 는 각각 ($m \times m$), ($m \times k$) 계수 행렬이며 ζ 는 평균이 0인 오차항 벡터이다. 그리고, 브랜드 잠재변수에 대한 모형은 다음과 같은 다항 로짓모형(multinomial logit model)을 가정한다.

$$\eta_i^B = b_i^T \eta + \gamma_i^T \xi + \zeta_i, \quad i = 1, \dots, b-1 \quad (6)$$

여기서, b_i 와 γ_i 는 추정하여야 할 계수벡터이다.

예를 들어 <Figure 1>과 같이 1개의 외생잠재변수인 가치동인(value drivers ξ)와 2개의 내생잠재변수, 즉 고객만족(customer satisfaction η_1) 및, 잠재고객(potential customers η_2), 그리고 브랜드 선택(brand choice)이 있는 모형에 대하여 구조방정식 모형은 다음과 같이 기술된다.

<외생잠재변수의 측정 모형>

$$\begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} \lambda_{11}^x \\ \lambda_{21}^x \\ \lambda_{31}^x \end{pmatrix} \xi + \begin{pmatrix} \delta_1 \\ \delta_2 \\ \delta_3 \end{pmatrix} \quad (7)$$

<내생잠재변수의 측정 모형>

$$\begin{pmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \end{pmatrix} = \begin{pmatrix} \lambda_{11}^y & 0 \\ \lambda_{21}^y & 0 \\ 0 & \lambda_{32}^y \\ 0 & \lambda_{42}^y \end{pmatrix} \begin{pmatrix} \eta_1 \\ \eta_2 \end{pmatrix} + \begin{pmatrix} \epsilon_1 \\ \epsilon_2 \\ \epsilon_3 \\ \epsilon_4 \end{pmatrix} \quad (8)$$

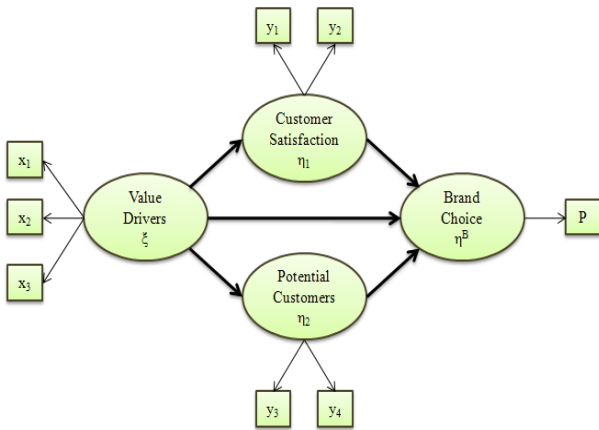


Figure 1. An example of structural equation model including brand choice

<구조 모형>

$$\begin{pmatrix} \eta_1 \\ \eta_2 \end{pmatrix} = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} \eta_1 \\ \eta_2 \end{pmatrix} + \begin{pmatrix} \gamma_1 \\ \gamma_2 \end{pmatrix} \xi + \begin{pmatrix} \zeta_1 \\ \zeta_2 \end{pmatrix} \quad (9)$$

브랜드 수가 3이라면 브랜드 잠재변수는 다음과 같이 표현되며,

$$\eta_i^B = \log(P_i/P_3), i = 1, 2 \quad (10)$$

이에 대한 다항로짓 모형은 아래와 같다.

$$\eta_i^B = b_{1i}\eta_1 + b_{2i}\eta_2 + \gamma_i\xi + \zeta_i, i = 1, 2 \quad (11)$$

3. PLS를 이용한 제안 모형의 추정

제 2장의 식 (1), 식 (2), 식 (5)로 기술한 구조방정식 모형에서 계수행렬 A^x, A^y, B, T 를 추정하여야 하며 동시에 식 (6)의 모형을 추정하여야 한다. 내부/외부 잠재변수의 구분없이 j번째 잠재변수와 관련된 측정변수들의 데이터 행렬을 X_j 라 하자. 즉, j번째 잠재변수에 q_j 개의 측정변수가 있으며 총 n개의 관측수가 있다면 X_j 는 $(n \times q_j)$ 행렬이 된다. 이 때, 브랜드 잠재변수는 별도로 고려한다. 제 2장에서 언급하였듯이 총 b개의 브랜드가 있을 때, (b-1)개의 브랜드 잠재변수가 추가되며 이에 대한 측정변수는 선택확률 $P_i (i = 1, \dots, b)$ 이다. 행렬은 필요에 따라 평균화(mean centered) 하거나 표준화(standardized)할 수 있다.

