

# 자동차 시장 내 최적의 리베이트 수준 결정을 위한 계층 베이지 분산구성 모형

Hierarchical Bayes Variance Components Model for Developing Optimal Rebate Program in the Automotive Industry

장광필 • Chang, Kwangpil

이 연구의 목적은 포장 소비재 시장에서의 A.C. Nielsen 패널 자료와 유사한 수준의 개별적 자동차 거래 자료(disaggregate transaction data)를 사용하여 마케팅 관리적 차원에서 지역별로 축적된 자료의 구조에 적합한 모형화 방식을 통해 최적의 리베이트 수준을 결정하는 방법론을 제시하는 것이다. 기존의 계층 베이지 모형의 적용은 1단계에서 최소의 분석단위인 소비자의 브랜드 선택을 로짓(logit)이나 프로빗(probit)모형의 선택확률로 모형화하고 2단계에서 소비자별 모수( $\beta^i$ )의 선행분포(prior)를  $MVN(\mu, \Sigma)$ 의 다변량 정규분포로 가정하고 3단계에서 선행분포의 모수  $\mu$ 와  $\Sigma$ 의 초 선행분포(hyper prior)를 각 각 다변량 정규분포와 Wishart분포로 가정하는 방식이었다. 이러한 3단계 모형화 방식으로 추정된 소비자별 모수를 의사결정의 근거로 사용하였다. 이 논문에서는 기존의 3단계 베이지 모형을 자동차 시장에 응용하고 확장하여 1단계에서 최소의 분석단위를 zip code로 가정하여 zip code별 소비자의 자동차 모델선택을 로짓모형의 선택확률로 모형화하고, 2단계에서 zip code별 모수  $\beta_r^i$ 의 분포를 지역(regional) 평균을 중심으로 분포된다는 가정 하에  $MVN(\mu_r, \Sigma_r)$ 의 다변량 정규분포로 가정하고, 3단계에서 지역평균  $\mu_r$ 의 분포를 전국(national) 평균을 중심으로 분포된다는 가정 하에  $MVN(\mu_n, \Sigma_n)$ 의 다변량 정규분포로 가정하고, 4단계에서  $\mu_n$ 과  $\Sigma_n$ 의 초 선행 분포를 각 각 다변량 정규분포와 Whishart분포로 가정하였다. 제안된 4단계 계층 베이지 모형은 분산구성 모형(Variance Components Model)으로서 zip code별 모수의 이질성을 지역내 이질성과 지역간 이질성으로 분해하여 반영함으로써 실증분석 결과 모형적합도와 예측타당성 측면에서 기존의 3단계 모형에 비해 우월한 것으로 나타났다. 실증분석 대상은 미국 내 미니 밴 시장으로 하였고 다양한 프로모션 수단 중에 가장 빈번히 사용되는 리베이트를 중심으로 zip code별로 서로 다른 리베이트 모수를 추정하여 최적의 리베이트 수준을 추정하였다. 시뮬레이션을 통해 관행적으로 집행되어온 일률적인 리베이트(uniform blanket rebate)보다 서로 다른 zip code별로 차별적 리베이트를 제공했을 때 수익성이 높다는 것을 실증적으로 보였으며, 3 단계 모형보다 4단계의 분산구성 모형에 근거한 의사결정이 보다 높은 수익을 제공하는 것으로 나타났다.

핵심주제어: 선택모형, 리베이트, 자동차산업, 베이지안 모형

이 논문은 2003년도 한국학술진흥재단의 지원에 의하여 연구되었음(KRF-2003-003-B00213)

장 광 필 | 서울시립대학교 경영학부 부교수(kpchang@uos.ac.kr)

## ABSTRACT

The current study attempts to design optimal rebate program for manufacturers in the automobile industry, who have to determine the combinations of incentives to be used (e.g., consumer rebates and dealer incentives) in a wide variety of automobile makes, models and regional markets. Allocating promotional budget to regional offices is further complicated by the need to evaluate and react to the conflicting information provided by different district managers, each pushing for a greater slice of the promotional budget.

This study develops an approach to capture zip code-level response to the marketing mix that is based on a Hierarchical Bayes multinomial logit model (random coefficient model) of automobile choice and show how it can be used for developing optimal rebate program in the automobile market. The approach requires us to hierarchically structure the data according to geographical proximity, e.g., zip codes under a particular region (e.g., DMA: Designated Marketing Area) to account for similarities within a region and differences across regions. In contrast to scanner panel data, the long inter-purchase times in this category provide only one observation per buyer in the sample, which makes it difficult to directly apply previously developed methods to account for consumer heterogeneity. We overcome this limitation by treating the transactions from a particular zip code as a “purchase string” and estimating choice model parameters at a zip code level using the MCMC method. This permits us to implement a micro approach to the development of the marketing mix such that responses to the marketing mix reflect differences in consumer price and promotion sensitivities across geographical areas.

In this approach the zip code level parameters in a region are distributed according to a normal distribution that is centered on a region-specific mean, with each regional mean, in turn, being distributed around a national mean, which is drawn from the distribution specified by the hyper-parameters. The priors for the variance covariance matrices of the parameters also follow the same hierarchical, set-up. We label this set up as the NRZ (National-Regional-Zip code) model. In this model, the first stage of the hierarchical structure, is a multinomial logit choice model. In the second stage we specify the multivariate normal distribution  $MVN(\mu_r, \Sigma_r)$  for the zip code-level parameters in the region  $r$ . In the third stage we specify the multivariate normal distribution  $MVN(\mu_n, \Sigma_n)$  for region-specific means,  $\mu_r$  and the Wishart distribution for  $\Sigma_r$ . In the final (fourth) stage we specify the hyper prior distribution for  $\mu_n$  and  $\Sigma_n$ . This four-stage approach results in the Variance Components model that accounts for both between-region ( $\Sigma_n$ ) and within-region heterogeneity ( $\Sigma_r$ ).

The data for this study come from Power Information Network (PIN), an affiliate of J.D. Power and Associates. PIN collects sales transaction data from a sample of dealerships in the major

metropolitan areas in the U.S. These are retail transactions, i.e., sales or leases to final consumers, excluding fleet sales. Each observation in the PIN database contains the transaction date, the manufacturer, model year, make, model, trim and other car information, the transaction price, consumer rebates, the interest rate, term, amount financed (when the vehicle is financed or leased), etc. We use data for 2000 model-year minivans for the period October 1999 - September 2000 (52 weeks). Fifteen models are included in the study: Chevrolet Astro, Dodge Caravan, Dodge Grand Caravan, Chrysler Grand Voyager, Pontiac Montana, Honda Odyssey, Nissan Quest, GMC Safari, Toyota Sienna, Oldsmobile Silhouette, Chrysler Town&Country, Chevrolet Venture, Mercury Villager, Chrysler Voyager and Ford Windstar. These models in the study account for 96% of category (i.e., minivan segment) unit sales.

Empirical results show that the proposed four-stage model (NRZ) outperformed a typical three-stage model (NZ) that accounts for only within-region heterogeneity (by assuming that all zip codes belong to one big nation) in terms of model fit in both calibration and holdout samples. The superiority of the NRZ model can be attributable to more thorough corrections of heterogeneity which is due to between- and within-differences in responses to the marketing mix. The managerial setting that we study provides some natural ways to structure the prior distribution of the zip code level parameters to accommodate their differences and similarities. Comparing the empirical performance of competing models based on different prior distributions provides insights into the best ways to borrow information across the different zip codes.

The approach permits managers the flexibility to use different promotional instruments in different ways. Thus, a general promotional incentive, such as manufacturer cash-back promotion, can be constrained to be equal across different zip codes in the same region to address antitrust concerns, while a targeted coupon mailing could be customized for a particular zip code.

We examine a realistic (but hypothetical) problem faced by one particular manufacturer, which involves setting the optimal rebate levels to be offered in various markets. First, we determine the optimal rebate level according to each model assumption (i.e., homogeneous MNL model, NRZ model assumptions). Given the model-specific optimal rebate level, we calculate expected profits by using the NRZ model parameters since in terms of model fit, the NRZ model is the best and most likely to describe market responses close to reality. We show how the model parameters can be used to determine these rebate levels and demonstrate that the profits obtained with the NRZ model assumption are significantly higher than those with the uniform blanket rebate program.

