



Munich Personal RePEc Archive

Analysis of housing choice using multinomial probit model – Bayesian estimation

Park, Sang Soo and Lee, Chung-Ki

Korea University

December 2011

Online at <https://mpra.ub.uni-muenchen.de/37150/>

MPRA Paper No. 37150, posted 08 Jan 2013 23:10 UTC

베이지안 추정법을 이용한 주택선택의 다항프로빗 모형 분석*

박상수** · 이충기***

목 차

- I. 서 론
- II. 선택대안 및 변수의 구성
- III. 다항프로빗 모형
- IV. 실증분석 결과
- V. 결 론

환경, 에너지, 교통 분야의 수요 분석 및 예측에 많이 응용되는 이산선택모형(discrete choice model)은 대부분 로지스틱 분포를 가정하고 있으며, 주택선택과 관련한 선행연구들도 대안들 간 독립성을 가정하는 다항로짓(multinomial logit) 모형의 다양한 형태로 분석되었다. 이에 비해 본 연구는 대안들간의 독립성 가정을 완화한 다항프로빗(multinomial probit)모형을 설정하였고, 이를 추정하기 위해 Gibbs sampling을 이용한 베이지안 추정방법을 주택선택 문제에 적용하였다. 연구결과 대안들 간의 독립성 가정이 충족되지 않음을 확인하였고,

▷ • 논문투고일 : 2011. . • 수정일 : 2011. . • 게재확정일 : 2011. .

* 이 논문은 고려대학교 교내연구비 및 2008년 정부(교육과학기술부)의 재원으로 한국연구재단의 지원을 받아 수행된 연구임(KRF-2008-327-B00144).

** 고려대학교 경제학과 조교수(제1저자), Tel:(02)3290-2227, Email: starpac@korea.ac.kr

*** 고려대학교 경제연구소 연구교수(교신저자), Tel:(02)3290-1675, Email: chungki@korea.ac.kr

기준대안이 아닌 다른 대안의 비용이 기준대안보다 높아지면 소비자가 그 대안을 선택할 가능성이 떨어지고 기준대안을 선택할 가능성이 높아지는 것으로 나타났다. 그리고 대안들의 비용이 동일할 경우 소득이 높아질수록 아파트-소형에 비해 아파트-대형을 선호하는 것으로 나타났다. 한편 주거환경요인변수 중에서 구입당시의 가격, 주택가격 상승에 대한 기대심리, 교육여건 등은 유의수준 5%에서 유의적인 계수값이 존재하였다. 이는 각 변수가 각각의 대안과 기준대안 사이의 선택에서 어떤 역할을 하는지를 보여주는 것으로 일례로서 향후 집값상승에 대한 기대가 큰 가구일수록 대형단독주택에 비해 소형아파트나 소형단독주택을 선택할 확률이 높음을 의미한다.

핵심주제어 : 주택선택, 다항프로빗 모형, 베이지안 추정법

경제학문헌분류기호 : C5, D7, H4

I. 서론

주거 환경의 급속한 변화에 따라 주택 수요자의 사회·경제적 특성과 주택선택 간의 관계를 규명하기 위한 다양한 연구들이 진행되었으며, 최근 연구로 이충기·이주석(2008)은 6대 광역도시를 대상으로 주택가격의 상승기대, 교육여건, 쾌적한 환경, 교통의 편리성, 직장 접근성 등 주거환경적 요소가 주택선택에 미치는 영향을 수도권과 비수도권으로 구분하여 분석하였다. 구체적인 모형의 형태는 조금씩 차이가 있지만 장성수·윤혜정(1998), 최막중·임영진(2001), 정의철(2002), 김정수·이주형(2004) 등 대부분의 연구에서와 마찬가지로 로지스틱 분포를 가정한 다항로짓(multinomial logit; MNL) 모형을 추정모형으로 이용하였다.

그러나 MNL 모형은 선택확률의 도출이 용이한 반면, 무관한 대안들의 독립성(Independence of irrelevant alternatives; IIA)이라는 너무 제약적인 가정으로 인하여 대안 간 대체관계가 비현실적인 경우가 발생하는 문제를 내포하

고 있다(McFadden, 1984). 이러한 문제를 해결하기 위해 교란항을 보다 유연하고 일반적인 상관관계를 가정한 발전된 형태의 방법론이 고려될 수 있다. 예를 들어, Generalized extreme value model(Train, 1986), Heteroskedastic extreme value model(Bhat, 1995), 다항프로빗(multinomial probit; MNP) 모형 등이 그 예이다. 이 중 MNP 모형은 MNL 모형의 IIA 가정을 하지 않는다는 점이 특징이다(Wooldridge, 2002).

MNP 모형은 추정방법의 어려움 때문에 거의 사용되지 않았으나 McFadden(1989), Keane(1994), McCulloch and Rossi(1994) 등이 제안한 시뮬레이션 기법으로 계산상의 효율성이 개선되었다. MNP의 추정에 있어 고전적인 방법론(Classical approach)은 일반적으로 최우추정법(maximum likelihood estimation)에 기초를 두고 있으나, 대안의 수가 많아질수록 다변량 정규분포의 적분과 계산상의 어려움 때문에 그 적용이 다소 제한적이다. 이에 비해 베이지안 방법론(Bayesian approach)은 Albert and Chib(1993), Nobile(1998), McCulloch et al.(2000) 등에 의해 발전되었으며, 추정 파라메타의 사후분포(posterior distribution)에서의 표본 추출을 Gibbs sampling 방법을 이용함으로써 유한 표본추정(finite sample inference)을 한다. 동 방법론은 대안의 수가 많은 MNP 문제의 경우에도 무난히 사용될 수 있으며, 우도함수 및 선택확률의 구성과 추정을 할 필요가 없고, 점근적 분석(asymptotic analysis)에 의존하지 않고 선택확률, 탄력성 등 연구자가 원하는 어떠한 변수라도 분포에 대한 비교적 정확한 근사를 할 수 있다는 장점을 가지고 있다. Geweke(1994)에 따르면 동 방법론은 다른 시뮬레이션 방법론들에 비해 우수한 것으로 평가하고 있다.

한편, 이길순(1990)은 주택의 보유형태, 규모, 유형의 세 가지 주택특성을, Cho(1997)는 보유형태와 유형을, Huang and Clark(2002)는 보유형태가 주택선택에 미치는 영향을 분석하였다. 한편, 주택의 주거환경변수로서 김한수의 2명(1998)은 교통, 주택가격을, 최막중·임영진(2001)과 김재홍(2006)은 교통과 쾌적성을 중요한 변수로서 분석에 포함시켰다. 이에 본 논문은 주택의 보유여부, 규모, 유형과 함께 주거환경요인변수를 분석의 대상으로 선정하였다.

본 연구는 MNP의 Gibbs sampling을 이용한 베이지안 방법론을 주택선택 문제에 적용하고, 이를 위해 교란항의 분산-공분산 행렬의 첫 번째 원소를 사전

적으로 1로 설정하는 McColluch et al.(2000)의 완전 식별 모형을 실증분석에 응용한다. 이후 본 논문의 구성은 다음과 같다. 제II장에서는 주택선택 대안 및 변수의 구성에 대해 설명하고, 제III장은 본 연구의 추정모형인 다항프로빗 모형에 대해 구체적으로 살펴본다. 그리고 제IV장에서는 실증분석결과를 제시하고, 마지막으로 제V장은 결론으로 할애한다.

II. 선택대안 및 변수의 구성

본 논문에서 사용된 자료는 2006년 7월 6대 광역도시의 600세대를 대상으로 전문리서치회사를 통해 개별 방문하여 설문조사한 결과이다. 이 중 주택가격과 소득에 관한 정보를 제대로 응답하지 않은 17명을 제외한 583명의 설문자료를 분석하기로 한다. 선택대안의 구성은 이충기·이주석(2008)의 구분을 원용하여 주거형태와 크기를 결합해 아파트-중소형(AS), 아파트-중대형(AL), 비아파트-중소형(HS), 비아파트-중대형(HL)의 4가지 결합선택 대안을 구성하였다. 주택 또는 아파트의 넓이가 99.3m²(30평) 미만은 중소형, 그 이상은 중대형으로 구분하였다. 분석에 사용한 583명의 응답자 중 각각의 대안을 선택한 빈도 및 비율과 대안별 변수들의 기술통계량은 다음의 <표 1>, <표 2>와 같다.