다음은 PLS를 이용한 제안 모형의 추정을 위하여 Noonan and Wold(1982)가 제시한(또는 Guinot et al., 2001 참조) 구조방정식모형에 대한 알고리즘을 수정한 것이다. 아래 알고리즘에서 가중치 벡터 w_j 들은 계수 행렬 A^x 및 A^y 와 관련되며, 잠재변수간의 회귀계수 β_{ji} 는 계수행렬 B 및 T 와 연결된다.

단계 0 : (가중치 초기화)

j번째 잠재변수 (브랜드 잠재변수 제외)의 관련 측정변수들과의 가중치 벡터 w_j 를 임의의 값으로 초기화 시키되 가중치 합을 1이 되게 한다. 예를 들어, <Figure 1>의 모형에서 잠재변수 ξ 에 세 개의 측정변수가 있으므로 $w_j^T = (1/3, 1/3, 1/3)$ 로 놓을 수 있다.

단계 1 : (잠재변수의 외부 추정치 산출)

다음과 같이 j번째 잠재변수의 외부 추정치 (V_j)를 구한다.

$$V_j = (X_j w_j)^*$$

여기서 (*)은 계산된 값을 표준화함을 의미한다. 한편, 브랜드 잠재변수 η_i^B 에 대한 외부 추정치는 아래와 같이 산출한다.

$$V_i^B = \log(P_i/P_b), i = 1, \dots, b-1$$

단, $P_i = 0 (i = 1, \dots, b)$ 인 경우에는 다른 작은 값(예를 들어 0.01)으로 대체하고 확률 합이 1이 되도록 다시 조정후 후의 외부 추정치를 산출한다.

단계 2 : (잠재변수의 내부추정치 산출)

외부 추정치간의 상관관계와 잠재변수들의 관계를 이용해 다음과 같이 j번째 잠재변수의 내부 추정치(Z_j)를 구한다.

$$Z_j = \left(\sum_{i: \beta_{ji} \neq 0} e_{ji} V_i \right)^*$$

여기서 β_{ji} 는 j번째 잠재변수가 i번째 잠재변수에 주는 영향을 나타내는 계수로 단계 5에서 추정되는 값이다. e_{ji} 는 잠재변수의 외부추정치들의 관계를 나타내는 값으로 centroid, factorial, 그리고 structural의 세 가지 방법으로 계산할 수 있으나 (Tenenhaus et al., 2005 참조) 편의상 아래와 같이 structural 방식을 사용한다.

$$e_{ji} = \begin{cases} V_i \text{와 } V_j \text{의 상관계수,} & V_j \text{가 } V_i \text{에 영향줄 때} \\ V_j \text{설명 회귀식에서 } V_i \text{의 계수} & V_j \text{가 } V_i \text{의 영향받을 때} \end{cases}$$

j번째 잠재변수가 브랜드 잠재변수에도 영향을 미치는 경우 이의 내부추정치는 다음과 같이 산출한다.

$$Z_j = \left(\sum_{i: \beta_{ji} \neq 0} e_{ji} V_i + \sum_{k=1}^{b-1} e_{jk} V_k^B \right)^*$$

단계 3 : (가중치 산출)

가중치 w_j 를 새로 구한다. 측정변수와 잠재변수의 단순회귀 분석을 수행하여 그 계수를 w_j 로 사용하면 된다. 따라서,

$$w_j = \frac{1}{n} X_j^T Z_j$$

단계 4 : (반복)

외부 추정치 혹은 내부 추정치가 수렴할 때까지 단계 1~단계 3을 반복한다. 수렴하는 경우, 단계 5를 수행한다.

단계 5 : (구조모형의 추정)

내생 잠재변수 η_j 에 대하여 수렴된 외부 추정치를 사용하는 다음의 다중회귀모형으로부터 잠재변수들 간의 계수 β_{ij} 를 추정한다.

$$V_j = \sum_{i \neq j} \beta_{ij} V_i + \zeta_j, j=1, \dots, m$$

브랜드 잠재변수에 대해서는 다음과 같은 다항로짓모형으로부터 회귀계수를 도출한다.

$$V_i^B = b_i^T v^I + \gamma_i v^E + \zeta_i, i=1, \dots, b-1$$

위에서 v^I 는 관련되는 내부 잠재변수의 외부추정치 벡터, v^E 는 외부 잠재변수의 외부추정치 벡터를 의미한다.