**Key words:** Choice models, Promotions, Automobiles, Hierarchical Bayes.

## 1. 서론

대부분의 기업에서 세일즈 프로모션은 소비자의 구매 촉진을 위한 중요한 수단으로 사용된다. 한 예로서, 미국 자동차 시장의 경우 년 간 400억불 정도의 금액이 소비자 리베이트, 달러 인센티브, APR(이자율)인하 등에 사용되는 것으로 보고되고 있다. 따라서 각종 프로모션에 투입되는 예산규모로 보아 자동차 시장의 경우 예산의 효율적 운용이 기타 산업에 비해 더욱 절실히 요구되며, 수익에 미치는 영향 역시 클 것으로 예상된다. 그러나 국내 시장은 물론이고 미국 시장에서도 리베이트 프로그램이나 파이낸싱 조건 등이 거주 지역별(zip code별) 소비자 반응의 이질성을 고려하지 않고 일률적으로 집행되거나, 차별적 집행 역시 계량화된 소비자 민감도에 근거한다기보다 마케팅 관리자의 직관 내지는 업계관행에 따라 집행되는 경우가 일반적이다. 이러한 관행의 원인은 가격이나 리베이트에 대한 소비자 민감도를 파악할 만한 개별적 거래 자료(disaggregate transaction data)가 가용하지 않은 측면도 있었지만, 지역별 거래량의 차이 때문에 모든 자료를 합하여(pooling) 평균적인 소비자 민감도를 추정한 결과이기도 하다.

지난 20년 간, 포장 소비재(frequently purchased consumer packaged goods)시장에서의 가격이나 판촉에 대한 소비자 반응 또는 민감도에 대한 연구는 널리 알려진 A.C. Nielsen의 패널 자료를 통해 진행되어 왔다(Guadagni and Little 1983, Gupta 1988, Chintagunta 1993). 이러한 연구를 확장해서 Tellis and Zufryden (1995)과 Silva-Risso, Bucklin and Morrison(1999)은 프로모션 칼렌다를 기획하는 의사결정 지원모델(decision support model)을 제안했다. 이와 더불어 소비자 반응

의 이질성(response heterogeneity)을 세그먼트 수준(Kamakura and Russell 1989, Bucklin, Gupta and Siddarth 1998)과 가구 수준(household-level, 예를 들어, Rossi, McCulloch, and Allenby 1996, Chang, Siddarth and Weinberg 1999)에서 반영하는 모형도 제안되었다. 특히, Rossi, McCulloch, and Allenby(1996)가 제안한 모형은 가구별 가격민감도의 차이를 고려해서 가구별로 차별적인 판촉을 기획할 수 있는 가능성을 보여주었고, Ansari and Mela (2003)는 인터넷상에서 그 모형을 적용하는 예를 제시하였다. 반면에, 자동차와 같은 내구재 시장에서의 가격 및 판촉에 대한 소비자 민감도에 대한 연구는 상대적으로 적은 편이다. Colombo and Morrison(1989)이 상표전환 행렬을 사용하여 자동차 시장에서의 경쟁관계를 분석한 바 있지만 방법론상 마케팅믹스 변수를 포함하지 않은 분석이었다. Berry, Levinsohn and Pakes(1995)와 Sudhir(2001)도 최근에 자동차 시장 내에서의 소비자 가격민감도를 추정했지만, 총합 거래 자료(aggregate transaction data)를 사용했기 때문에 효율적인 프로모션 기획에 대한 실질적인 제안을 하지 못했다. 장광필(2003) 역시 자동차 거래 자료로부터 고려상표군을 반영한 시장구조를 추론하는 방법론을 제시했지만 프로모션 기획에 대한 내용은 다루지 않았다.

이 연구의 목적은 포장소비재 시장에서의 A.C. Nielsen 패널자료와 유사한 수준의 개별적 거래 자료(disaggregate transaction data)를 사용하여 마케팅 관리적 차원에서 지역별로 축적된 자료의 구조에 적합한 모형화 방식을 통해 zip code별 마케팅믹스에 대한 소비자 민감도의 차이를 반영한 최적의 리베이트 수준을 결정하는 방법론을 제시하는 것이다. 패널 자료를 사용한 기존의 베이스 모형의 적용은 대부분의 경우

1 Annual Percentage Rate로서 미국 내 자동차회사의 경우 자동차 구매 시 financing을 해주는 회사를 소유하고 이자율을 조정함으로써 APR을 촉진수단으로 사용하고 있다

2 1999년 9월부터 2000년 10월까지 미국 미니 밴 시장에서의 리베이트 수준은 1,500불에서 2,000불 사이로 나타났으며, 대부분의 자동차메이커가 그러한 수준에서 리베이트를 제공했다.

3단계로 모형화 하였다. 즉, 1단계에서 최소의 분석단위인 가구(household)의 브랜드 선택을 로짓(logit)이나 프로빗(probit)모형의 선택확률로 모형화하고 2단계에서 가구별<sup>3</sup> 모수( $\beta^h$ )의 선행분포(prior distribution)를  $MVN(\mu, \Sigma)$ 의 다변량 정규분포로 가정하고, 3단계에서 선행분포의 모수  $\mu$ 와  $\Sigma$ 의 초 선행분포(hyper prior distribution)를 각 각 다변량 정규분포와 Wishart분포로 가정하는 방식이었다. 이 연구에서는 그러한 방식을 응용하고, 확장하여 자동차 시장에서의 거래 자료에 적합한 형태로 모형화를 시도하였다. 기존의 패널 자료를 사용한 베이스 모형이 가족 구성원의 구매 자료를 기초로 하여 가구를 최소의 분석단위로 사용한 것처럼, 이 연구에서는 zip code를 최소의 분석단위로 가정하고 같은 zip code내의 고객별 거래 자료를 기초로 하여 모형화를 시도하였다. 기존의 베이스 모형의 3단계 모형화와 다른 부분은 zip code별 모수의 선행분포를 한 단계에서 두 단계로 확장하여 모두 4단계로 모형화한 것이다 (i.e. a four-stage hierarchical model, p. 241 in Carlin and Louis 2000). 부연 설명하면, 전국을 몇 개의 지역 (region), 예를 들면, DMA(Designated Marketing Area)로 구성되어 있다고 가정하고, zip code별 모수가 각 지역 평균을 중심으로 분포되고, 지역 평균은 다시 전국평균을 중심으로 분포되어 있다고 가정하는 것이다. 4단계 모형은 분산구성 모형 (Variance Components Model)으로서 zip code별 모수의 이질성을 지역내 이질성과 지역간 이질성으로 분해하여 반영하는 특징을 갖는다. 3단계 모형과 4단계 분산구성 모형과의 구체적인 차이점에 대하여는 II장 연구모형을 통하여 보다 자세히 설명하기로 한다.

제안된 모형의 실증분석 대상은 미국 내 미니 밴 시장으로 하고 다양한 프로모션 수단 중에 가장 빈번히 사용되는 리베이트를 중심으로 zip code별로 서로

다른 리베이트 모수를 추정하여 최적의 리베이트 수준을 결정하는 방법론을 제시하고자 한다. 현재까지 국내에서는 A. C. Nielsen 패널자료와 유사한 수준의 자동차 거래 자료가 수집되고 있지 않지만, 미국 내에서는 그러한 자료가 수집되고 실제 프로모션 기획에 사용되고 있는 실정인 바 (예를 들어, 미국의 J. D. Power & Associates사가 전국의 딜러로부터 수집하는 자동차 거래 자료의 분석을 통해 기아 자동차를 비롯한 미국의 우수 자동차 메이커의 현지 소비자 인센티브 수준결정), J. D. Power & Associates사의 자료를 지원받아 제안된 모형의 타당성을 검증하고자 한다. 시뮬레이션을 통해 관행적으로 집행되어온 일률적인 리베이트(uniform blanket rebate)와 서로 다른 zip code별 차별적 리베이트를 제공했을 때 수익성의 차이를 보여줌으로써 소비자 민감도에 따른 가격차별의 중요성에 주의를 환기시키고자 한다.