<표 1> 선택대안 빈도 및 비율

구 분	응답자 수	아파트- 중소형(AS)	아파트- 중대형(AL)	비아파트- 중소형(HS)	비아파트- 중대형(HL)
자가	432(74.1)	112(19.2)	155(26.6)	64(11.0)	101(17.3)
전·월세	151(25.9)	47(8.0)	11(1.9)	79(13.6)	14(2.4)
총계	583(100.0)	159(27.2)	166(28.5)	143(24.6)	115(19.7)

주 : 1) () 안은 583명에 대한 비율을 나타내며, 단위는 % 임.

<표 2> 대안별 변수들의 평균과 표준편차

대안	응답자 연령(년)	가구주 비율(%)	보유 비율(%)	주택 규모(평)	연 가구소득 (천만원)	전세등가액 (천만원)
AL	44.6 (7.5)	52.4 (50.1)	93.4 (24.9)	34.5 (5.6)	4.82 (1.62)	11.27 (8.61)
AS	40 (7.4)	51.6 (50.1)	70.4 (45.8)	23.7 (2.7)	3.77 (1.03)	7.66 (6.18)
HL	44.4 (7.6)	53.0 (50.1)	87.8 (32.8)	37.9 (10.5)	4.42 (1.52)	9.25 (7.04)
HS	41.3 (7.8)	58.7 (49.4)	44.8 (49.9)	22.1 (4.5)	3.27 (1.11)	5.00 (3.20)

주: 1) () 안은 표준편차이며, 전세등가액에 대한 자세한 설명은 2절에서 상술함.

1. 변수 설명

추정모형에 사용되는 변수는 크게 종속변수인 선택변수(choice variable)가 있고, 설명변수는 개인특성변수(individual specific variables)와 대안특성변수(alternative specific variables)로 구분된다. 선택변수는 가용한 대안들 AL, AS, HL, HS의 4가지 중 개인 i 가 어떤 것을 선택했는지를 말해주는 변수이다. 한편, 동일한 개인이 대안과 무관하게 하나의 값을 갖는 개인별 변수로는 주택보유여부 더미변수와 가구주 연령과 같은 개인특성변수(X_i)와 주택선택 시 고려되는 주택가격, 교육여건, 교통 편의성 등과 같은 주거환경요인변수(P_i)로 구분하였다. 그리고 대안특성변수는 대안선택에 따른 비용으로 응답자료를 이용하여 전세등가액을 추산하여 사용하였다.

본 연구에서 사용한 개인특성변수 중 설명을 요하는 변수는 주택보유여부이다. 대안들 간의 선택에서 주택을 소유하고자 하는 사람과 전세 또는 월세로 거주하려는 사람의 선택은 다를 수 있다. 따라서 주택소유여부를 고려하여 모형을 수립하여야 할 것이다.

주택소유를 모형에 포함시키는 첫 번째 방법으로 개별 가구의 의사결정이 주택소유형태-주거형태-크기를 동시에 고려해 동시적 의사결정을 한다고 보아 대안의 구성 자체에 포함시키는 것을 생각할 수 있다. 이 방법을 선택할 경우 본 연구의 대안은 총 8개(각 대안별로 소유-전·월세로 구분) 또는 12개(각 대안

별로 소유-전세-월세로 구분)가 될 것이다.

다른 방법으로는 개별 가구가 외생적 요인에 의해 주택의 소유여부를 먼저 결정한 후에 주거형태-크기를 결합한 대안을 선택한다는 순차적 의사결정을 한다고 보아 주거형태-크기를 결합한 대안을 선택할 상황에서는 이미 소유-전·월세 여부는 결정되어 있다고 보는 것이다. 이 경우 결합대안의 선택단계에서 주택의 소유여부는 모형의 설명변수로 간주될 수 있다.

본 논문에서는 후자의 접근방법을 사용해서 주택소유여부를 설명변수로 고려하였다. 이는 저자들이 개념적으로 후자가 더 현실적이라고 생각한 까닭이 아니라 전자의 접근을 취해 대안을 8개 또는 12개로 구성할 경우 AL과 HL의 전·월세에 해당하는 데이터가 너무 적어(각각 11개, 14개) 의미있는 추정치를 얻을 수 없다는 현실적인 문제 때문이었다. 추후 보다 광범위한 데이터 수집을 할 수 있다면 현재의 모형보다 현실적인 대안의 선택과정을 모형화하고 이를 추정하는 작업을 하여야 할 것이다.

주거환경요인변수는 개인특성변수 이외에 개인들의 주택의사결정에 영향을 미치는 요인으로 이충기·이주석(2008)의 연구와 유사하게 구입당시 주택가격(P1), 자가상승기대(P2), 교육여건(P3), 교통 편의성(P4), 환경의 쾌적성(P5), 직장 접근성(P6), 그리고 치안(P7)과 같이 총 7가지 변수를 고려하였다. 또한, 설문조사를 통해 만족스럽거나 중요하다고 생각하는 항목 3개를 선택하여 만족도/중요성에 따라 1-3점의 점수를 부여토록 하였으며 점수가 커질수록 만족도/중요도가 증가한다는 것을 의미한다. 다음의 <표 3>은 위의 7가지 주거환경요인변수에 대한 기술통계량들이다.¹⁾

1) 이 기술통계량은 분석에 이용된 설문 응답자 583 명만으로 계산하였다.

<표 3> 주거환경요인변수의 기술통계량

변수		P1	P2	P3	P4	P5	P6	P7	
평균		0.88	0.72	1.15	1.45	1.21	0.3	0.17	
표준편차		1.23	1.1	1.17	1.12	1.17	0.72	0.49	
점수별 빈도	3	119 (20.41)	69 (11.84)	111 (19.04)	129 (22.13)	124 (21.27)	20 (3.43)	5 (0.86)	
	2	50 (8.58)	84 (14.41)	117 (20.07)	171 (29.33)	98 (16.81)	31 (5.32)	15 (2.57)	
	1	57 (9.78)	42 (7.2)	102 (17.49)	119 (20.41)	136 (23.33)	52 (8.92)	52 (8.92)	
	0	357 (61.23)	388 (66.55)	253 (43.40)	164 (28.13)	225 (38.59)	480 (82.33)	511 (87.65)	
대안별 점수별 응답자 수	AL	3	38	20	36	34	35	1	1
		2	11	35	38	43	26	5	4
		1	18	12	31	28	41	14	16
		0	99	99	61	61	64	146	145
	AS	3	22	27	40	32	29	6	2
		2	15	21	37	48	22	9	3
		1	14	14	22	34	41	19	9
		0	108	97	60	45	67	125	145
	HL	3	36	10	10	26	30	2	1
		2	9	16	20	39	18	7	3
		1	9	4	27	26	26	5	14
		0	61	85	58	24	41	101	97
	HS	3	23	12	25	37	30	11	1
		2	15	12	22	41	32	10	5
		1	16	12	22	31	28	14	13
		0	89	107	74	34	53	108	124

주: 1) () 안은 583 명에 대한 중복 응답 비율(%)을 나타낸다.

2. 대안특성변수의 도출

본 논문에서 대안특성변수는 대안선택에 따르는 비용 개념으로 주택매매가격 또는 전·월세비용으로 설정하였다. 이 변수들의 문제는 대안특성변수에서 개인 당 실제로 선택된 대안의 가격(매매가격 또는 전월세비용)은 설문응답 자료를 통해 알 수 있으나 선택되지 않은 대안들의 가격에 대해서는 알 수 없다는 점이

다. 본 연구에서는 선택되지 않은 대안들의 가격을 적절한 회귀모형을 구성해 추정하여 추정치를 이용하였다. 대안특성변수의 계산 및 추정에 대해서는 제II 장 말미에 상술한다.