단계 1에서 브랜드 잠재변수의 외부추정치를 구하는 다른 방안을 고려할 수 있다. 즉, 잠재변수의 외부추정치를 바탕으로 식 (6)의 다항로짓모형에 적합시킨 후, 식 (4)에 따라 확률값을 추정하여 다음과 같이 산출하는 방안이다.

$$V_i^B = \log(\hat{P}_i / \hat{P}_b), i=1, \dots, b-1 \tag{12}$$

이 방안은 입력된 확률값을 직접 사용하는 대신 모형의 추정확률 값을 사용하는 것으로 확률이 0일 때 발생하는 문제를 방지할 수 있다. 이를 대안이라 칭하자. 위에서 제안된 방안은 잠재변수의 외부추정치가 알고리즘 수행 중에 변하지 않는 반면, 식 (12)의 대안의 경우 다른 잠재변수의 외부추정치가 달라짐에 따라 브랜드 잠재변수의 외부추정치 역시 변한다는 특징이 있다. 후자의 경우 장점이 될지 단점이 될지는 명확하지 않다. 이에 따라 제 5장의 시뮬레이션에서 두 방안의 성능을 시험하였다. 결과적으로 서로의 성능이 유사하나 제안방안이 다소 좋게 나타났다.

4. 적용 사례

<Figure 2>는 이동통신 서비스 산업에서 브랜드별 구매확률을 분석할 수 있는 구조방정식 모형을 나타낸 것이다. 고객만족도를 비롯한 총 9개 잠재변수들이 소비자의 구매확률에 영향을 준다는 가정을 두고 있다. 이 중 LV15(가입)는 다시 LV1(프로세스) 과 LV2(상담직원)로부터 영향을 받으며, LV16, LV17,

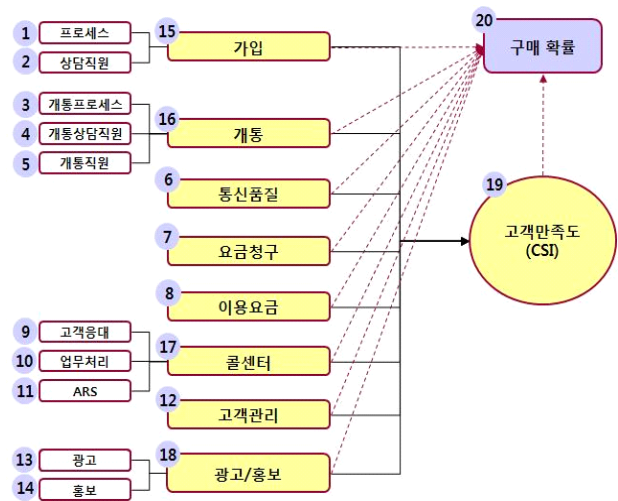


Figure 2. Customer satisfaction index (CSI) and brand choice model for mobile telecommunication industry

Table 1. Regression coefficients between LVs of structural model in <Figure 2>

	LV15	LV16	LV17	LV18	LV19
LV1	0.3493	0	0	0	0
LV2	0.6505	0	0	0	0
LV3	0	0.2077	0	0	0
LV4	0	0.2471	0	0	0
LV5	0	0.5453	0	0	0
LV6	0	0	0	0	0.2332
LV7	0	0	0	0	0.0328
LV8	0	0	0	0	0.0018
LV9	0	0	0.2731	0	0
LV10	0	0	0.3864	0	0
LV11	0	0	0.3405	0	0
LV12	0	0	0	0	0.1553
LV13	0	0	0	0.6546	0
LV14	0	0	0	0.3456	0
LV15	0	0	0	0	0.0727
LV16	0	0	0	0	0.0595
LV17	0	0	0	0	0.0667
LV18	0	0	0	0	0.0873
LV19	0	0	0	0	0

LV18 또한 영향을 받는 상위 잠재변수가 존재한다. 측정변수로는 총 62개의 항목이 있으나 복잡도를 피하기 위하여 여기에 표시하지 않았다. 측정변수 62개 항목과 3개의 브랜드에 대한 구매확률 데이터는 총 1500명의 소비자를 대상으로 설문조사를 수행하여 얻어졌다.