이 논문의 구성은 다음과 같다. 제 II장에서는 연구모형을 제안하고, 제 III장에서는 실증분석을 실시한다. 제 IV장에서는 실증분석 결과를 설명하고, 마지막 V장에서는 결론과 함께 향후 연구방향을 제시한다.

## II. 연구모형

전국적인 리베이트 프로그램을 시행해야 하는 마케팅 의사결정자의 경우 소비자의 리베이트에 대한 민감도의 차이를 고려하여 리베이트 수준을 결정하는 것이 이상적일 것이다. 예를 들어, 전국이 20여개의 DMA로 구분되고, 각 DMA는 수십 개의 zip code로 구성된다고 하면, 첫 번째로 직면하게 되는 문제는 최소의 분석단위를 zip code로 가정할 경우 zip code별로 서로 다른 소비자 민감도(즉, 선택모형의 가격이나 리

3 패널자료를 사용할 경우 패널의 구매 자료는 사실 패널이 속한 가구(household) 구성원의 선호를 반영한 구매 자료이기 때문에 소비자별이란 표현보다는 가구별이란 표현이 보다 정확하다고 볼 수 있다. 따라서 가구별 모수를 추정한다는 의미는 가족 구성원의 민감도는 동일한 것으로 가정하는 것이다.

베이트 같은 변수의 모수)의 추정이 가능한가의 문제이다. 기존의 계량경제학적인 방법은 특정 zip code내에서의 거래 자료가 부족할 경우 신뢰할 만한 민감도의 추정이 어렵기 때문에 다른 zip code의 거래 자료와 통합하여 지역별 평균 민감도를 추정하는 것이 일반적이었다. 이 경우 zip code별 민감도의 차이는 무시되고, 지역별 차이만 반영하게 된다. 이러한 문제점을 극복하고 최소의 분석단위, 즉, zip code수준의 민감도를 추정할 수 있는 방법으로서 계층 베이스 모형(Hierarchical Bayes Model, e.g., Rossi, McCulloch and Allenby 1996, Chang, Siddharth and Weinberg 1999)을 적용할 수 있다. 서론에서 언급했듯이 기존의 계층 베이스 모형을 자동차 거래 자료에 3단계로 적용한다면, 1단계에서 zip code별 소비자의 자동차 선택행동을 로짓모형의 선택확률로 모형화하고, 2단계에서 zip code별 모수  $\beta_r^*$ 의 분포를 전국(national) 평균을 중심으로 분포된다는 가정 하에  $MVN(\mu, \Sigma)$ 의 다변량 정규분포로 가정하고, 3단계에서  $\mu$ 와  $\Sigma$ 의 초 선택분포를 각각 다변량 정규분포와 Wishart분포로 가정할 수 있다(이후 이런 방식의 3단계 모형을 NZ모형으로 명한다.). 3단계 모형화의 다른 대안으로서 이 연구에서 제안하는 모형은 zip code수준의 모수는 지역 평균을 중심으로 분포되고, 지역 평균은 다시 전국 평균을 중심으로 분포된 구조를 갖는 모형이다. 1단계에서 zip code별 소비자의 자동차 선택행동을 로짓모형의 선택확률로 모형화하고, 2단계에서 zip code별 모수  $\beta_r^*$ 의 분포를 지역(regional) 평균을 중심으로 분포된다는 가정 하에  $MVN(\mu_r, \Sigma_r)$ 의 다변량 정규분포로 가정하고, 3단계에서 지역평균  $\mu_r$ 의 분포를 전국(national) 평균을 중심으로 분포된다는 가정 하에  $MVN(\mu_n, \Sigma_n)$ 의 다변량 정규분포로 가정하고, 4단계에서  $\mu_n$ 과  $\Sigma_n$ 의 초 선택분포를 각각 다변량 정규분포와 Wishart분포로 가정하는 것이다(이후 이런 방식의 4단계 모형을 NRZ모형으로 명한다.).

자동차 시장의 경우, 포장 소비재와는 달리, 구매주기가 평균 6년 이상이 되기 때문에 1년간의 거래 자료상에는 소비자당 최대 1건의 거래 자료가 기록된다. 따라서 소비자별 마케팅믹스에 대한 민감도를 추정하기에는 어려움이 있기 때문에 선행연구(Steenburgh, Ainslie and Engebretson 2003)에서 제안한 대로 최소의 분석단위로서 zip code수준이 적절하다고 판단된다. 즉, 같은 zip code에 거주하는 소비자는 동일한 민감도를 갖는 것으로 가정하였다. zip code를 최소의 분석단위로 하는 것은 지리적 세분화의 원칙에도 부합되고, 각종 센서스자료나 인구통계학적 특성이 zip code수준에서 요약된 점을 고려할 때 정당화될 수 있다고 판단된다. 그러나 같은 zip code에 거주하는 소비자의 거래를 총합한다고 해도, zip code별로 거래량, 즉, 정보의 양은 차이가 날 수 밖에 없기 때문에 부족한 거래량을 갖는 zip code의 경우 다른 zip code의 정보를 통계적으로 차용할 수 있는 절차가 필요하게 된다. 가장 적절한 방법으로는 Bayesian random effects framework를 적용하는 것이다.

Bayesian random effects framework를 적용하는 방식에 있어서 이 연구에서 제안하는 NRZ모형과 기존의 NZ모형을 고려해 볼 수 있다. 한 지역(region)의 마케팅책임자가 자동차 리베이트 액수를 결정할 때 그 지역의 거래 자료뿐만 아니라 다른 지역의 자료를 모두 포함하여 zip code별 모수를 추정하고 zip code별 최적의 리베이트를 결정하는 방식이다. NZ모형은 zip code별 모수가 전국 평균을 중심으로 분포되어 있다고 가정함으로써 다른 zip code의 정보를 전국 평균을 통하여 공유하며 zip code간 이질성은 전국 수준의 공분산 행렬에 의해 반영한다. 반면에 NRZ모형은 zip code별 모수가 지역 평균을 중심으로 분포되고, 지역 평균은 다시 전국 평균을 중심으로 분포되어 있다고 가정하는 구조이다. NZ모형이 zip code간 이질성을 전국수준의 공분산 행렬 하나로 반영하는 반면에 NRZ모형은

zip code가 속한 지역 정보를 통해 지역내 이질성과 지역간 이질성으로 분해하여 반영하는 차이점을 가진다 (i.e., 분산구성 모형). 다음은 계층 베이스 모형의 가장 전형적인 형태인 NZ모형과 이 연구에서 제안하는 NRZ모형을 수식을 통하여 비교하고 차이점에 대하여 자세히 설명하였다.

## 1. NZ모형:

선행연구에서 3단계로 적용한 계층 베이스 모형의 전형적인 형태로서 zip code수준의 모수는 전국 평균을 중심으로 분포되어 있다고 가정한다.

1 단계: zip code  $z$ 에 거주하는 소비자  $c$ 가 구매시점  $t$ 에 자동차 모델  $j$ 를 선택할 확률은 다항로짓모형을 적용하여 다음과 같이 정의한다. (NRZ모형과 동일).

$$P_t^c(j | \beta_r^z, \mathbf{X}_{jt}^z) = \frac{\exp(\beta_r^z \mathbf{X}_{jt}^z)}{\sum_k \exp(\beta_r^z \mathbf{X}_{kt}^z)} \quad (1)$$

여기서,  $\mathbf{X}_{jt}^z$ 는 소비자  $c$ 의 자동차 모델선택에 영향을 미칠 수 있는 충성도변수, 즉, 메이커 충성도, 모델 충성도 등과 마케팅믹스 변수로서, 가격, 리베이트, APR 등이 포함된다.  $\beta_r^z$ 는 지역  $r$ 에 속한 zip code  $z$ 에 거주하는 소비자의  $\mathbf{X}_{jt}^z$ 에 대한 민감도(즉, 가격민감도, 리베이트민감도 등을 의미한다)를 나타낸다.

