

주관적 사전확률을 반영한 부동산 대량평가모형 개발

Incorporating Subjective Priors into Mass Appraisal Modeling

이창로 Lee Changro¹⁾, 박기호 Park Keyho²⁾

Abstract

The government-announced land price has a variety of usage from providing land information to market to serving as assessment value for property. In this study, we investigate an alternative model to enhance the accuracy of this government-announced land price. The current model for the price estimation is based on a standard comparison table of price, which is a special case of hedonic pricing models. This comparison table is constructed from the ordinary least squares result, which is one of frequentist approaches. The frequentist approach, however, can only produce an output as good as the quality of input data and if the data is not enough in amount or its quality is poor, its output necessarily deteriorates. We overcome this constraint of data by incorporating substantive knowledge of real estate experts (subjective priors) into mass appraisal modeling. The improved accuracy of our model is verified by the cross-validation simulation method. Attractiveness of a Bayesian model lies in the fact that it provides a logically transparent method for combining prior information with data, rather than rejecting the knowledge in experts' heads just because it is subjective or qualitative.

Keywords: Mass Appraisal, Expert Knowledge, Subjective Priors, Bayesian Estimation, Standard Comparison Table of Price

I. 서론

정부가 매년 발표하는 공시지가(公示地價)는 부동산 시장에 지가 정보를 제공하고, 일반 토지거래의 기준이 되게 하며, 다양한 행정 목적에 활용하기 위하여 1989년에 최초 도입되었다. 이러한 공시지가는 표준지(標準地) 공시지가와 개별 공시지가로 다시 구분되는데, 표준지 공시지가는 전국 약 3,000만 필지 중 대표성이 인정되어 선정된 필지(50만 필지, 표

본 비율 약 1.7%)를 말하며 감정평가사의 정밀평가(single-property appraisal)를 통하여 가격을 산정, 공시하게 된다. 반면 개별 공시지가는 일종의 헤도닉 가격모형(Hedonic Pricing Model)인 '토지가격 비준표(土地價格 比準表)'를 활용하여 가격을 일시에 대량 산정한다. 이때 토지가격 비준표는 50만 표준지의 특성 정보와 가격 정보에 근거하여 작성된다.

따라서 발표되는 대부분의 공시지가는 개별 공시지가에 해당되며, 이러한 개별 공시지가의 적정성은

1) 서울대학교 지리학과 박사과정(제1저자) | Ph.D. Candidate, Dept. of Geography, Seoul National Univ. | Primary Author | k25125@kab.co.kr

2) 서울대학교 지리학과 교수 및 국토문제연구소 겸무 연구원(교신저자) | Prof., Dept. of Geography, Seoul National Univ., Researcher, Institute for Korean Regional Studies | Corresponding Author | khp@snu.ac.kr

비준표 모형에 좌우된다고 할 수 있다. 비준표 모형이 부적절하게 구성된 경우 그러한 모형의 결과로 산출된 개별 공시지가 역시 부정확할 수밖에 없으며, 이는 잘못된 지가 정보의 제공, 조세부담의 불공평, 납세자 행정소송에 따른 행정력 낭비 등 수많은 부작용을 초래하게 된다.

현행 비준표 모형은 헤도닉 가격모형에 기반한 다중회귀모형으로 최소제곱법(Ordinary Least Squares: OLS)으로 추정한 회귀계수를 가격 격차율 형태로 제시하고 있다. 이러한 고전적 접근법(classical or frequentist approach)은 모형에 투입된 ‘데이터만큼 좋은 결과가 나오게 되어 있으며 따라서 데이터의 질이 미흡하면 모형을 통해 산출된 결과 역시 미흡할 수밖에 없다.

데이터(통상 ‘표본’)의 질이 미흡하거나 양이 충분하지 않은 경우 이러한 부족 부분을 연구자가 보유한 경험이나 지식, 또는 과거에 수행된 선행연구 결과로 보완하려는 방법이 바로 베이지안 접근법(Bayesian approach)이다. 베이지안 접근법은 표본으로부터 얻을 수 없는 정보를, 연구자 또는 관련 전문가가 지닌 해당 분야의 축적된 지식으로 모형에 반영함으로써 획득할 수 있다.