전술한 바와 같이 본 논문에서 선정한 대안은 주거형태-규모의 혼합대안이나 각각의 대안에는 주택을 보유한 사람들과 전·월세로 거주하는 사람들이 혼재되어 있으며, 그에 따라 개인이 실제로 지불한 대안별 비용이 주택가격, 전세, 월세로 다양하다. 따라서 개인들이 보고한 매매가, 전세, 월세 등을 대안별 가격으로 그대로 이용할 수 없어 저자들은 매매가와 월세를 전세등가액으로 환산하여 환산된 값을 대안별 가격으로 이용하기로 하였다.²⁾

먼저 매매가의 전세등가액은 매매가에 매매가 대비 전세가비율(이하 매매전세비율)을 곱한 값을 이용하였다. 손쉽게 이용할 수 있는 규모별 매매전세비율 통계가 없어 국민은행에서 발표하는 주택가격지수 통계를 이용하여 아래에 상술하는 절차를 통해 아파트 및 단독주택 규모별 매매전세비율을 추산하였다.³⁾

우선 앞의 국민은행 통계에서는 2007년 1월부터 2008년 12월 사이 전국평균 아파트 매매가격대비 전세비율 0.537과 2007년 1월부터 2008년 12월 사이 아파트 규모별 매매가격지수 및 전세가격지수를 이용해 동 기간의 규모별 전국 평균 아파트 매매전세비율의 추산치를 계산하였다.⁴⁾ 단독주택의 매매전세비율은 자료가 가용하지 않아 단독주택의 경우에도 아파트 매매전세비율 0.537을 그대로 적용하였고, 단독주택 규모별 매매가격지수 및 전세가격지수를 이용하여 추

2) 매매가, 전세, 월세 등 서로 다른 성격의 실제 지불액을 동일한 성격의 비용으로 통합하는 문제는 결합대안을 주거형태-규모로 구성하지 않고 주거형태-규모-보유형태로 구성해도 동일하게 발생한다. 이상적으로는 구매와 임차에 따른 기대이익흐름의 현재가치를 추정하는 방식으로 환산해야 하겠으나 이를 추정할 자료가 없어 본 논문에서는 이하에 상술하는 전세등가액의 추산이라는 다소 조악한 방식으로 환산하였다. 보다 완벽한 추산방식에 대한 연구는 추후 과제로 남긴다.

3) 국민은행 발표 주택가격지수 시계열자료 (http://land.kbstar.com/quics?asfilecode=5023&_nextPage=page=B008838)를 활용하여 추산하였다.

4) 정확한 대안특성변수의 추산을 위해서는 설문에 응한 사람들이 실제로 자신들의 주거형태-규모를 선택한 시점에서의 가격변수 및 가구소득을 이용하여야 할 것이다. 하지만 설문조사의 한계로 인해 각 응답자들이 주거형태-규모를 선택한 시점에 대한 가격자료 및 가구소득 자료를 구할 수 없어 연구가 진행되던 시점의 시장조사 결과를 이용하였다. 이런 제한점으로 인해 본 연구의 정량적 결과를 해석하는 데에는 많은 주의가 요구된다.

산하였다.

2007-2008년 종류별 규모별 매매가격지수(2008.12=100)와 전세가격지수(2008.12=100), 그리고 이들을 이용해 구한 매매전세비율의 추산치 다음의 <표 4>와 같다.

<표 4> 매매전세비율의 추산치

2007.1-2008.12 평균	아파트 대형	아파트 중형	아파트 소형	단독 대형	단독 중형	단독 소형
매매가격지수(A)	103.2	99.8	96.0	98.8	98.4	97.6
전세가격지수(B)	100.9	100.5	98.2	98.9	98.0	97.5
매매전세비율 추산치 (0.537*B/A)	0.525	0.541	0.549	0.538	0.535	0.526

국민은행 통계조사에서 아파트와 단독주택의 규모는 아파트의 경우 전용면적을 기준으로 62.8㎡ 미만은 소형, 62.8㎡ 이상~95.9㎡ 미만은 중형, 95.9㎡ 이상은 대형으로 구분하고, 단독주택의 경우 대지면적과 건물면적을 기준으로 다음의 <표 5>와 같이 구분하고 있다.

<표 5> 단독주택의 주택규모 분류

구 분		건물 면적		
		95.9㎡ 미만	95.9㎡ 이상 ~162.0㎡ 미만	162.0㎡ 이상
대지 면적	228.1㎡ 미만	소	중	대
	228.1㎡ 이상 ~327.3㎡ 미만	중	중	대
	327.3㎡ 이상	대	대	대

주: 1) 자료: 국민은행 발표 주택가격지수 시계열자료

(http://land.kbstar.com/quics?asfilecode=5023&_nextPage=page=B008838).

하지만 본 연구에서는 설문에 응한 사람들의 전용면적이나 대지면적 등의 정보가 가용하지 않아 응답자가 답한 평형정보만을 이용해 다음의 <표 6>과 같이 분류하여 위의 매매전세비율 추산치를 적용하였다.

<표 6> 평형정보에 따른 주택규모 분류

	대형	중형	소형
아파트	30평 이상	20평 이상-30평 미만	20평 미만
비아파트	50평 이상	30평 이상-50평 미만	30평 미만

예컨대 응답자가 45평의 비아파트에 거주 중이고 매매가격이 4억원이라고 답을 했을 경우 $4억 \times 0.5335 = 2억 1천 4백만원$ 이 매매가격의 전세등가액으로 간주된다.

월세의 전세등가액은 이창무외 2명(2007)의 식 (5-1)을 이용해 완전연세전세(RC)로 전환하는 방식을 이용하였다. 우선 설문조사에 보증금 자료가 없는 까닭에 동 논문의 <부록 3>과 <부록 4>에 있는 서울지역 전체의 단독다가구 및 연립다세대의 보증금(D) 대비 월세(M)비율의 평형별 평균을 구해 100으로 나누어 평형별 월세보증금 비율($M/D=f$) 값을 추정한 후, 설문조사에 응한 월세거주자의 월세비용을 앞에서 구한 f 값으로 나누어 D값을 계산하였다. 다음으로 동 논문의 <부록 5>와 <부록 6>의 서울지역 전체 통계를 이용해 단독다가구 및 연립다세대의 전월세 전환율의 평형별 평균값을 구한 후 이를 100으로 나누고 1을 더해 전월세전환율 r 을 구하였다. 다음의 <표 7>은 위의 과정을 거쳐 구한 평형별 f 값과 r 값이다.

<표 7> 평형별 f 와 r

평형	f	r
19평 이하	$0.01 \times (4.82 + 5.15) / 2 = 0.0499$	$1 + 0.01 \times (13.02 + 12.63) / 2 = 1.1283$
20평대	$0.01 \times (3.16 + 3.06) / 2 = 0.0311$	$1 + 0.01 \times (11.16 + 11.16) / 2 = 1.1116$
30평대	$0.01 \times (2.76 + 2.79) / 2 = 0.0278$	$1 + 0.01 \times (10.38 + 10.32) / 2 = 1.1035$
40평대	$0.01 \times (2.82 + 2.59) / 2 = 0.0271$	$1 + 0.01 \times (10.58 + 10.78) / 2 = 1.1068$

주: 1) 설문에 응한 사람들 중 50평 이상의 월세거주자가 없어 40평대까지만의 환산계수를 계산하였음.

본 논문에서 이 값들을 이용해 월세를 완전연세(RC)로 전환한 방식은 다음과 같다. 월세금액이 월 M원일 경우,

$$D = M/f \tag{1}$$

$$RC=12*M+D*r=M*(12+r/f) \tag{2}$$

이상의 방식으로 추산된 전세등가액(YRent_Equiv)은 선택된 대안의 가격의 의미를 가진다.