2 단계:  $\beta_r^z$ 는 전국 평균,  $\mu$ 를 중심으로 다음과 같은 다변량 정규분포를 갖는다고 가정한다.

$$\beta_r^z = \mu + \epsilon_r^z, \quad \epsilon_r^z \sim MVN(\mathbf{0}, \Sigma) \quad (2)$$

3 단계:  $\mu$ 과  $\Sigma^{-1}$ 은 각각 다음과 같은 분포를 갖는

다고 가정한다.

$$\mu \sim MVN(\boldsymbol{\eta}, \mathbf{C}) \quad (3)$$

$$\Sigma^{-1} \sim Wishart((\rho \mathbf{R})^{-1}, \rho) \quad (4)$$

NZ모형에서는 전국 평균을 중심으로 zip code별 모수가 분포되고 zip code간 이질성은  $\Sigma$ 를 통하여 반영된다.

## 2. NRZ모형:

선행연구에서 적용한 3단계 계층 베이스 모형과는 달리 4단계의 구조를 갖는, 보다 일반화된 모형으로서 다음과 같은 계층 구조를 갖는다.

1 단계: 지역  $r$ 에 속한 zip code  $z$ 에 거주하는 소비자  $c$ 가 구매시점  $t$ 에 자동차 모델  $j$ 를 선택할 확률은 다항로짓(Multinomial Logit)모형을 적용하여 다음과 같이 정의한다.

$$P_t^c(j | \beta_r^z, \mathbf{X}_{jt}^z) = \frac{\exp(\beta_r^z \mathbf{X}_{jt}^z)}{\sum_k \exp(\beta_r^z \mathbf{X}_{kt}^z)} \quad (5)$$

2 단계:  $\beta_r^z$ 는 지역 평균(민감도),  $\mu_r$ 을 중심으로 다음과 같은 다변량 정규분포를 갖는다고 가정한다.

$$\beta_r^z = \mu_r + \epsilon_r^z, \quad \epsilon_r^z \sim MVN(\mathbf{0}, \Sigma_r) \quad (6)$$

지역내 이질성은  $\Sigma_r$ 에 의해 반영된다.

3 단계:  $\mu_r$ 는 전국 평균(민감도),  $\mu_n$ 을 중심으로 다음과 같은 다변량 정규분포를 갖고,  $\Sigma_r^{-1}$ 은 다음과 같은 Wishart분포를 갖는다고 가정한다.

$$\mu_r = \mu_n + \delta_r, \quad \delta_r \sim MVN(0, \Sigma_n) \quad (7)$$

$$\Sigma_r^{-1} \sim Wishart((\rho_n \Sigma_n)^{-1}, \rho_n) \quad (8)$$

여기서  $\rho_n$ 은 Wishart분포의 자유도를 나타낸다. 지역간 이질성은  $\Sigma_n$ 에 의해서 반영된다.

4 단계:  $\mu_n$ 과  $\Sigma_n^{-1}$ 은 각각 다음과 같은 초 선택분포를 갖는다고 가정한다.

$$\mu_n \sim MVN(\eta, C) \quad (9)$$

$$\Sigma_n^{-1} \sim Wishart((\rho R)^{-1}, \rho) \quad (10)$$

여기서  $\eta$ ,  $C$ ,  $\rho$ ,  $\rho_n$ ,  $R$ 은 초 선택(hyper prior)분포의 모수로서 사후분포가 적절(proper)하고, 3단계 이하의 모수들을 추정할 때 선택분포의 영향을 최소화하도록(diffuse) 정해진다 (Gelfand et al. 1990, Allenby et al. 1998, Chang et al. 1999).

NRZ모형과 NZ모형의 차이점을 설명하기 위해 수식 (7)을 수식 (6)에 대입하면 다음과 같다.

$$\beta_r^z = \mu_n + \delta_r + \epsilon_r^z, \quad \beta_r^z \sim MVN(\mu_n, \Sigma^*) \quad (11)$$

$$\Sigma^* = \Sigma_n + \Sigma_r \quad (12)$$

NZ모형의 수식 (2)와 NRZ모형의 수식 (11)은 동일한 형태로 표현된다. 그러나 중요한 차이는 수식 (12)에 나타난 바와 같이 공분산 행렬이  $\Sigma_n$ (지역간 이질성)와  $\Sigma_r$ (지역내 이질성)으로 분해되는 것이다. 즉 NRZ모형에서 zip code의 이질성은 zip code가 속한 지역내 이질성과 지역간 이질성을 모두 반영하는 것이다. 이러한 특징 때문에 NRZ모형과 같은 4단계 모형을 분산구성 모형 (variance components model)으로 부른다 (Carlin and Louis 2002). NRZ모형의 추정방법은

부록에 제시된 바와 같다.

### 3. HMNL(Homogeneous Multinomial Logit:

동질적인 다항로짓)모형: 위에 언급한 계층 베이스 모형(NRZ와 NZ모형)과 비교되는 모형으로서 모든 zip code의 모수(즉 민감도)는 동일하다고 가정하는 모형이다. 모든 zip code별 자료를 총합(pool)하여 공통모수를 추정하는 모형이다. 계층 베이스 모형이나 Latent Class모형이 소비자 민감도의 차이를 반영한 모형이라면, HMNL모형은 소비자 민감도의 이질성을 무시한 모형이라고 할 수 있다.

## III. 실증분석

### 1. 자료

이 연구에 사용된 자료는 미국 내 자동차업계 전문조사 연구기관인 J. D. Power & Associates사로부터 제공 받은 것이다. 이 기관은 미국 전역의 딜러로부터 자동차 거래정보(자동차가격, 리베이트액수, APR, Residual value, 고객의 인구통계학적 정보, 리스/파이낸싱 같은 거래조건을 포함한 약 200개의 변수)를 일별로 전송받아 그 자료를 근거로 자동차메이커에 컨설팅서비스를 제공하고 있다. 실증분석은 미국 전역에서 1999년 9월부터 2000년 10월까지 판매된 2000년형 미니 밴 시장을 대상으로 했다. 이 자료의 중요한 특성은 기존의 자동차 시장을 대상으로 한 연구에서 사용한 총합 자료와는 달리 개별적 거래 자료라는 것이다. 또한 A.C. Nielsen의 패널자료와 마찬가지로, 자동차 모델선택에 영향을 미칠 수 있는 변수의 횡단면적/시계열적

4 간단한 수식 유도를 통해 저자에게 3단계와 4단계 모형화의 차이점이 분산분해 (variance decomposition)에 있음을 지적해준 심사자에게 감사의 뜻을 전한다.



변이(cross-sectional/longitudinal variation)를 모두 포함하고 있는 장점을 가지고 있다. 미국 전역에서 판매된 모든 미니 밴 모델 중에서 시장점유율 상위 15개의 모델을 실증분석에 포함하였고, 자동차 메이커별 모델명, 시장점유율과 기타 주요변수의 평균치는 <Table 1>에 제시된 바와 같다. 실증분석에 포함된 지역(DMA)은 23개로서 각 지역별로 총 2,314개의 zip code가 분포되어 있으며 각 지역에 속한 zip code의 숫자는 <Table 2>에 나타난 바와 같다. 23개 지역에서 제공된 총 67,106개의 거래 자료를 무작위로 구분

하여 모수추정 (60,395개)에 사용하고 나머지 자료 (6,711개)는 예측타당성을 검증하는데 사용하였다.

## 2. 모형에 포함된 변수

포장 소비재시장에서 브랜드 선택모형 적용 시 주로 사용된 충성도 변수(Siddarth, Bucklin and Morrison 1995, Bronnenberg and Vanhonacker 1996, Chiang, Chib, and Narasimhan 1999)와 마찬가지로 형태의 변수로서 자동차모델 선택에도 중요한 영향을 미칠 것으로 예상

<Table 1> Market Shares by Model and Means of Marketing Mix Variables

Vehicle Model	Market Share	Price (\$)	Rebate (\$)	APR (%)
Astro	0.042	22,819	889	6.956
Caravan	0.090	20,144	1,563	7.316
Grand Caravan	0.135	25,167	1,654	6.399
Grand Voyager	0.045	23,376	---	7.810
Montana	0.027	26,650	1,692	5.591
Odyssey	0.159	26,508	---	8.507
Quest	0.053	23,488	1,000	7.811
Safari	0.014	23,550	806	7.405
Sienna	0.107	25,826	---	8.609
Silhouette	0.020	27,950	1,662	5.524
Town & Country	0.077	29,556	1,813	5.683
Venture	0.049	25,718	1,660	5.773
Villager	0.014	24,574	1,971	5.600
Voyager	0.038	20,250	1,481	7.356
Windstar	0.131	26,153	1,500	4.145

<Table 2> Number of Zip Codes Across DMAs

DMA	# of zip codes	DMA	# of zip codes
Baltimore/Washington	133	Northern California	256
Cleveland	79	Orlando	82
Colorado	65	Pennsylvania	166
Detroit	37	Phoenix	32
Georgia	79	Pittsburgh	23
Illinois/Indiana	180	Seattle/Portland	63
Indianapolis	30	South Texas	48
Miami	109	Southern California	301
Minneapolis	62	Tampa	55
Missouri	27	Tennessee	26
New England	162	Texas	64
New York	235		

되는 변수의 조작적 정의는 다음과 같다.