고전적 접근법과 베이지안 접근법의 오랜 논쟁의 핵심은 사전 확률분포(prior probability distribution)의 주관성에 있다. 기본적으로 고전적 접근법에서는 사전 확률분포라는 개념이 존재하지 않으나, 베이지안 접근법은 전문가의 지식 등을 사전 확률분포를 통해 모형에 반영하고자 한다. 따라서 사전 확률분포의 ‘주관성’에 대한 수많은 비판이 제기되었으나(Dennis, 1996), 베이지안 접근법은 이러한 주관적 지식을 아예 존재하지 않는 것으로 간주하거나 ‘정성적 판단’이라 하여 배척하기보다는 그러한 주관적 지식이 어

떻게 모형에 반영되었는지 적어도 투명하게 설명할 수 있다는 점에서(McCarthy, 2007: 225) 장점이 보다는 많은 접근법이라 할 수 있다.

이러한 베이지안 접근법은 모형 구성이 복잡하고 계산량이 방대하여 적극적으로 응용되지 못했으나 1990년대 들어 컴퓨터 성능의 발달, 관련 소프트웨어의 보급 등으로 최근에는 여러 분야에서 매우 활발하게 적용되고 있다. 초기에는 사전 확률분포의 주관성 비판을 최소화하기 위해 추정하려는 모수(parameters)에 정보를 거의 투입하지 않는 사전분포³⁾를 사용하였으나 최근에는 주관적 지식을 명시적으로 반영한 연구 사례들이 점차 늘고 있다(Crome et al. 1996; Martin et al. 2005; Gill and Walker, 2005; Levy and Crawford, 2009; Scholten, 2013 등).

본 논문에서는 베이지안 추론의 원리에 대해 간략히 설명하고, 사례 지역을 선정하여 부동산 가격추정에서 고전적 접근법과 베이지안 접근법으로 산출한 모형의 성능을 실증적으로 비교하고자 한다. 즉 전문가의 지식을 반영한 비준표 모형의 개선 가능성에 대해 살펴본다. 또한 이러한 베이지안 접근법이 부동산 가격공시제도나 감정평가 분야에 있어 유용하게 활용될 수 있는 가능성에 대해 논의하고자 한다.

II. 이론적 검토 및 선행연구 고찰

베이지안 통계(Bayesian statistics)는 베이즈 법칙(Bayes' rule)에 기반한 것으로 Y를 관찰값 벡터(vector of observed values), θ 를 추정할 모수 벡터(vector of parameters)라 하면 다음과 같이 베이즈 법칙을 표현할 수 있다.

$$p(\theta|Y) = \frac{p(Y|\theta)p(\theta)}{p(Y)} \quad \langle \text{식 1} \rangle$$

3) 모호 사전분포(vague prior) 또는 무정보적 사전분포(non-informative prior)라 함.

베이시안 통계와 기존 통계(classical or frequentist statistics)의 가장 큰 차이는 θ 의 추정방식에 있다. 기존의 통계적 접근에서 θ 는 확률변수(random variable)가 아니며 고정된 값을 갖는다(다만 그 값을 사전에 알지 못할 뿐이다). 따라서 기존의 통계적 접근에서 θ 에 대해 확률분포를 논하는 것은 의미가 없다. 반면, 베이시안 통계에서 θ 는 확률변수이며 그 분포는 θ 에 대한 불확실성을 표현한다. <식 1>에서 $p(\theta)$ 는 사전 확률분포(prior probability distribution)라 하며 데이터를 관찰하기 전 θ 에 대해 연구자가 가지는 사전 지식(또는 신념)이라 할 수 있다. $p(Y|\theta)$ 는 우도(likelihood)라 하며 최우추정법(Maximum Likelihood Method)에서의 우도와 동일한 개념이다. $p(Y)$ 는 주변확률(marginal probability)로서 원칙적으로 다음과 같은 적분을 통하여 계산할 수 있다.

$$p(Y) = \int p(Y|\theta) p(\theta) d\theta \quad \langle \text{식 2} \rangle$$

마지막으로 $p(\theta|Y)$ 는 사후 확률분포(posterior probability distribution)로서 데이터를 관찰한 후 θ 에 대해 업데이트된 연구자의 현재 지식(또는 신념)을 의미한다. 이러한 사후 확률분포의 대표값(평균)은 θ 에 대한 최적의 추정치로 해석되며, θ 의 분산은 그러한 추정치에 대한 불확실성을 나타낸다(Plant, 2012: 450).