MNP모형에 대해 상술한 III절에서 알 수 있는 바와 같이, MNP모형을 추정하기 위해서는 선택되지 않은 대안의 가격을 추정해야 하는데 본 연구에서는 실제 지불된 주택가격, 월세 등을 앞에서 상술한 방식으로 전세등가액으로 환산한 후 이를 지역별, 대안별로 평균을 구해 (실제로는 $\ln(\text{YRent_Equiv})$ 의 평균) 그 값을 선택되지 않은 대안의 가격으로 간주하였다. 다음의 <표 8>은 그 평균값들을 구하기 위해 $\ln(\text{YRent_Equiv})$ 를 설문조사가 실시된 지역 더미변수(총 6개 지역, D1 ~ D6)와 선택대안을 선정할 때 사용한 기준에 맞추어 판단한 중소형 더미(S, 중소형일 경우 1), 아파트 더미(A, 아파트일 경우 1), 그리고 각 변수들의 상호작용 변수를 설명변수로 하여 선행회귀모형을 추정한 것이다.⁵⁾

이 표를 이용하면, 예컨대, D1 지역의 소형 아파트(AS)의 $\ln(\text{YRent_Equiv})$ 의 평균값은 $8.446 + 0.316 - 0.286 - 0.213 + 0.923 + 0.489 - 0.328 + 0.066 = 9.410$, AL 대안의 $\ln(\text{YRent_Equiv})$ 평균값은 $8.446 - 0.286 + 0.923 + 0.489 = 9.572$, HS 대안의 $\ln(\text{YRent_Equiv})$ 평균값은 $8.446 + 0.316 + 0.923 - 0.328 = 9.357$, HL 대안의 $\ln(\text{YRent_Equiv})$ 평균값은 $8.446 + 0.923 = 9.369$ 가 된다. 따라서 D1 지역에 거주하는 어떤 응답자가 AS를 선택했을 경우 그 응답자의 AS 대안의 $\ln(\text{YRent_Equiv})$ 는 실제 지불한 주택가격 또는 전월세가격의 전세등가액을 사용하지만 선택하지 않은 AL, HS, HL 대안의 $\ln(\text{YRent_Equiv})$ 값으로는 각각 9.547, 9.357, 9.369를 이용

5) 지역 더미변수 D1-D6는 각각 서울 및 수도권, 인천, 광주, 대구, 부산, 대전 등 6대 광역 도시권을 나타낸다.

하게 된다.

<표 8> 전세등가액(YRent_Equiv) 추산식

종속변수: ln(YRent_Equiv)				
설명변수	계수추정치	표준오차	t-값	p-값
상수항	8.446	0.101	83.41	0
S	0.316	0.125	2.52	0.012
A	-0.286	0.131	-2.18	0.03
A*S	-0.216	0.168	-1.28	0.2
D1	0.923	0.134	6.9	0
D1*A	0.489	0.184	2.66	0.008
D1*S	-0.328	0.174	-1.88	0.06
D1*A*S	0.066	0.241	0.28	0.783
D2	0.455	0.418	1.09	0.276
D2*A	0.027	0.444	0.06	0.951
D2*S	-0.603	0.445	-1.36	0.175
D2*A*S	0.484	0.512	0.94	0.345
D3	1.058	0.255	4.15	0
D3*A	-1.042	0.293	-3.55	0
D3*S	-0.994	0.371	-2.68	0.008
D3*A*S	1.081	0.427	2.53	0.012
D4	0.804	0.216	3.72	0
D4*A	-0.274	0.263	-1.04	0.299
D4*S	-0.999	0.30	-3.33	0.001
D4*A*S	1.051	0.382	2.75	0.006
D5	0.592	0.157	3.78	0
D5*A	-0.207	0.228	-0.91	0.363
D5*S	-0.368	0.442	-0.83	0.406
D5*A*S	0.438	0.511	0.86	0.392
Adj R ²	0.484			
F(23,559)	22.83			
표본 수	583			

III. 다항프로빗모형

이산선택모형에서 흔히 이용되는 확률효용함수 모형을 이용해 다항프로빗모형의 설정과 추정에 대해 설명해 보자.⁶⁾ 확률효용함수 모형에서는 개인 i 는 대안 j 를 선택하는 경우 효용 U_{ij} 를 얻는다고 가정하고, 개인은 각 대안으로부터의 효용을 비교해 최대의 효용을 주는 대안을 선택한다고 간주한다. 본 연구에서는 개인의 효용 U_{ij} 가 다음과 같은 형태를 갖는다고 가정한다.

$$U_{ij} = \beta_{0j} + (X_i', S_i')\beta_{1j} + P_{ij}'\gamma + e_{ij}, \quad i = 1, \dots, N, \quad j = 0, \dots, J \quad (3)$$

여기서 i 는 개인, j 는 대안을 말하며 S_i 는 주거환경요인변수벡터 ($p \times 1$), X_i 는 주거환경요인변수를 제외한 개인별 변수벡터 ($r \times 1$), P_{ij} 는 대안특성변수벡터 ($q \times 1$)이다. 그리고 β_{0j} 는 대안 j 의 확률효용함수의 상수항이다.

개인 i 는 가능한 $J+1$ 개의 대안 중에서 효용이 가장 큰 대안을 선택하므로, 개인 i 의 선택을 C_i 라고 하면 개인의 선택함수는 다음과 같다.

$$C_i = s \in \{0, 1, \dots, J\} \text{ iff } \max_{j=0, \dots, J} U_{ij} = U_{is} \quad (4)$$

여기서 대안 0은 기준대안(baseline alternative)이라고 불린다.

통상적인 2변수(bivariate) 프로빗 모형과 유사한 이유로 위의 확률효용함수 체계는 식별되지 못하며 대신 다음과 같이 대안간의 효용의 차이들만 식별될 수 있다. i 가 대안 j 로부터 얻는 효용과 대안 0으로부터 얻는 효용의 차 $V_{ij} = U_{ij} - U_{i0}$ 를 정의하고 두 대안 간의 교란항의 차 $\epsilon_{ij} = e_{ij} - e_{i0}$ 를 정의하여 위의 확률효용함수 모형을 다음과 같이 나타내자.

6) 보다 자세한 내용은 Cameron and Trivedi(2007)을 참조하기 바란다.

$$\begin{aligned}
 V_{ij} &= \beta_{0j} - \beta_{00} + (X_i', S_i')(\beta_{1j} - \beta_{10}) + (P_{ij} - P_{i0})\delta + \epsilon_{ij} \\
 &= \gamma_{0j} + (X_i', S_i')\gamma_{1j} + W_{ij}'\delta + \epsilon_{ij}, \quad i = 1, \dots, N, j = 1, \dots, J
 \end{aligned}
 \tag{5}$$

여기서 γ_{0j} 는 대안 j 와 기준대안 간의 상수항의 차이, γ_{1j} 는 대안 j 와 기준대안의 공변량(covariates) 벡터 (X_i', S_i') 의 계수값의 차이가 되며 이들 값들은 대안별로 차이가 있을 수 있다. 한편 δ 는 대안 j 와 기준대안간의 대안특성변수의 차이값의 계수로 기준대안을 제외한 총 J 개의 대안 전체에 걸쳐 하나의 값을 갖는다.

V_{ij} 를 이용하여 개인 i 의 선택함수를 다시 표시하면 다음과 같다.

$$C_i = \begin{cases} s \in \{1, 2, \dots, J\} & \text{if } \max_{j=1, \dots, J} V_{ij} = V_{is} > 0; \\ 0 & \text{otherwise.} \end{cases}
 \tag{6}$$

따라서 i 가 대안 $s > 0$ 를 선택할 확률은

$$\begin{aligned}
 &\Pr[C_i = s | X_i, S_i, P_{ij}, j = 0, \dots, J] \\
 &= \Pr[\max_{j=1, \dots, J} V_{ij} = V_{is} > 0 | X_i, S_i, P_{ij}, j = 0, \dots, J]
 \end{aligned}
 \tag{7}$$

한편 i 가 기준대안을 선택할 확률은

$$\begin{aligned}
 &\Pr[C_i = 0 | X_i, S_i, P_{ij}, j = 0, \dots, J] \\
 &= \Pr[\max_{j=1, \dots, J} V_{ij} \leq 0 | X_i, S_i, P_{ij}, j = 0, \dots, J]
 \end{aligned}
 \tag{8}$$

MNP에서는 교란항 벡터 $\epsilon_i = (\epsilon_{i1}, \dots, \epsilon_{iJ})'$ 가 평균이 $0_{(J \times 1)}$ 이고 분산-공분산 행렬이 Σ 인 다변량 정규분포를 한다는 가정 하에 위의 확률효용함수의 모수들과 분산공분산행렬을 추정한다. 만일 원래의 교란항벡터 (e_{i0}, \dots, e_{iJ}) 의 각 원소들이 상호독립적이라면 Σ 의 대각의 원소는 모두 동일한 값인 $Var(e_{i0}) > 0$

가 된다. 따라서 Σ 의 대각외 원소들이 동일한 양의 값이라는 가설이 기각되면 각 대안간의 독립성 가정은 기각된다.