$ZIPMS_c^j$ : 소비자  $c$ 가 거주하는 zip code내 모델  $j$ 의 시장점유율로서 Lattin과 Bucklin(1989)의 정적인(static) 상표충성도 측정치와 동일하다.

$LASTMODEL_{jt}^c$ : 소비자  $c$ 가 교환(trade-in)하는 모델과 모델  $j$ 가 동일할 경우 1, 아니면 0의 값을 갖는다. Lattin과 Bucklin(1989)의 동적인(time-varying) 최근 상표충성도(last brand loyalty)와 동일하다.

$LASTMAKE_{jt}^c$ : 소비자  $c$ 가 교환하는 모델의 메이커와 브랜드  $j$ 의 메이커가 동일한 경우 중 그 모델과 모델  $j$ 가 다를 경우에만 1, 아니면 0의 값을 갖는다.

기타, 모형에 포함된 중요한 마케팅믹스 변수는 다음과 같다.

$PRICE_{jt}$ : 옵션가격을 제한 모델  $j$ 의 자동차가격 (vehicle price after market option)

$REB_{jt}$ : 소비자에게 제공되는 모델  $j$ 의 리베이트 액수

$APR_{jt}$ : 자동차 구입비 대출금의 연 이자율

이상의 변수들을 수식(1)의  $X_{jt}$ 에 포함하여 부록에 설명한 절차에 따라 모수를 추정하고 모수 추정치를 근거로 최적의 리베이트 수준을 추정하였다.

### 3. 최적의 리베이트 수준추정

zip code별 모수,  $\beta_r^z$ 의 함수로 표현된 수식(1)의 선택 확률을 근거로 리베이트 수준에 따른 zip code별 수익을 추정하고, 최적의 리베이트 수준을 결정하였다. 특정 리베이트 수준에 따른 수익이 실현될 확률은 수식(1)의 선택확률과 동일하다. 따라서 zip code  $z$ 의 리베

이트 수준,  $REB^z$ 에 따른 수익은 다음과 같이 계산할 수 있다.

$$profit^z(\beta_r^z, REB^z) = \left[ \sum_{n=1}^N \sum_{c \in z} [P_t^c(j | \beta_r^z, REB^z)(margin_j - REB^z)] \right] / N \quad (22)$$

여기서  $\beta_r^z$ 은 사후분포에서 샘플링한  $N$ 개의 추정치중  $n$ 번째로서 수식(22)는 모수의 사후분포로부터 추정한 평균수익을 의미한다. 실제로 집행된 리베이트 액수를 포함하는 범위로서 0달러부터 3000달러까지 100달러씩 리베이트 액수를 증가시키면서 계산된 평균수익 중 최고의 평균수익을 올리는 리베이트 액수를 최적 리베이트로 간주하였다. 수식(22)를 근거로 zip code별 최적 리베이트 수준,  $REB^{z*}$ 가 결정되면 지역  $r$ 의 최적 수익은 다음과 같이 계산된다.

$$profit_r^*(\beta_r^z) = \sum_{z \in r} profit^z(\beta_r^z, REB^{z*}) \quad (23)$$

만일 zip code 수준에서 리베이트의 최적 수준을 결정하지 않고, 같은 지역에 속한 zip code는 모두 동일한 리베이트 민감도를 갖고 있다는 가정 하에 지역수준의 최적 리베이트를 결정하고 zip code별로 동일한 리베이트를 제공한다면, 실제 zip code수준에서 실현되는 수익은 최적일 수 없을 것이다.

## IV. 분석결과

부록에 설명된 방법에 따라 추정된 모형들의 적합도와 모수추정치는 각각 <Table 3>과 <Table 4>에 제시된 바와 같다. zip code별 소비자 반응의 이질성을

5 메이커가 동일한 경우 1의 값을 갖게 되면,  $LASTMODEL_{jt}^c$  과 상관관계 1이 되므로, 모델은 다르지만 동일한 메이커일 경우에만 1의 값을 갖도록 조작한다.

<Table 3> Model Fit

Model	Calibration Log Marginal Density (LMD)	Holdou Mean Absolute Deviation (MAD)
HMNL	-135,430,118	0,820
NZ	-127,964,002	0,809
NRZ	-119,102,143	0,775

<Table 4> Parameter Estimates by Model

Variables	HMNL	Hierarchical Bayes Model	
		NZ ( $\mu$ )	NRZ ( $\mu_n$ )
Zip M/S	0,072*	0,083 (0,004)	0,078 (0,012)
LastModel	2,492*	2,673 (0,029)	2,997 (0,056)
LastMake	1,255*	1,269 (0,019)	1,306 (0,053)
Log(Price)	-0,882*	-1,426 (0,094)	-1,865 (0,106)
Log(Rebate)	0,459*	0,469 (0,039)	0,494 (0,093)
APR	-0,119*	-0,143 (0,005)	-0,187 (0,020)

\* significant at  $\alpha=0,05$  (HMNL), ( ): standard deviations of the posterior distribution for population parameters,  $\mu$  and  $\mu_n$  (NZ, NRZ)

반영하지 못한 HMNL모형은 모형적합도면에서 계층 베이스 모형보다 열등한 것으로 나타났다. 계층 베이스 모형 중에는 이 연구에서 제안한 NRZ모형이 NZ 모형보다 설명력(Calibration Log Marginal Density 참고)이나 예측 타당성(Holdout Mean Absolute Deviation 참고)에서 모두 우월한 것으로 나타났다. 베이스 요인 (Bayes Factor)에 의한 모형선택도 NRZ모형이 압도적

으로 우월한 것으로 나타났다 ( $BF=\exp(8,862)>100$ , Kass and Raftery 1995). 즉, zip code별 모수의 이질성을 반영할 때 지역내 이질성뿐만이 아니라 지역간 이질성도 반영하는 것이 모형적합도를 향상시킬 수 있는 요인이라는 점을 확인시켜주는 결과이다. 지역간 이질성( $\Sigma_n$ )과 지역내 이질성( $\Sigma_r$ )을 나타내는 공분산 행렬은 각각 <Table 5>와 <Table 6>에 제시된 바와

<Table 5> Between-Region Heterogeneity Matrix ( $\Sigma_n$ )

	Zip M/S	LastModel	LastMake	Log(Price)	Log(Rebate)	APR
Zip M/S	0,0015	0,0066	-0,0045	0,0057	0,0035	0,0029
LastModel	0,0066	1,0775	0,0277	-0,0196	-0,0438	0,0207
LastMake	-0,0045	0,0277	0,2809	0,0016	-0,0141	0,0226
Log(Price)	0,0057	-0,0196	0,0016	0,7155	0,0719	0,0439
Log(Rebate)	0,0035	-0,0438	-0,0141	0,0719	0,0663	0,0023
APR	0,0029	0,0207	0,0226	0,0439	0,0023	0,0402

\* The submatrix for brand-specific intercepts is excluded.