예를 들어, Y 를 특정 지역의 주택가격, θ 를 그러한 주택가격의 평균이라 하고 전문가 A는 해당 지역의 주택가격 평균, 즉 θ 를 880만 원/m²~920만 원/m² 정도로 매우 확실하게 믿고 있다고 하자.⁴⁾ ‘매우 확실’하다는 것을 이러한 신념의 표준편차가 20 정도로 비교적 작다고 표현한다면, 전문가 A의 사전 확률은 다음과 같이 나타낼 수 있다.

$$p_A(\theta) \sim N(900, 20^2) \quad \langle \text{식 3} \rangle$$

반면, 전문가 B는 주택가격을 이보다 낮은 수준으로 생각하고 있지만 그러한 생각에 확신이 크지 않다고 하자. 즉, θ 를 720만 원/m²~880만 원/m² 정도로 생각하되, 이러한 신념의 표준편차를 80 정도로 비교적 크다고 표현한다면 전문가 B의 사전확률은 다음과 같이 표현할 수 있다.

$$p_B(\theta) \sim N(800, 80^2) \quad \langle \text{식 4} \rangle$$

$p_A(\theta)$ 및 $p_B(\theta)$ 는 전문가 A, B가 갖는 사전 확률분포를 나타낸다. 이제 거래 사례가 다수 발생하여 그러한 거래 사례들의 평균이 850만 원/m², 표준편차가 40이었다면 이러한 관찰 결과는 다음과 같은 우도 함수로 표현할 수 있다.

$$l(Y|\theta) \sim N(850, 40^2) \quad \langle \text{식 5} \rangle$$

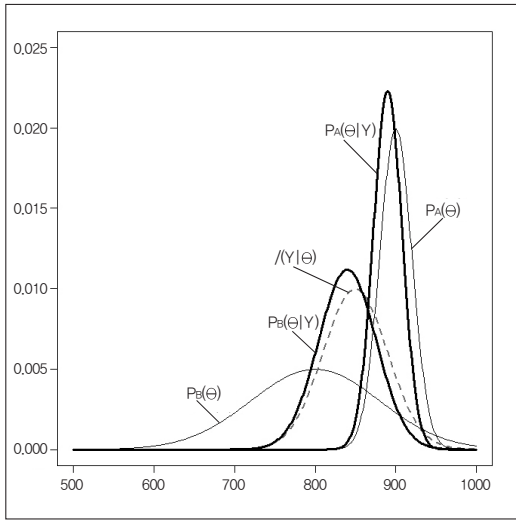
<그림 1>은 전문가 A 및 B의 사전 확률분포가 데이터 관찰 이후 어떻게 변하였는지, 즉 사후 확률분포의 형태를 보여준다. 전문가 A는 데이터 관찰 이후 자신의 예측치를 조금 하향 조정한 반면(900만 원 → 890만 원), 전문가 B는 자신의 생각에 확신이 크지 않았으므로 데이터 관찰 이후 예측치를 큰 폭으로 상향조정하였음을 알 수 있다(800만 원 → 840만 원).

이러한 베이시안 접근법은 현재 다양한 분야에서 활용되고 있으며, 특히 모호 사전분포(vague prior)가 아닌 주관적 지식을 명시적으로 반영한 연구 사례는 1990년대 이후 점차 증가하고 있는 추세다.

즉, 환경 및 생태학 분야에서 보전 가치가 높은 지역의 개발을 둘러싸고 전문가의 의견을 담아 자연

4) 이 예는 Box and Tiao(1973: 15) 및 Plant(2012: 451)가 제시한 맥락을 약간 변형한 것임.