<Figure 2>의 모델에 대하여 본 논문에서 제안한 방법으로 분석한 결과가 아래 <Table 1>과 <Table 2>에 정리되어 있다.

Table 2. Multinomial logistic regression coefficients for brand choice probability

	Logit1	Logit2
Constant	0.4035	1.4188
LV6	-0.0178	0.0475
LV7	-0.0414	-0.0341
LV8	0.0622	0.0631
LV12	-0.1258	0.0267
LV15	0.0652	0.0782
LV16	0.0071	0.0212
LV17	0.0038	-0.0312
LV18	-0.0852	-0.0919
LV19	0.0802	0.1947

<Table 1>은 식 (5)의 추정결과를 나타낸 것으로, <Figure 2>의 구조 모형에서 실선 화살표로 나타낸 잠재변수 간 관계를 설명하는 회귀계수이다. 예를 들어, LV2는 LV1에 비해 약 2배만큼 LV15에 영향을 미치고 있다. 또한 LV6 (통신품질)이 다른 요인보다 고객만족도 (LV19)에 가장 큰 영향을 주고 있음을 알 수 있다.

<Table 2>는 식 (6)의 추정결과를 나타낸 것이며, <Figure 2>에서 점선 화살표로 나타낸 브랜드 구매확률과 관련된 다항로짓모형의 계수를 정리한 것이다. Logit1은 브랜드 3대비 브랜드 1의 로그 승산비에 대한 모형의 계수이며, Logit2는 브랜드 3대비 브랜드 2의 로그 승산비에 대한 모형의 계수이다. 음수의 경우 해당 잠재변수가 한 단위 상승할 때 승산비가 감소함을 의미하며, 양수의 경우 승산비가 증가함을 의미한다. 결과적으로 LV19 (고객만족도)가 브랜드 1 또는 브랜드 2의 구매 확률에 가장 크게 영향을 줌을 볼 수 있다.

5. 시뮬레이션

5.1 실험계획

<Figure 1>과 같이 1개의 외생잠재변수 ξ 와 2개의 내생잠재변수 η_1, η_2 를 갖는 모형을 고려하자. 대상 브랜드는 3개이며 브랜드 선택 잠재변수벡터는 (η_1^B, η_2^B) 이다. 실제 (true) 구조 모형이 아래와 같다고 가정하고 잠재변수 및 측정변수를 생성시킨다.

$$\eta_1^B = 0.7\eta_1 + 0.4\eta_2 + 0.3\xi + \zeta_1$$

$$\eta_2^B = 0.4\eta_1 + 0.5\eta_2 + 0.2\xi + \zeta_2$$

여기서 오차항 ζ_i ($i = 1, 2$)는 평균이 0, 분산이 σ_{ζ}^2 (추후 두 가지 값을 고려-Case 1의 경우 $\sigma_{\zeta}^2 = 0.1$, Case 2의 경우 $\sigma_{\zeta}^2 = 0.2$)

인 정규분포로부터 랜덤하게 생성한다.

외생잠재변수 ξ 는 평균이 0, 분산이 1인 정규분포로부터 랜덤으로 생성하며, 내생잠재변수는 아래로부터 생성한다.

$$\eta_1 = 0.5\xi + \zeta$$

$$\eta_2 = 0.6\xi + \zeta$$

여기서 ζ 는 평균 0, 분산 0.1인 정규분포로부터 생성한다. 실험결과분석에서 위와 같은 실제 가정된 잠재변수들 간의 계수와 추정된 계수의 비교가 이루어진다.

잠재변수에 해당하는 측정변수를 생성하기 위하여 아래와 같은 두 가지 실험을 설계한다. 두 실험에서 평균벡터는 동일하며 분산-공분산행렬만을 변화시킨다.