〈Table 6〉 Within-Region Heterogeneity Matrix ( $\Sigma_r$ )

	Zip M/S	LastModel	LastMake	Log(Price)	Log(Rebate)	APR
Zip M/S	0,0062	-0,0055	0,0004	-0,0018	-0,0025	0,0013
LastModel	-0,0055	5,6961	0,2126	-0,0354	0,2278	-0,0217
LastMake	0,0004	0,2126	1,9402	-0,4619	0,1078	0,0016
Log(Price)	-0,0018	-0,0354	-0,4619	2,6837	-0,7237	0,0388
Log(Rebate)	-0,0025	0,2278	0,1078	-0,7237	0,1982	-0,0051
APR	0,0013	-0,0217	0,0016	0,0388	-0,0051	0,1217

〈Table 7〉 VPC (Variance Partition Coefficients)

	Zip M/S	LastModel	LastMake	Log(Price)	Log(Rebate)	APR
$VPC = \frac{\sigma_n}{\sigma_n + \sigma_r}$	0,1948	0,1591	0,1265	0,2105	0,2507	0,2483

같다. 이 행렬의 대각선 요소를 각각  $\sigma_n$ ,  $\sigma_r$ 이라 하고 VPC<sup>6</sup>(Variance Partition Coefficient)를 계산한 결과는 <Table 7>에 제시된 바와 같다. VPC가 1인 경우, 지역내 이질성은 존재하지 않고, 지역간 이질성만 존재함을 의미한다. 즉, 같은 지역내에 속한 zip code모수는 서로 차이가 없으며 그 지역의 모수 평균 ( $\mu_r$ )과 동일하며 지역간 평균차이만 존재한다는 의미이다. 반면에 VPC가 0인 경우, 지역간 이질성은 존재하지 않고 NRZ모형은 NZ모형과 동일한 모형이 된다. 추정된 VPC결과를 보면, 마케팅 믹스 변수 (가격, 리베이트, APR)들의 VPC는 20%이상으로서 zip code간 이질성의 20% 이상이 지역간 (DMA간) 이질성으로부터 기인되는 것으로 해석된다. zip code간 이질성의 대부분, 즉, 80%정도는 지역내 이질성으로 기인되지만 여전히 상당부분 (20%)의 zip code간 이질성은 지역간 차이에서 비롯됨을 의미한다.

HMNL모형과 계층 베이스 모형의 근본적인 차이점은 시장의 반응(소비자 민감도)을 HMNL모형에 근거해서 이해한다면 최적의 리베이트 수준을 결정함에 있어서 전국적으로 균일한 리베이트 액수를 결정한다는 것이고, 반면에 계층 베이스 모형에 근거하여 시

장의 반응을 가정한다면, zip code 수준에서 최적의 리베이트 수준을 결정한다는 것이다. 각각의 모형이 시장의 반응에 대하여 서로 다른 가정을 하고 있으므로 그러한 가정 하에 추정된 최적의 리베이트 수준에 따라 예상되는 수익을 비교하였다. 분석대상 자동차 모델로서 Dodge Caravan을 선택했으며 분석대상 지역은 5개의 DMA(Designated Marketing Area)를 선정하였다. 수익을 계산하는 절차는 다음과 같다. 우선 각각의 모형이 가정하는 바에 따라 최적의 리베이트 수준을 결정하였다 (개별 모형의 추정된 모수에 근거하여 최적의 리베이트 수준을 결정하였다). HMNL모형의 경우 최적 리베이트는 900불, NZ모형의 경우 zip code별 최적 리베이트의 평균은 1,025불, NRZ모형의 경우 zip code별 최적 리베이트의 평균은 1,102불로 결정되었다. 모형별로 예상하는 수익은 각각 8,788만불, 9,039만불, 9,693만불로 나타났다. 이러한 차이는 마케팅 의사결정자가 가정하는 모형에 따라 최적의 리베이트 수준이 달라지고, 각각의 모형이 추정하는 최적 이윤도 달라짐을 의미한다. 그러나 각각의 모형이 가정하는 바에 따라 추정한 최적의 리베이트는 실제로는 최적일 수 없다 (NRZ모형의 경우는 제외). 왜

6  $VPC = \frac{\sigma_n}{\sigma_n + \sigma_r}$ , 0과 1사이에 있으며 총분산중 지역간 이질성의 비율을 나타냄.

나하면 모형적합도 측면에서 볼 때 실제 수익은 NRZ 모형이 예측하는 시장의 반응에 근접하게 실현될 것이기 때문이다. 따라서 최적의 리베이트 수준은 개별 모형의 모수에 근거해서 추정하고, 실현되는 수익은 NRZ모형의 모수에 근거하여 다시 계산하였다. 계산된 결과는 <Table 8>에 제시된 바와 같다.

NRZ모형의 모수에 근거하여 예상수익을 분석한 결과, 리베이트를 하지 않은 경우 2,341만불, 연구기간내 최빈 리베이트 액수인 1500불의 리베이트를 시행한 경우 8,425만불, HMNL모형의 가정에 따라 지역 내 동일한 리베이트(900불)를 시행한 경우 9,044만불, NZ 그리고 NRZ모형의 서로 다른 가정에 따라 zip code별로 최적의 리베이트를 추정하고 NRZ모형의 모수에 근거하여 실현된 수익을 총합한 경우 각 각 9,204만, 9,693만불의 수익을 내는 것으로 추정되었다. 분석지역, 분석대상 자동차 모델에 따라 실현되는 수익은 달라지겠지만, 서로 다른 가정 하에 결정된 최적의 리베이트를 시행하였을 때 NRZ모형에 따라 실현되는 각 모형별 수익의 차이는 여전히 존재할 것이며 NRZ모형에 근거한 최적의 리베이트 수준이 최대의 수익을 보장할 것이다.

기존의 선택모형을 적용한 실증논문의 경우 자료가 수집된 지역 혹은 도시(예를 들어, AC Nielsen의 패널 자료의 경우, Sioux Falls)에 국한하여 모형을 추정하고, 가구별 모수의 분포를 유한의 지지점(latent class 모형의 finite support points)이나 다변량 정규분포(계층

베이즈 모형의 경우)로 가정하여 추정된 세그먼트별 혹은 가구별 모수를 바탕으로 의사결정하는 방식을 사용하였다. 즉, 다른 지역의 자료는 가용성의 문제도 있지만 가용하더라도 사용할 방법에 대하여 고려하지 않았다는 것이다. 예를 들어, 경기지역에서 자동차 리베이트 수준을 결정하려면 경기지역의 거래 자료만 가지고 의사결정을 했다는 것이다. 이 논문에서 제안하는 적용방식의 유리한 점은 경기지역뿐만이 아니라 전국적인 자료가 가용하다면 4단계 모형화(NRZ모형)를 통해서 다른 지역의 정보도 함께 활용함으로써 모형적합도가 향상되어 보다 합리적인 의사결정을 할 수 있다는 것이다.

## V. 결론과 향후 연구방향

zip code별로 서로 다른 리베이트 민감도를 가정한 계층 베이즈 모형이 동일한 민감도를 가정한 HMNL모형보다 모수 추정샘플과 예측용 유보샘플에서 모두 우월한 것으로 나타났다. 아울러 계층 베이즈 모형 중, 기존의 3단계 NZ모형에 비해 4단계 NRZ모형이 지역내 이질성뿐만 아니라 지역간 이질성을 반영함으로써 설명력이나 예측타당성을 향상시킬 수 있음을 보여주었다.

수익성을 비교하는 시뮬레이션에서 나타난 바와 같이 최적의 리베이트 수준을 결정하기 위해서는 zip

<Table 8> Profitability Analysis (unit: \$)

DMA (Designated Marketing Area)	Zero Rebate	Blanket Rebate (\$1500)	HMNL Optimal Profit	NZ Optimal Profit	NRZ Optimal Profit
Cleveland	21,823	201,398	203,345	205,283	225,407
Indianapolis	7,953,237	23,058,732	23,621,356	24,239,809	26,482,357
Illinois/Indiana	11,367,766	37,661,356	42,013,215	42,574,528	43,943,245
Detroit	600,454	8,342,659	8,628,372	8,996,794	9,280,247
Minneapolis	3,473,554	14,987,275	15,983,257	16,025,646	16,994,355
TOTAL	23,416,834	84,251,420	90,449,545	92,042,060	96,925,611

code별로 서로 다른 리베이트 민감도를 반영하는 것이 중요하다고 판단된다. 왜냐하면 HMNL모형의 가정에 근거해서 zip code에 관계없이 모두 동일한 리베이트 민감도를 가정하고 일률적인 리베이트를 제공하는 것보다 이 연구에서 제안한 모형에 근거해서 zip code별로 최적의 리베이트 수준을 결정하는 것이 항상 수익성을 향상시킬 수 있기 때문이다.