2변수 프로빗 모형에서 교란항의 분산을 1로 정규화한 계수추정치를 얻는 것과 같은 이유로 MNP에서도 여러 가지 방법으로 정규화한 계수추정치를 얻는다. 본 논문에서는 분산-공분산 행렬 Σ 의 (1, 1) 원소를 1로 정규화하는 방식을 취하였다.

MNL 모형이 통상적인 최우추정법을 이용하여 추정할 수 있으며 많은 계량경제분석 소프트웨어에서 추정에 관련된 명령어를 지원하는 데 비해 MNP모형에서는 우도함수를 구성하는데 필요한 식 (7) 과 (8)의 해석적 형태가 주어지지 않고 적분형태로만 구해지는 까닭에 통상적인 최우추정법이 아닌 simulated maximum likelihood estimation(SMLE)을 이용하거나 Bayesian 추정법의 한 가지인 Gibbs 샘플링을 이용한 마코프 체인 몬테카를로 기법(Markov chain Monte Carlo; MCMC)으로 추정한다.

이 방법들 중, McCulloch and Rossi(1994)와 Geweke et al.(1994)에 의하면, SMLE를 이용한 추정보다는 Gibbs sampling을 이용한 MCMC 추정방법이 모형의 계수 추정량의 finite sample 분포를 보다 더 정확하게 얻을 수 있다. 이런 이유로 본 연구에서는 Gibbs sampling을 이용한 MCMC 기법을 이용하였으며 김연배 · 신성윤(2002)과 Kim et al.(2003) 등의 연구에서도 MNP 모형을 추정하는 데에 Gibbs sampling 기법을 이용한 바 있다.

본 연구에서는 Imai and van Dyk(2005)의 방법을 구현한 통계패키지 R의 MNP package를 이용하였다. Gibbs sampling에 관한 보다 세부적인 논의는 부록으로 첨부한다. 저자들의 code는 저자들에게 문의 시 가용하다.

IV. 실증분석 결과

확률효용함수 모형을 추정하기 위해 사용한 대안특성변수는 위에서 구한 전세등가액(YRent_Equiv)을 가구소득(Inc_H)으로 나눈 값의 자연대수값

($\ln(YRent_Equiv/Inc_H)$)을 이용하였다. 이 값이 클수록 가구소득에 비해 선택대안들의 비용이 비싸다는 것을 의미한다.

개인특성변수로는 주택보유여부 더미변수(O, 자가주택 일 경우 1)와 가구주 연령(Age)을 이용하였다. 설문조사 상 가구주 연령을 직접 묻지 않은 관계로 응답자가 가구주가 아닌 경우에는 4를 더한 후 10으로 나누어 이를 가구주 연령으로 간주하여 사용하였다. 기타 응답자의 교육수준, 가구구성원 수, 거주지역이 도시인지 여부, 응답자의 직업 등의 변수들이 포함될 수 있으나 분석결과 추정에 포함된 모든 대안에서 유의성이 부족하거나 Gibbs sampling과정에서 오류가 발생하는 등의 문제가 있었다.⁷⁾ MNP의 속성상 하나의 개인특성변수에 J개의 계수가 추정되는 까닭에 모형에 포함되는 변수를 줄일 필요성이 있기도 하여 위의 두 변수만을 포함하였다. 이하의 추정에서 기준대안은 'HL'을 이용하였으며, HL이 기준대안임을 명확하게 하기 위해 이하의 분석에서 대안들은 AL:HL, AS:HL, HS:HL와 같이 표기하였다.

1. 기본 모형: 주거환경요인변수를 포함하지 않은 모형

우선 본 소절에서는 앞서 언급한 변수들만으로 구성된 주거환경요인변수를 포함하지 않은 기본 모형의 추정결과는 다음의 <표 9>와 같다.

<표 9> 주거환경요인변수를 제외한 모형 추정결과

	AL:HL	AS:HL	HS:HL
Intercept	-1.111*	0.176	0.011
O	0.929*	0.315*	-0.214
Age	0.042	-0.082*	-0.010
$\ln(YRent_Equiv/Inc_H)$	-0.217*		

7) 특정한 대안모형을 MNP 로 추정하는 경우 계수의 유의성이 없다는 것의 의미는 해당 변수가 그 대안모형에서 설명력이 없다는 것이 아니라 해당변수의 계수가 그 대안모형과 기준대안 모형에서 차이가 없다는 것을 의미하는 것이다. 따라서 어떤 변수가 추정에 포함된 모든 대안에서 유의적이지 않다는 것이 그 변수가 특정한 대안의 의미있는 설명변수일 개연성까지 부정하는 것은 아니다.

주: 1) * 표시된 변수들은 해당 파라미터의 posterior 분포의 95% credible region에 0이 포함되지 않는다는 것을 나타냄. 이는 전통적인 추정결과 해석에서 5%에서 유의적이라는 것과 유사한 의미를 가지는 까닭에 이후의 논의에서는 ‘유의적’이라는 표현을 사용함. 자세한 내용은 <부록>을 참조하기 바람.

$\ln(\text{YRent_Equiv}/\text{Inc_H})$ 의 계수추정치는 -0.217로 음(-)의 값을 보이고 있으며 유의수준 5%에서 유의적이다. 이 변수의 계수가 음(-)이라는 것은 식 (5)에서의 $\delta = \partial E[V_{ij}|X_i, S_i, W_{ij}] / \partial W_{ij}$ 가 음(-)이라는 것을 의미하므로, P_{ij} 가 P_{i0} 에 비해 줄어들 경우 $V_{ij} = U_{ij} - U_{i0}$ 는 평균적으로 증가함을 말해준다. 이는, 대안 j 의 가구소득 대비 전세등가액(주거비용)이 기준대안에 비해 줄어들 경우 기준대안에 비해 대안 j 를 선택할 확률이 커짐을 의미한다. 다른 변수의 계수 추정치에 대한 설명은 주거환경요인변수를 포함한 모형의 추정결과를 설명할 때 하기로 한다. 한편 이 모형의 분산-공분산 행렬 추정결과는 다음의 <표 10>과 같다.

<표 10> 주거환경요인변수를 제외한 모형의 분산공분산 행렬 추정 결과

	AL:HL	AS:HL	HS:HL
AL:HL	1	0.515*	-0.113
AS:HL		0.302*	-0.080
HS:HL			0.040*

분산-공분산행렬의 비대각요소 추정결과는 0.515, -0.113, -0.080으로 0.515는 유의수준 0,05에서 0과 유의적으로 다른데 비해 다른 두 요소 추정치는 0과 유의적으로 다르지 않으며 부호 또한 음(+)으로 추정되었다. 이 결과는, 비록 엄밀한 통계적 검정은 아니지만, 대안간의 독립성가정이 충족되지 않음을 시사하며 대안간의 독립성을 가정하지 않는 MNP 추정이 독립성을 가정하는 MNL모형에 비해 더 현실적임을 시사한다.

2. 주거환경요인변수를 포함한 모형

본 모형에 주거환경요인변수를 포함한 추정결과는 다음과 같다. 총 7가지의

주거환경요인변수 중에서 P1-P5까지 5가지의 변수만을 포함하였다. 주거환경요인변수를 구성한 방법상 7개의 변수를 모두 더하면 상수항과 완전다중공선성이 발생하므로($\sum_{k=1}^7 P_{ki} = 6$) 적어도 1개의 변수를 제외하여야 할 뿐 아니라 P6과 P7은 분석에 이용된 응답자중 각각 82.33%와 87.65%가 0의 응답을 하여 0의 응답이 28-67%에 불과한 다른 변수들에 비해 상대적인 중요성이 약하다고 판단하였기 때문이다. 다음의 <표 11>은 주거환경요인변수를 포함한 모형의 추정 결과이다.