<실험 I>

외생잠재변수 ξ 에 대한 3개의 측정변수를 아래와 같은 평균 벡터와 분산-공분산 행렬을 갖는 다변량 정규분포로부터 생성한다.

$$\text{평균벡터} = \begin{pmatrix} 0.5\xi \\ 0.3\xi \\ 0.2\xi \end{pmatrix}$$

$$\text{분산-공분산행렬} = \begin{pmatrix} 0.1 & 0.08 & 0.05 \\ 0.08 & 0.1 & 0.08 \\ 0.05 & 0.08 & 0.1 \end{pmatrix}$$

즉, 외생잠재변수 ξ 에 대한 3개의 측정변수에 대한 실제 가중치를 (0.5, 0.3, 0.2)로 둔 것이다. 실험결과분석의 하나로 실제 가중치를 추정된 가중치와 비교하고자 한다. 그리고, η_1 에 대한 측정변수 y_1, y_2 는 아래와 같은 평균 벡터와 분산-공분산 행렬을 갖는 다변량 정규분포로부터 생성하며,

$$\text{평균벡터} = \begin{pmatrix} 0.6\eta_1 \\ 0.4\eta_1 \end{pmatrix},$$

$$\text{분산-공분산행렬} = \begin{pmatrix} 0.1 & 0.07 \\ 0.07 & 0.1 \end{pmatrix}$$

η_2 에 대한 측정변수 y_3, y_4 는 아래와 같은 평균 벡터와 분산-공분산 행렬을 갖는 다변량 정규분포로부터 생성한다.

$$\text{평균벡터} = \begin{pmatrix} 0.2\eta_2 \\ 0.5\eta_2 \end{pmatrix},$$

$$\text{분산-공분산행렬} = \begin{pmatrix} 0.1 & 0.07 \\ 0.07 & 0.1 \end{pmatrix}$$

<실험 II>

측정변수를 생성할 때 평균은 실험 I과 동일하나 분산 및 공분산 값들을 약간 증대시켜 동일한 실험을 수행하였다. 외생잠재변수 ξ 에 대한 3개의 측정변수를 아래와 같은 평균 벡터

와 분산-공분산 행렬을 갖는 다변량 정규분포로부터 생성한다.

$$\text{평균벡터} = \begin{pmatrix} 0.5\xi \\ 0.3\xi \\ 0.2\xi \end{pmatrix},$$

$$\text{분산-공분산행렬} = \begin{pmatrix} 0.2 & 0.16 & 0.1 \\ 0.16 & 0.2 & 0.16 \\ 0.1 & 0.16 & 0.2 \end{pmatrix}$$

그리고, η_1 에 대한 추정변수 y_1, y_2 는 아래와 같은 평균 벡터와 분산-공분산 행렬을 갖는 다변량 정규분포로부터 생성하며,

$$\text{평균벡터} = \begin{pmatrix} 0.6\eta_1 \\ 0.4\eta_1 \end{pmatrix},$$

$$\text{분산-공분산행렬} = \begin{pmatrix} 0.2 & 0.14 \\ 0.14 & 0.2 \end{pmatrix}$$

η_2 에 대한 추정변수 y_3, y_4 는 아래와 같은 평균 벡터와 분산-공분산 행렬을 갖는 다변량 정규분포로부터 생성한다.

$$\text{평균벡터} = \begin{pmatrix} 0.2\eta_2 \\ 0.5\eta_2 \end{pmatrix},$$

$$\text{분산-공분산행렬} = \begin{pmatrix} 0.2 & 0.14 \\ 0.14 & 0.2 \end{pmatrix}$$

마지막으로 브랜드 선택확률 P_1, P_2, P_3 를 아래로부터 생성한다.

$$P_i = \frac{\exp(\eta_i^B)}{1 + \exp(\eta_1^B) + \exp(\eta_2^B)} + u_i, \quad i = 1, 2$$

$$P_3 = \frac{1}{1 + \exp(\eta_1^B) + \exp(\eta_2^B)} + u_3$$

여기서 u_i 는 (0, 0.1)의 구간으로부터 랜덤으로 생성한다. 위의 확률은 합이 1이 되도록 조정한다.

위에서 생성한 외생/내생 잠재변수 값은 모형의 추정에 사용하지 않으며 단지 추정변수인 x_i ($i = 1, 2, 3$), y_i ($i = 1, 2, 3, 4$), P_i ($i = 1, 2, 3$) 만을 사용한다. 표본크기 (관측 수)는 50, 100, 500, 1000의 네 가지를 고려하며, 각각의 경우에 대해 시뮬레이션 정확도를 높이기 위해 500번씩 반복적으로 실험을 수행한 후 평균값으로 결과를 얻는다.