이 연구에서 제안된 모형은 기존의 계량경제학적 모형이나 3단계 계층 베이스 모형의 한계를 극복하고 보다 일반화된 구조를 갖고 있으므로, 자동차시장 이외의 기타 산업분야(예를 들면, 전국적인 체인을 갖고 있는 할인점)에서도 지역별 정보량의 차이가 있는 경우 4단계 모형화를 통해 보다 광범위한 정보를 반영한 모수 추정이 가능할 것으로 기대된다. 또한 추정된 모수(시장의 민감도)를 근거로 프로모션에 대한 합리적 의사결정뿐만 아니라 지역별 판매반응 함수에 따른 영업사원의 최적 수준을 결정하는 것과 같은 분야에도 적용 가능할 것으로 판단된다.

영업사원 배분문제에 대한 고전적인 모형 콜 플랜(CALL PLAN, Lodish 1971)을 예로 들면 지역별 관리자가 영업력을 배분하는 경우 관리 지역 내의 점포 방문횟수에 대한 점포별 판매반응이 중요한 의사결정의 근거가 된다. 관리자의 판단에 의존하는 결정 미분법(decision calculus) 대신 거래 자료를 통해 판매반응 함수의 모수를 추정할 경우, 지역 내의 점포별로 충분한 자료가 축적되어 추정이 가능하다면 모르겠지만, 새로 개척한 점포는 자료의 부족으로 판매반응 함수를 추정하기 어려울 수 있다. 이러한 경우 NRZ모형의 적용은 자료가 부족한 점포의 판매반응함수일지라도 다른 지역의 정보를 망라하여(pooling) 모수 추정이 가능하며 그 점포가 속한 지역내 이질성을 반영할 뿐만 아니라 다른 지역과의 차이도 반영함으로써 보다 합리적인 의사결정을 할 수 있도록 도움을 줄 것이다.

마지막으로 향후 연구방향으로서, 공간모형(spatial

model, Bronnenberg and Sismeiro 2002)을 통한 실증 분석과 제안된 모형화 방식의 비교 연구를 생각해 볼 수 있다. 공간모형의 적용은 서로 다른 zip code의 수요함수를 추정함에 있어서 zip code간 거리(location differential)를 모수에 반영하는 방식으로서, 모수 공분산행렬의 요소(element)를 zip code간 거리의 베셀(Bessel) 함수로 가정하고 베셀 함수의 모수를 추정하는 방식이다. 베셀 함수와 같이 zip code별 모수의 공분산이 zip code간 거리에 따라 감소 후 다시 증가하는 형태로 모형화할 수도 있고, 보다 단순한 단조감소의 형태로 모형화할 수도 있다. 또는 이 연구에서 제안한 방식처럼 zip code간 거리가 구체적으로 모수 추정에 반영되지는 않지만 지역간 이질성과 지역내 이질성을 반영하는 방식으로 모수를 추정할 수도 있다. 이렇게 다양한 방식으로 추정한 결과를 비교하여 장단점을 정리하는 것도 흥미로운 연구주제가 될 것으로 생각한다.

## 부록: NRZ모형 추정방법

NRZ모형과 같은 계층 베이스 모형의 경우 추정해야 할 모수의 선형분포를 가정하고 실제 데이터의 정보를 반영하는 우도(likelihood)와 조합된 사후분포에 근거한 추론을 기초로 한다. 선형분포와 사후분포는 다음과 같다.

### 1. 선형분포 (Prior Distribution):

이 연구에서 가정한 초선형분포와 선형분포는 II. 연구모형.에서 설명한 다변량 정규분포와 Wishart분포이다(Gelfand et al. 1990). 추정해야 할 모수, 즉,  $\beta_r^0$ ,  $\mu_r$ ,  $\Sigma_r$ ,  $\mu_n$ ,  $\Sigma_n$  이외의 고정된 모수(fixed parameter)는

추정해야 할 모수의 값에 실질적인 영향을 주지 못하지만 (즉, diffuse prior을 구성하도록), 선행분포가 적절하도록 (즉, proper prior을 구성하도록, 따라서, 사후분포도 적절하도록), 선행연구의 예를 따라 값을 정하였다 (Gelfand et al. 1990, Allenby et al. 1998). 그 값은 다음과 같다.  $\eta=0$ ,  $C=1000I$ ,  $\rho=21$  (i.e.,  $R$ 의 차원+1),  $\rho_n=21$  (i.e.,  $\Sigma_n$ 의 차원+1),  $R=15I$ . 선행분포로부터 유도된 사후분포에 근거해서 Markov Chain Monte Carlo (MCMC) 방식에 따라 깃스 샘플링과 Metropolis-Hastings Algorithm을 혼용하여 모수를 추정하였다. 추정하고자 하는 모수의 (m+1)번째 추정치는 다음과 같은 사후분포로부터 추출하였다 (Geman and Geman 1984, Gelfand and Smith 1990, Tierney 1994, Chib and Greenberg 1995).

## 2. 사후분포 (Posterior Distribution):

(1)  $\beta_r^{(m+1)}$ 는 다음과 같은 사후분포로부터 추출한다.

$$p(\beta_r^{(m+1)} | \mu_r^{(m)}, \Sigma_r^{(m)}, y^z) \propto f(y^z | \beta_r^{(m)}) \phi(\beta_r^{(m)} | \mu_r^{(m)}, \Sigma_r^{(m)}) \quad (1)$$

여기서  $y^z$ 는 데이터를 의미하고,  $\phi(\cdot | \cdot)$ 는 다변량 정규밀도함수(multivariate normal density)이고  $f(\cdot | \cdot)$ 는 NRZ모형 1단계의 선택확률로부터 계산된 우도함수이다. 우변은 다변량 정규분포의 밀도함수와 로짓 우도함수의 곱으로서, 알려진 사후분포가 아니므로 Metropolis-Hastings algorithm의 random walk version에 따라  $\beta_r^{(m+1)}$ 를 추출한다(Tierney 1994, Chib and Greenberg 1995, Green 1995). 먼저 (m)번째 iteration에서 추출한  $\beta_r^{(m)}$ 에 오차항  $u$ 를 더해서  $\beta_r^{(m+1)}$ 의 후보 추정치(candidate draw)  $\beta_r^{(c)}$ 를 다음과 같이 정의한다.

$$\beta_r^{(c)} = \beta_r^{(m)} + u \quad (2)$$

여기서  $u$ 는  $MVN(0, V)$  분포에서 추출한다.  $V$ 는 Chiang et al. (1999)의 제안에 따라, MLE로 추정된 HMNL 모수의 information matrix를 근거로 한다.  $\beta_r^{(c)}$ 를  $\beta_r^{(m+1)}$ 로 채택할 확률은 다음과 같이 정의된다.

$$\min \left[ \frac{f(y^z | \beta_r^{(c)}) \phi(\beta_r^{(c)} | \mu_r^{(m)}, \Sigma_r^{(m)})}{f(y^z | \beta_r^{(m)}) \phi(\beta_r^{(m)} | \mu_r^{(m)}, \Sigma_r^{(m)})}, 1 \right] \quad (3)$$

좌변의 비율이 1이상이면  $\beta_r^{(c)}$ 를  $\beta_r^{(m+1)}$ 로 채택하고, 미만이면 그 비율을 확률로 하는 이항분포에서 Bernoulli trial한 결과로 채택/기각을 결정한다.  $\beta_r^{(c)}$ 가 기각되면  $\beta_r^{(m+1)} = \beta_r^{(m)}$ 로 정한다.

(2)  $\mu_r^{(m+1)}$ 는 다음과 같은 사후다변량 정규분포로부터 깃스 샘플링한다.

$$MVN[D(Z\Sigma_r^{-1(m)}B_r + \Sigma_n^{-1(m)}\mu_n), D] \quad (4)$$

여기서  $D = (Z\Sigma_r^{-1(m)} + \Sigma_n^{-1(m)})^{-1}$ ,  $B_r = \frac{1}{Z} \sum_z \beta_r^{(m)}$  이고  $Z$ 는 지역  $r$ 에 속한 총 zip code의 수이다.