<표 11> 주거환경요인변수를 포함한 모형 추정 결과

	AL:HL	AS:HL	HS:HL
Intercept	-3.295*	-0.02	-0.145*
O	1.642*	-0.01	0.058*
Age	0.317*	-0.005	0.011*
P1	0.06	-0.037*	0.007
P2	0.017	0.020*	0.009*
P3	0.255*	0.016	0.010*
P4	0.024	-0.017	0.006
P5	0.098	-0.005	0.007
ln(YRent_Equiv/Inc_H)	-0.093*		

주거환경요인변수를 포함한 경우에도 ln(YRent_Equiv/Inc_H)의 계수는 유의적으로 음(-)의 값을 가지는 것으로 추정된다. <표 9>에서와 마찬가지로 대안 j 의 전세등가액이 기준대안의 전세등가액에 비해 줄어들 경우 기준대안에 비해 대안 j 를 선택할 확률이 커짐을 의미한다.

개인별 변수인 주택보유여부(O)는 대부분 유의적인 것으로 나타났다. 이는 주택을 보유하기로 한 사람의 의사결정은 주택을 임대하기로 한 사람의 의사결정과 다르게 이루어짐을 의미한다. AL:HL과 HS:HL이 유의적으로 양인데 이는 주택을 보유하기로 마음을 먹고 있는 사람은 다른 조건이 동일하고 대안들의 비용이 동일한 경우 HL보다는 AL이나 HS를 더 선호한다는 것을 말해준다.

주거환경요인변수의 계수는, 다른 조건이 동일할 때, 즉 대안의 가격이 동일한 경우 각 변수가 각각의 대안과 기준대안 사이의 선택에서 어떤 역할을 하는지

보여준다. 계수 추정치가 양(+)인 경우 해당 주거환경요인변수를 중요하게 생각할수록 기준대안이 아닌 다른 대안에서 얻는 효용이 커짐을 의미하며, 계수 추정치가 음(-)인 경우 해당 주거환경요인변수를 중요하게 고려할수록 기준대안에서 얻는 효용이 커짐을 의미한다. 계수추정치가 0인 경우에는 해당 주거환경요인변수를 중요하게 생각하는 정도가 기준대안과 특정 대안의 비교에서 영향을 주지 않음을 말한다.

구입당시 주택가격(P1)은 AS:HL에서는 유의적으로 양(-)의 값을 갖는 것으로 추정되는 데에 비해 AL:HL 그리고 HS:HL에서는 유의적이지 않았다. 따라서 개별 가구가 P1을 중요하게 고려할수록 HL에 비해 AS를 선택할 가능성이 높지만 개별가구가 AL과 HL를 비교할 때에나 HS와 HL을 비교할 때에는 P1은 거의 영향을 주지 않는다고 할 수 있다.

지가상승기대(P2)은 개별가구가 AS와 HL을 비교할 때와 HS와 HL을 비교할 때에는 중요한 고려요소가 되며 P2를 중요하게 생각하는 가구일수록 HL에 비해 AS나 HS에서 얻는 효용이 높다고 할 수 있다. 하지만 P2는 AL과 HL을 비교할 때에는 거의 영향을 주지 않는 것으로 나타났다.

교육여건(P3)의 계수는 AL:HL과 HS:HL에서 유의적으로 양(+)의 값을 가진다. 따라서 교육여건을 중요하게 생각하는 가구일수록 HL보다는 AL 혹은 HS의 효용이 높음을 알 수 있다.

다른 두 환경변수(교통 편의성(P4), 환경의 쾌적성(P5))는 대안들간의 선택에서 특별히 더 중요한 역할을 하지 않는 것으로 드러났다. 모형의 해석 상 주의할 점은 MNP 추정 시 계수의 유의성이 없다는 의미는 해당 변수가 설명력이 없다는 것이 아닐 수 있다는 점이다. MNP모형은 특정 대안 모형과 기준대안 모형의 차이를 추정하는 까닭에 계수의 유의성이 없다는 것은 단순히 해당변수의 계수가 특정한 대안모형과 기준대안 모형에서 차이가 없다는 것을 의미한다. 따라서 유의적이지 않은 주거환경요인변수 2가지는 구매자가 구매행위를 할 때 영향을 미치지 않는다는 것이 아니라 AS, AL, HS, HL의 4가지 대안에 동일한 영향을 준다는 것을 의미한다. 이 모형의 분산-공분산 행렬 추정결과는 다음의 <표 12>와 같다.

<표 12> 주거환경요인변수를 포함한 모형의 분산공분산행렬 추정 결과

	AL:HL	AS:HL	HS:HL
AL:HL	1	0.089*	0.066*
AS:HL		0.036*	-0.001
HS:HL			0.006*

<표 10>에서와 유사하게 <표 12>에 나타난 분산공분산 행렬의 대각의 요소 추정치들 또한 유의수준 0.05에서 유의적인 요소와 유의적이지 않은 요소가 혼재되어 있고, 부호도 다른 바, 역시 대안간의 독립성가정이 충족되지 않음을 시사한다.

V. 결 론

본 연구는 개별가구가 거주형태별로 아파트-중소형(AS), 아파트-중대형(AL), 비아파트-중소형(HS), 비아파트-중대형(HL)의 선택대안 중 하나의 대안을 선택하는 것으로 보고 개별 가구의 대안선택에 어떤 요소가 영향을 주는지를 알아보기 위해 다항프로빗(MNP) 모형을 수립 Gibbs sampling 방법으로 추정하였다.

각 대안에서 주택의 소유형태에 따른 차이(소유, 전세, 월세)를 보정하기 위해 주택구매가격 또는 월세액의 전세등가액을 구해 대안간 비용에 해당하는 ‘전세등가액과 가구소득의 비’를 이용하였고, 선택되지 않은 대안들의 ‘전세등가액과 가구소득의 비’는 회귀분석을 이용해 추정하였다. 실증분석의 결과로부터 다음과 같은 사항들이 확인할 수 있다.

첫째, 대안특성변수($\ln(\text{YRent_Equiv}/\text{Inc_H})$)의 계수추정치는 두 모형에서 모두 유의적이고 음(-)이다. 대안특성변수가 각 대안들의 비용임을 고려하면, 이 추정결과는 기준대안이 아닌 다른 대안의 비용이 기준대안보다 높아지면 소비자가 그 대안을 선택할 가능성이 떨어지고 기준대안을 선택할 가능성이 높아진다는 것을 말해준다. 이는 소비자이론과 일치하는 결과라고 할 수 있다.

둘째, 주택을 보유하기로 한 사람의 의사결정은 주택을 임대하기로 한 사람의 의사결정과 다르게 이루어지는 것으로 보이며, 이는 주택을 보유하기로 마음을 먹고 있는 사람은 대안들의 비용이 동일한 경우 AS보다는 HL을 더 선호하고 AS보다 HS를 덜 선호한다.

셋째, 7가지의 주거환경요인변수 중에서 구입당시의 가격(P1), 지가상승의 기대감(P2), 그리고 교육여건(P3) 유의수준 5%에서 유의적인 계수값이 존재하였고 나머지 변수들은 유의수준 5%에서 유의적이지 않았다.

넷째, 주거환경요인변수를 포함한 모형과 포함하지 않은 모형에서 모두 분산-공분산 행렬의 대각외 원소 추정치들은 서로 다른 것으로 나타났다. 이는 대안들간의 독립성 가정(IIA)이 현실의 선택을 모형화 하는데 지나치게 큰 제약을 주는 것임을 말해주며, MNP등 IIA가정을 완화한 모형을 세우는 것이 보다 현실적임을 말해준다.

하지만 다음과 같은 본 연구의 한계로 인해 본 연구의 결과를 그대로 소비자들의 주택유형-규모 선택 문제에 직접적으로 적용하는 데에는 주의가 요구된다. 첫째, 본 연구에 사용된 자료는 2006년 6월에 행해진 설문조사에서 얻어진 것이므로 최신의 자료를 사용할 경우에도 동일한 결과가 도출될 것인지에 대해 면밀한 검토 및 추가적 연구가 필요하다.

둘째, 본 연구에서 사용된 대안특성변수는 2007년 1월부터 2008년 12월까지 가격자료의 평균값으로 구한 것으로, 설문에 응한 사람들의 주택형태-규모의 선택시점과는 차이가 있다. 따라서 본 연구에서 얻어진 결과는 일종의 경향성으로 해석하여야 하며 이를 정량적으로 해석하기에는 무리가 있다.