5.2 실험결과 분석

생성된 데이터로부터 추정된 모형의 성능은 실제로 가정된 잠재변수간의 모형에 가까운 정도로 평가할 수 있다. 위에서 언급하였듯이 실험결과는 가정된 모형의 계수 또는 가중치와 추정값의 비교를 위해 제곱근오차(root mean squared error; RMSE)

Table 3. 실험 I-1 결과 ($\sigma_c^2 = 0.1$)

	관측 수	RMSE		
		내생잠재변수 계수	브랜드잠재변수 계수	추정변수 가중치
제안 방법	50	0.1217	0.2171	0.0882
	100	0.0949	0.2094	0.0832
	500	0.0592	0.2018	0.0789
	1000	0.0548	0.2015	0.0786
대안	50	0.1285	0.2195	0.1195
	100	0.0964	0.2100	0.1179
	500	0.0576	0.2029	0.1172
	1000	0.0517	0.2026	0.1170

Table 4. 실험 I-2 결과 ($\sigma_c^2 = 0.2$)

	관측 수	RMSE		
		내생잠재변수 계수	브랜드잠재변수 계수	추정변수 가중치
제안 방법	50	0.1325	0.2051	0.0938
	100	0.0990	0.1910	0.0894
	500	0.0620	0.1824	0.0849
	1000	0.0534	0.1806	0.0856
대안	50	0.1358	0.2061	0.1260
	100	0.1011	0.1948	0.1251
	500	0.0617	0.1836	0.1238
	1000	0.0533	0.1827	0.1243

로 표현한다. 결과에 내생잠재변수 모형의 계수와 브랜드 잠재변수 모형의 계수를 구분하여 보고한다. 이 때 계수의 수에 차이가 있으므로 계수수를 나누어 RMSE를 산출한다. 따라서 RMSE는 하나의 계수에 해당하는 평균 오차로 해석될 수 있다. 실험의 종류는 총 4가지가 되며, 실험 I-1은 실험 I에 의하여 추정변수들을 생성하고 브랜드 잠재변수 생성 시 Case 1 ($\sigma_c^2 = 0.1$)의 사용을 의미한다. 다른 실험 번호도 유사하게 설명된다.

<Table 3>은 실험 I-1의 결과를 나타낸 것이며, <Table 4>는 실험 I-2의 결과를 정리한 것이다. <Table 3>과 <Table 4>에서 보듯이, 제안한 방법이 대안에 비해 성능이 다소 좋은 것을 알 수 있다. 특히 추정변수 가중치 추정면에서 제안한 방법이 대안에 비해 월등함을 볼 수 있다. 제안한 방법은 관측수가 커질수록 그 성능이 좋아 짐을 또한 볼 수 있다. 특히, 내생잠재변수의 계수 추정면에서 큰 변화가 있으며, 브랜드잠재변수 계수 또는 추정변수 가중치 추정면에서는 관측수 증대가 큰 영향을 주지 않음을 볼 수 있다.

<Table 5>는 실험 II-1의 결과를 나타낸 것이며, <Table 6>은 실험 II-2의 결과를 정리한 것이다. 앞에서 언급한 바와 같이 실험 II는 추정변수 생성시 실험 I에 비하여 분산이 증대되었다. 이에 따라 계수 추정과 연관된 RMSE 또한 커짐을 볼 수 있

Table 5. 실험 II-1 결과 ($\sigma_{\epsilon}^2 = 0.1$)

	관측 수	RMSE		
		내생잠재 변수 계수	브랜드잠재변 수 계수	측정변수 가중치
제안 방법	50	0.1969	0.2256	0.0826
	100	0.1749	0.2153	0.0668
	500	0.1640	0.2097	0.0542
	1000	0.1597	0.2092	0.0528
대안	50	0.2137	0.2262	0.1188
	100	0.1966	0.2184	0.1169
	500	0.1860	0.2115	0.1146
	1000	0.1816	0.2101	0.1145

Table 6. 실험 II-2 결과 ($\sigma_{\epsilon}^2 = 0.2$)

	관측 수	RMSE		
		내생잠재 변수 계수	브랜드잠재변 수 계수	측정변수 가중치
제안 방법	50	0.2021	0.2105	0.0841
	100	0.1789	0.2010	0.0720
	500	0.1646	0.1923	0.0596
	1000	0.1603	0.1913	0.0582
대안	50	0.2298	0.2128	0.1252
	100	0.2084	0.2039	0.1210
	500	0.1870	0.1940	0.1206
	1000	0.1862	0.1924	0.1204

다. 다만, 측정변수 가중치의 경우는 관측치가 많은 상태에서는 실험 I에 비해 RMSE가 감소함은 흥미로운 사실이다.