(3)  $\Sigma_r^{-1(m+1)}$ 는 다음과 같은 사후Wishart분포로부터 깃스 샘플링한다.

$$Wishart \left[ \left[ \sum_{z=1}^Z (\beta_r^{(m)} - \mu_r^{(m)})(\beta_r^{(m)} - \mu_r^{(m)})' + \rho_n \Sigma_n^{(m)} \right]^{-1}, Z + \rho_n \right] \quad (5)$$

(4)  $\mu_r^{(m+1)}$ 는 다음과 같은 사후다변량 정규분포로부터 깃스 샘플링한다.

$$MVN[A(K\Sigma_n^{-1(m)}B_n + C^{-1}\eta), A] \quad (6)$$

여기서  $A = (K\Sigma_n^{-1(m)} + C^{-1})^{-1}A = (K\Sigma_n^{-1(m)} + C^{-1})^{-1}$ ,

$B_n = \frac{1}{K} \sum_r^K \mu_r^{(m)}$ 이고  $K$ 는 총 지역의 수이다.

(5)  $\Sigma_n^{-1(m+1)}$ 는 다음과 같은 사후Wishart분포로부터  
 깃스 샘플링한다.

$$Wishart \left[ \left[ \sum_{r=1}^K (\mu_r^{(m)} - \mu_n^{(m)})(\mu_r^{(m)} - \mu_n^{(m)})' + \rho R \right]^{-1}, K + \rho \right] \quad (7)$$

위의 절차에 따라 추출된 일련의 추정치는  
 “burn-in” 단계(초기 값으로부터 수렴하기까지의 단계)  
 를 거쳐서 안정화된 Markov Chain을 형성하게 되는데,  
 수렴된 이후의 추정치중 추정치 간의 자기상관관계를  
 제거하기 위해 몇 회의 iteration을 건너뛰면서(i.e.,  
 thinning the chain, 본 연구에서는 10회) 최종 추정치  
 $N$ 개 (본 연구에서는 1000개)를 추출하고, 추정치의 분  
 포를 이용하여 통계적 추론을 한다.

<최초투고일: 2007년 12월 10일>

<수정일: 1차: 2008년 5월 13일>

<수정일: 2차: 2010년 7월 27일>

<게재확정일: 2010년 8월 5일>

## 참고문헌

장광필 (2003), “고려상표군을 반영한 자동차 시장구조 분석  
 모형,” *마케팅연구*, 18(2), 53-67.

Allenby, Greg M., Neeraj Arora, and James L. Ginter (1998),  
 “On the Heterogeneity of Demand,” *Journal of Marketing  
 Research*, 35 (August), 384-389.

Ansari, Asim M., and Carl F. Mela (2003), “E-Customization,”  
*Journal of Marketing Research*, 30 (May), 131-145.

Berry, S., J. Levinsohn and A. Pakes (1995), “Automobile Prices

in Marketing Equilibrium,” *Econometrica*, 63(4), 841-890.

Bronnenberg, Bart J. and Catarina Sismeiro (2002), “Using  
 Multimarket Data to Predict Brand Performance in Markets  
 for Which No or Poor Data Exist,” *Journal of Marketing  
 Research*, 39 (Feb), 1-17.

Bronnenberg, Bart J. and Wilfried R. Vanhonacker (1996),  
 “Limited Choice Sets, Local Price Response, and Implied  
 Measures of Price Competition,” *Journal of Marketing  
 Research*, 33(May), 163-173.

Bucklin, Randolph E., Sunil Gupta and S. Siddarth, (1998),  
 “Determining Segmentation in Sales Response Across  
 Consumer Purchase Behaviors,” *Journal of Marketing  
 Research*, 35 (May), 189-197.

Carlin, Bradley P. and Thomas A. Louis (2000), *Bayes and  
 Empirical Bayes Methods for Data Analysis*, 2nd ed.,  
 Chapman & Hall/CRC.

Chang, Kwangpil, S. Siddarth, and Charles B. Weinberg  
 (1999), “The Impact of Heterogeneity in Purchase Timing  
 and Price Responsiveness on Estimates of Sticker Shock  
 Effects,” *Marketing Science*, 18(2), 178-192.

Chiang, Jeongwen, Siddartha Chib and Chakravarthi  
 Narasimhan (1999), “Markov Chain Monte Carlo Models of  
 Consideration Set and Parameter Heterogeneity,” *Journal of  
 Econometrics*, 89(1-2), 223-248.

Chib, S. and E. Greenberg (1995), “Understanding the Metropolis-  
 Hastings Algorithm,” *American Statistician*, 49, 327-335.

Chintagunta, Pradeep (1993), “Investigating Purchase Incidence,  
 Brand Choice and Purchase Quantity Decisions of  
 Households,” *Marketing Science*, 12 (Spring), 184-208.

Colombo, Richard A., and Donald G. Morrison, (1989), “A  
 Brand Switching Model with Implications for Marketing  
 Strategy,” *Marketing Science*, 8 (Winter), 89-99.

Gelfand, A. E. and A. F. M. Smith (1990), “Sampling Based  
 Approaches to Calculating Marginal Densities,” *Journal of  
 the American Statistical Association*, 85, 398-409.

Gelfand, A. E., S. E. Hills, A. Racine-Poon, and A. F. M.  
 Smith (1990), “Illustration of Bayesian Inference in Normal  
 Data Models Using Gibbs Sampling,” *Journal of the*



- American Statistical Association*, 90, 972-985.
- Geman, S., and D. Geman (1984), "Stochastic Relaxation, Gibbs Distributions and the Bayesian Restoration of Images," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 6, 721-741.
- Green, Peter J. (1995), "Reversible Jump Markov Chain Monte Carlo Computation and Bayesian Model Determination," *Biometrika*, 82(4), 711-732.
- Guadagni, Peter and John D. C. Little (1983), "A Logit Model of Brand Choice Calibrated on Scanner Data," *Marketing Science*, 2 (Summer), 203-38.
- Gupta, Sunil (1988), "Impact of Sales Promotions on When, What and How Much to Buy," *Journal of Marketing Research*, 25 (November), 342-55.
- Kamakura, Wagner A. and Gary J. Russell (1989), "A Probabilistic Choice Model of Market Segmentation and Elasticity Structure," *Journal of Marketing Research*, 26 (November), 379-90.
- Kass, Robert E and Adrian E. Raftery (1995), "Bayes Factors," *Journal of the American Statistical Association*, 90(430), 773-795.
- Lattin, James M. and Randolph E. Bucklin (1989), "Reference Effects of Price and Promotion on Brand Choice Behavior," *Journal of Marketing Research*, 26 (August), 299-310.
- Lodish, Leonard M. (1971), "CALLPLAN: An Interactive Salesman's Call Planning System," *Management Science*, 18, no. 4, pt. 2 (December), 25-40.
- Rossi, Peter E., Robert E. McCulloch and Greg M. Allenby (1996), "The Value of Purchase History Data in Target Marketing," *Marketing Science*, 15 (4), 321-340.
- Siddarth, S., Randolph E. Bucklin, and Donald Morrison (1995), "Making the Cut: Modeling and Analyzing Choice Set Restriction in Scanner Panel Data," *Journal of Marketing Research*, 32 (August), 255-266.
- Silva-Risso, Jorge, M., Randolph E. Bucklin and Donald G. Morrison (1999), "A Decision Support System for Planning Manufacturers Sales Promotion Calendars," *Marketing Science*, 18 (3), 274-300.
- Steenburgh, Thomas J., Andrew Ainslie and Peder Hans Engbretson (2003), "Massively Categorical Variables: Revealing the Information in Zip Codes," *Marketing Science*, 22 (1), 40-57.
- Sudhir, K. (2001), "Competitive Pricing Behavior in the Automobile Market: A Structural Analysis," *Marketing Science*, 20(1), 42-60.
- Tellis, G. and F. Zufryden (1995), "Tackling the retailer decision maze: Which brands to discount, how much, when and why?," *Marketing Science*, 14(3), 271-300.
- Tierney, L. (1994), "Markov Chains for Exploring Posterior Distributions," *Annals of Statistics*, 22, 1701-62.