셋째, 본 연구는 주택의 소유형태를 주택의 거주형태 선택에서 분리된 독립변수(O)인 것으로 간주하였다. 가용한 자료의 한계로 인해 이와 같이 모형화 하였으나 향후 가용자료가 충분히 구해지는 경우에는 대안의 구성 단계에서 주택의 소유까지 고려해 대안을 구성하고 개별가구의 대안선택 행위를 분석하는 것이 보다 실질적인 연구가 될 것이다.

넷째, 본 연구에서 사용한 대안특성변수는 개별가구가 느끼는 대안간의 상대적 가격 또는 비용의 개념이다. 이 중 선택된 대안의 비용만이 실제로 연구자에게 관측되고 선택되지 않은 대안에 대한 비용은 연구자가 추산할 수밖에 없다.

비록 본 연구에서 이용한 추산과정을 상세하게 기술해 두었으나 단순한 회귀분석에 의한 추정이 아니라 보다 경제학적으로 엄밀한 추정방법을 연구하고 적용하는 것 또한 본 연구에서 하지 못한 의미있는 연구가 될 것이다.

參 考 文 獻

- 김연배 · 신성윤, “Multinomial probit 모형을 이용한 가정용 난방연료 선택에 관한 연구”, 『자원환경경제연구』, 제11권 제4호, 2002, pp. 609-632.
- 김재홍, “명시선호실험을 이용한 소득계층별 주거입지 속성가치의 격차 분석”, 『지역연구』, 제22권 제2호, 2006, pp. 3-26.
- 김정수 · 이주형, “가구특성에 따른 주택선택행태에 관한 연구”, 『국토계획』, 제39권 제1호, 2004, pp. 191-204.
- 김한수 · 임준홍 · 송홍수, “도심 주거지 선호성향에 관한 연구”, 『주택연구』, 제6권 제1호, 1999, pp. 137-155.
- 이창무 · 이진호 · 임성은, “단독 · 다가구 및 연립 · 다세대 임대시장분석-매물 DATA를 사용하여”, 『부동산학연구』, 제13집 제1호, 2007, pp. 25-47.
- 이길순, “로짓모형을 이용한 주택선택 결정요인 연구”, 연세대학교 대학원 박사학위 논문, 1990.
- 이충기 · 이주석, “주거환경이 주택유형과 주택규모 선택에 미치는 영향 분석-다항 로짓모형을 활용하여”, 『경제학연구』, 제56집 제3호, 2008, pp. 55-73.
- 장성수 · 윤혜정, “아파트선택가구의 사회 · 경제 및 주거특성에 관한 연구”, 『주택연구』, 제6권 제2호, 1998, pp. 5-32.
- 정의철, “도시가구의 주택접유형태 및 주택유형선택에 관한 연구”, 『주택연구』, 제10권 제1호, 2002, pp. 5-31.
- 최막중 · 임영진, “가구특성에 따른 주거입지 및 주택유형 수요에 관한 실증분석”, 『국토계획』, 제36권 제6호, 2001, pp. 69-81.

- Albert, J. H. and S. Chib, “Bayesian analysis of binary and polychotomous response data”, *Journal of the American Statistical Association*, Vol. 88, 1993, pp. 669-679.
- Bhat, C. R., “A Heteroscedastic Extreme Value Model of Intercity Travel Mode Choice”, *Transportation Research*, Vol. 29B, 1995, pp. 471-483.
- Cameron, A. C. and P. Trivedi, *Microeconometrics-Method and Applications*, Cambridge University Press, NY, USA. 2007.
- Casella, G. and E. I. George, “Explaining the Gibbs Sampler”, *The American Statistician*, Vol. 46, 1992, pp. 167-174.
- Cho, C. J., “Joint Choice of Tenure and Dwelling Type: a Multinomial Logit Analysis for the City of Chongju”, *Urban Studies*, Vol. 34, No. 9, 1997, pp. 1459-1473.
- Geweke, J., M. Keane, and D. Runkle, “Alternative Computational Approaches to Inference in the Multinomial Probit Model”, *The Review of Economics and Statistics*, Vol. 76, 1994, pp. 609-632.
- Huang, Y. and W. A. V. Clark, “Housing Tenure Choice in Transitional Urban China: A Multilevel Analysis”, *Urban Studies*, Vol. 39, No. 1, 2002, pp. 7-32.
- Imai, K. and D. A. V. Dyk, “A Bayesian Analysis of the Multinomial Probit Model Using Marginal Data Augmentation”, *Journal of Econometrics*, Vol. 124, 2005, pp. 311-334.
- Keane, M., “A Computationally Practical Simulation Estimator for Panel Data”, *Econometrica*, Vol. 62, 1994, pp. 95-116.
- Kim, Y., T-T Kim, and E. Heo, “Bayesian Estimation of Multinomial Probit Models of Work Trip Choice”, *Transportation*, Vol. 30. 2003, pp. 351-365.
- McConnell, K. E. and M. A. Rosado, “Valuing Discrete Improvements in Drinking

Water Quality through Revealed Preferences”, *Water Resources Research*, Vol. 36, 2000, pp. 1575-1582.

McCulloch, R. E. and P. E. Rossi, “An Exact Likelihood Analysis of the Multinomial Probit Model”, *Journal of Econometrics*, Vol. 64, 1994, pp. 207-240.

McCulloch, R. E., N. G. Polson and P. E. Rossi, “A Bayesian Analysis of the Multinomial Probit Model with Fully Identified Parameters”, *Journal of Econometrics*, Vol. 99, 2000, pp. 173-193.

McFadden, D., *Econometric Analysis of Qualitative Response Models, in Handbook of Econometrics*, Griliches, Z. and Intriligator, M (eds) North-Holland, New York, pp. 1395-1457. 1984.

McFadden, D. “A Method of Simulated Moments for Estimation of Discrete Response Models without Numerical Integration”, *Econometrica*, Vol. 57, 1989, pp. 995-1026.

Nobile, A., “A Hybrid Markov Chain for the Bayesian Analysis of the Multinomial Probit Model”, *Statistics and Computing*, Vol. 8, 1998, pp. 229-242.

Train, K., *Qualitative Choice Analysis: Theory, Econometrics and an Application to Automobile Demand*, Cambridge, Massachusetts: The MIT Press, 1986.

Wooldridge, J. M., *Econometric Analysis of Cross Section and Panel Data*, The MIT Press, Cambridge, MA, USA, 2002.

〈부록〉 Gibbs sampler를 이용한 MNP 모형의 Bayesian 추정

Bayesian estimation에서 모형의 추정이란 모형의 파라미터 벡터 $\theta = (\theta_1, \theta_2)$ 의 사후분포(posterior distribution)를 구하는 것이다. θ 의 사전분포(prior distribution)를 $p(\theta_1, \theta_2)$ 라 하고, 주어진 θ 에 대해 얻어지는 자료의 분포(즉, 최우추정법에서 말하는 우도함수)를 $p(Y|X, S, P, \theta_1, \theta_2)$ 라 하면 Bayes의 법칙에 따라 θ 의 사후분포 $p(\theta_1, \theta_2|X, S, P, Y)$ 는 다음과 같이 나타난다.

$$p(\theta_1, \theta_2|X, S, P, Y) \propto p(Y|X, S, P, \theta_1, \theta_2)p(\theta_1, \theta_2)$$

Bayesian estimation에서 흔히 이용되는 방법은 Markov Chain Monte Carlo(MCMC) 방법이다. 이는 s-번째의 $p(\theta_1, \theta_2|X, S, P, Y)$ 의 추정치를 s+1-번째의 추정을 위한 사전분포로 이용하여 반복적으로 $p(\theta_1, \theta_2|X, S, P, Y)$ 를 추정하는 기법이며 반복적으로 추정되는 사후분포 $p(\theta_1, \theta_2|X, S, P, Y)$ 이 특정한 기준에 의해 수렴할 때까지 반복이 진행된다. Gibbs sampling에서는 이러한 MCMC 방법을 사용하는 과정에서 (θ_1, θ_2) 의 결합사후분포(joint posterior distribution)를 직접 구하지 않고 $\theta_1|\theta_2$ 와 $\theta_2|\theta_1$ 의 조건부사후분포(conditional posterior distribution)를 구하고 이를 이용해 (θ_1, θ_2) 의 결합사후분포(joint posterior distribution)를 구하는 과정을 취한다(Casella and George, 1992).