6. 결론

본 연구에서는 고객만족 및 고객자산 등이 브랜드 선택확률에 미치는 영향을 분석하기 위하여 구조방정식 모형을 제시하며 모형의 추정을 위한 PLS 기반 알고리즘을 제안하였다. 본 모형에는 확률이 개입됨에 따라 기존의 알고리즘을 그대로 적용하기 어렵기 때문에 다항로짓 모형을 활용하도록 기존의 알고리즘을 수정 제안하였다.

제시된 모형 및 알고리즘을 이동통신사의 재무성과 분석에 적용하여 결과를 얻었으며, 보다 객관적인 알고리즘의 성능평가를 위해 가상의 모형 및 생성된 데이터를 바탕으로 시뮬레이션을 수행하였다. 시뮬레이션 결과는 가정된 실제 계수와

제안된 알고리즘으로 추정된 계수들을 비교하여 나타내었는데 대체로 합리적인 성능을 보이고 있는 것으로 판단되었다.

본 연구에서는 브랜드별 선택확률이 표본데이터로 얻어지는 것을 가정하고 있으나, 만약 특정 브랜드의 선택 여부가 지시변수 (0 또는 1)로 주어지는 경우에는 제안된 알고리즘을 수정할 필요가 있다. 따라서 이에 대한 추후 연구가 필요하다고 하겠다.

참고문헌

- Bagozzi, R. P. (1980), *Causal Models in Marketing*, John Wiley and Sons, New York.
- Berger, P. D. and Nasr, N. I. (1998), Customer lifetime value : marketing models and applications, *Journal of Interactive Marketing*, **12**(Winter), 17-30.
- Blattberg, R. C. and Deighton, J. (1996), Manage marketing by the customer equity test, *Harvard Business Review*, **74**(4), 136-144.
- Fornell, C. and Bookstein, F. L. (1982), Two structural equation models : LISREL and PLS applied to consumer exit-void theory, *Journal of Marketing Research*, **19**, 440-452.
- Fornell, C., Johnson, M. D., Anderson, E. W., Cha, J., and Bryant, B. E. (1996), The American customer satisfaction index : nature, purpose, and findings, *Journal of Marketing*, **60**, 7-18.
- Fornell, C. and Larcker, D. F. (1981), Evaluating structural equation models with unobservable variables and measurement error, *Journal of Marketing Research*, **18**, 39-50.
- Guinot, C., Latreille, J. and Tenenhaus, M. (2001), PLS path modeling and multiple table analysis, Application to the cosmetic habits of women in Ile-de-France, *Chemometrics and Intelligent Laboratory Systems*, **58**, 247-259.
- Hellier, P. K., Geursen, G. M., Carr, R. A., and Rickard, J. A. (2003), Customer repurchase intention : a general structural equation model, *European Journal of Marketing*, **37**, 1762-1800.
- Hulland, J. (1999), Use of partial least squares (PLS) in strategic management research : a review of four recent studies, *Strategic Management Journal*, **20**, 195-204.
- Jöreskog, K. A. and Sörbom, D. (1989), *LISREL 7 User's Reference Guide, 1st ed.*, Scientific Software, Mooresville, IN, USA.
- Martensen, A., Gronholdt, L., and Kristensen, K. (2000), The drivers of customer satisfaction and loyalty : cross-industry findings from Denmark, *Total Quality Management*, **11**, S544-553.
- Noonan, R. and Wold, H. (1982), "PLS path modeling with indirectly observed variables : A comparison of alternative estimates for the latent variable," In Joreskog, K., and H. Wold, (Eds.), *Systems under indirect observation : Causality, structure, prediction*, Amsterdam : North Holland Publishing, 1-54.
- Rust, R. T., Lemon, K. N., and Zeithaml, V. A. (2004), Return on marketing : using customer equity to focus marketing strategy, *Journal of Marketing*, **68**, 109-127.
- Tenenhaus, M., Vinzi, V. E., Chatelin, Y., and Lauro, C. (2005), PLS path modeling, *Computational Statistics and Data Analysis*, **48**, 159-205.