Gibbs sampling을 이용한 MNP모형의 Bayesian 추정의 개략적 절차는 다음과 같다. 먼저 식 (5)를 $j = 1, \dots, J$ 에 대해 중첩하여(stacking) 이를 벡터와 매트릭스 형태로 나타낼 수 있다.

$$V_i = R_i' \theta + \epsilon_i, \quad \epsilon_i \sim N(0_{(J \times 1)}, \Sigma), \quad i = 1, \dots, N..$$

이제 모형의 모수 $\theta = (\gamma_{01}, \dots, \gamma_{0J}, \gamma_{11}, \dots, \gamma_{1J}, \delta)'$ 의 사전분포를 평균 $\bar{\theta}$ 이고 공

분산행렬 A^{-1} 인 정규분포로 설정하고 모수 $\Sigma^{-1}(\equiv G)$ 의 사전분포를 모수 v 와 M 을 갖는 Wishart 분포로 설정하면 각각의 사전분포는 다음과 같다.

$$p(\theta|\bar{\theta}, A) \propto |A|^{0.5} \exp\left(-\frac{1}{2}(\theta - \bar{\theta})' A (\theta - \bar{\theta})\right),$$

$$p(G|v, M) \propto |G|^{(v-J-1)/2} \exp\left(-\frac{1}{2} GM\right).$$

V_i 는 관측되지 않는 변수이나 식 (5)는 주어진 θ 와 G 를 이용하여 V_i 를 임의 추출(random draw) 할 수 있음을 말해준다. 즉,

$$V_i|R_i \sim N(R_i'\theta, G^{-1}), i = 1, \dots, N. \quad (9)$$

이제 위의 사전분포와 임의추출된 V_i 를 이용하여 θ 와 G 의 사후분포를 구하면 각각 다음과 같다.

$$\theta|V, G \sim N(\hat{\theta}, \Sigma_\theta),$$

$$\Sigma_\theta = \left(\sum_{i=1}^n R_i' G R_i + A\right)^{-1}, \hat{\theta} = \Sigma_\theta \left(\sum_{i=1}^N R_i' G V_i + A\bar{\theta}\right),$$

$$G|\theta, V \sim \text{Wishart}(v + N, M + \sum_{i=1}^N (V_i - R_i'\theta)(V_i - R_i'\theta)').$$

주의할 점은 $G = \Sigma^{-1}$ 이고 Σ 의 (1,1)번째 원소를 1로 정규화 하므로 이를 만족하도록 G 또한 정규화 하여야 한다.

이상의 단계에서 사후분포를 구하는 과정을 보면 θ 의 사후분포를 구하는 경우 G 의 정보를 이용하고, 반대로 G 의 사후분포를 구하는 경우 θ 의 정보를 이용함을 알 수 있는데 이 과정을 Gibbs sampling이라 한다.

모형 추정은 다음의 단계를 거쳐 이루어진다.

< Step 1 > $\bar{\theta}, A, v, M$ 에 대해 임의의 초기값을 부여.

< Step 2 > 다음의 사전분포를 이용하여 첫번째 θ 와 G 를 추출:

$$p(\theta|\bar{\theta}, A) \propto \text{Ken} A \text{Ken}^{0.5} \exp\left(-\frac{1}{2}(\theta - \bar{\theta})' A (\theta - \bar{\theta})\right);$$

$$p(G|v, M) \propto \text{Ken} G \text{Ken}^{(v-J-1)/2} \exp\left(-\frac{1}{2} GM\right).$$

< Step 3 > 모수 θ 와 G 의 s -번째 추출을 $\theta^{(s)}$ 와 $G^{(s)}$ 라고 하면, $G^{(s)}$ 를 이용하여 $\epsilon_i^{(s)}$ 를 $N(0_{(J \times 1)}, (G^{(s)})^{-1})$ 로부터 추출하고 이를 $R_i' \theta^{(s)}$ 과 더해 $V_i, i = 1, \dots, N$ 의 s -번째 추출 $V_i^{(s)} = R_i' \theta^{(s)} + \epsilon_i^{(s)}, i = 1, \dots, N$ 를 구한다. 이때 $V_i^{(s)}$ 는 i 의 실제의 선택 C_i 와 일관되게 추출되어야 한다. 즉, 예컨대 $C_i = k > 0$ 이라면 $\max\{V_{i1}^{(s)}, \dots, V_{ik}^{(s)}, \dots, V_{iJ}^{(s)}\} = V_{ik}^{(s)} > 0$ 이 되도록 하여야 하며, 만일 $C_i = 0$ 이라면 $\max\{V_{i1}^{(s)}, \dots, V_{iJ}^{(s)}\} \leq 0$ 이 되도록 하여야 한다. 빠른 시간 안에 이런 조건을 충족하도록 $V_i^{(s)}, i = 1, \dots, N$ 를 추출하기 위해 적절한 truncation method를 쓰는 경우가 많다.

< Step 4 > $\theta^{(s)}$ 와 $V_i^{(s)}, i = 1, \dots, N$ 를 이용해 다음의 분포에서 $G^{(s+1)}$ 을 임의추출:

$$G^{(s+1)} \sim \text{Wishart}(v + N, M + \sum_{i=1}^N (V_i^{(s)} - R_i' \theta^{(s)})(V_i^{(s)} - R_i' \theta^{(s)})').$$

< Step 5 > $\theta^{(s)}, G^{(s+1)}, V_i^{(s)}, i = 1, \dots, N$ 를 이용하여

$$\Sigma_{\theta}^{(s)} = \left(\sum_{i=1}^n R_i' G^{(s+1)} R_i + A \right)^{-1} \text{ 과 } \widehat{\theta}^{(s)} = \Sigma_{\theta}^{(s)} \left(\sum_{i=1}^N R_i' G^{(s+1)} V_i^{(s)} + A \theta^{(s)} \right)$$

를 구하고 다음의 분포에서 $\theta^{(s+1)}$ 을 임의추출:

$$\theta^{(s+1)} \sim N(\widehat{\theta}^{(s)}, \Sigma_{\theta}^{(s)}).$$

< Step 6 > < Step 3 > - < Step 5 >의 과정을 충분히 큰 S 만큼 반복하여 $\{\theta^{(s)}, s = 1, \dots, S\}$ 와 $\{\Sigma^{(s)} (= [G^{(s)}]^{-1}), s = 1, \dots, S\}$ 를 얻은 뒤 적절한 $T (< S)$ 를 선정하여 초기 T 개의 추출을 버리고 $\{\theta^{(s)}, s = T+1, \dots, S\}$ 와 $\{\Sigma^{(s)}, s = T+1, \dots, S\}$ 만을 θ 와 Σ 의 사후분포로 간주한다. 본 연구에서는 $S = 60,000$, $T = 50,000$ 이다.

<표 9> - <표 12>은 위의 과정을 거쳐 추정된 사후분포의 평균값을 보고하고 있으며 각 모수들 사후분포의 2.5%-tile과 97.5%-tile이 0을 포함하고 있지 않은 경우 *표로 표기해 두었다.

Analysis of housing choice using multinomial probit model - Bayesian estimation -

Sang-Soo Park* · Chung-Ki Lee**

We employ a multinomial probit model to understand the housing choice of Koreans, more specifically to understand the decision factors that affect Koreans' housing decisions. The data used are from surveys conducted in July, 2006, and we use the Gibbs sampling technique in analyzing them. Apparently the housing decision is based on three dimensions—types of housing, size of places, and home ownership. A household may make its decision by simultaneously considering these three dimensions given its income and decision factors. Therefore, it would be ideal to categorize the alternatives based on these three dimensions. Due to the limited data, we treated the home ownership decision as if it were given outside the model or, in other words, as if a household made the decision before taking the type and size into account. From the results, we can see the followings: Firstly, the estimate of the coefficient of the alternative specific variables is negative and significant in all equations. Secondly, a household which decides to purchase housing is more likely to buy AS rather than HS. This is understandable considering Koreans' inclination toward apartments.

Keywords : Housing choice, multinomial probit model, Bayesian estimation

* Assistant professor, Department of Economics, Korea University

** Research professor, Institute of Economic Research, Korea University