그림 1_ 전문가 A 및 B의 주택가격에 대한 사전 확률분포, 우도 및 사후 확률분포



환경 보전에 따른 가치를 산정한 사례(Crome et al. 1996; Martin et al. 2005), 정치학 분야에서 저널리스트의 의견을 반영한 여론 분석 사례(Gill and Walker, 2005), 범죄지리학 분야에서 우범지역 파악을 위해 해당 지역 경찰관들이 생각하는 우범지역의 범위를 모형에 반영한 사례(Haining and Law, 2007), 교육·심리학 분야에서 과거 시험성적의 분포를 토대로 향후 학생들의 성적을 예측한 연구(Levy and Crawford, 2009) 등, 전문가의 생각을 담으려는 시도는 여러 분야에서 진행되고 있다.

부동산 대량평가 분야에서도 베이지안 접근법을 적용한 사례는 흔히 찾아볼 수 있으나 이러한 연구들은 대부분 전문가의 주관적 지식을 반영하기 위한 것이 아니라, 기존의 통계적 추론방법(예를 들어, 최우 추정법)으로는 모수를 추정하기 어렵거나 불가능하여 시뮬레이션 기법에 기반한 베이지안 접근법을 활용한 경우가 대부분이다. 따라서 사전 확률분포는 모호 사전분포를 부여한 경우가 대부분이다. Brunauer

et al.(2013)은 주택가격을 추정하기 위해 4개의 공간적 위계구조(house - municipality - district - county)를 갖는 다수준모형(multi-level model)을 적용하였는데, 4개의 위계구조 자체도 복잡한 모형이거니와, 종속변수와 비선형의 관계를 갖는 설명변수(시간의 추세)의 계수를 추정하기 위해 스플라인 보간(spline interpolation)을 추가하는 등 기존의 최우추정법으로는 모형을 적합시키기 어려워 베이지안 접근법을 활용하였다. Parent and Hofe(2013)는 산책로와 같은 편의시설이 주택 가치에 미치는 영향을 분석하였으며, 이때 오차항에 포함된 공간적 종속성(spatial dependence) 관련 모수(nugget, range, sill 등)의 추정이 복잡하여 베이지안 접근법을 활용하였다.

이와 같이 부동산 대량평가 분야에 있어 모형의 복잡성이 증대함에 따라 이러한 모형의 계수를 추정하기 위해 시뮬레이션 기법에 기반한 베이지안 접근법을 적용한 사례는 찾아 볼 수 있으나, 전문가의 지식을 사전 확률분포에 명시적으로 반영하여 분석한 사례는 드물다. 부동산 가격을 추정하기 위해 전문가의 지식을 반영한 베이지안 모형 적용은 본 논문이 최초의 시도라고 판단된다.

III. 사례 분석

1. 사례지역의 선정

본 연구에서는 서울시 광진구 소재 단독주택을 분석 대상으로 하였다. 광진구는 서울 동측 한강 북단에 위치한 자치구(면적 약 17km²)로서 인구수(약 37만 명), 구민 1인당 소득수준(지방세 납부액 16만 3,000 원/인, 2012년 기준)⁵⁾ 등에서 서울시 자치구 중 중간 정도에 속하는 지역이라 할 수 있다. 본 연구의 목적

5) 서울시 평균 193,000원/인(지방재정세계국, 2012).

표 1 _ 적정 실거래 신고 건의 선별

최초 건수	제외 건수				최종 건수
	금액 미기재	중복신고	지분거래*	특성 결측치**	
299	1	29	8	8	253

주: *토지대상상 면적 197.5m²의 주택 1건을 2.2m², 3.7m² 등으로 나누어 8번 신고.
 **거래 후 철거 등으로 거래시점 당시의 물적 특성(건물면적, 층수 등)을 확인할 수 없는 자료.

표 2 _ 기초 통계량(일부 주요 항목)

구분	최솟값	중위수	평균	최댓값	
거래금액(총액)	0.8억 원	7.1억 원	7.5억 원	33.0억 원	
거래단가(/m ²)	138만 원	496만 원	496만 원	1,017만 원	
토지면적(m ²)	13.2	140.5	148.7	560.7	
용도지역	1종 일반주거: 40건	2종 일반주거: 169건	3종 일반주거: 24건	준주거: 17건	개발제한: 3건

은 비준표 모형의 개선에 있으므로 비준표를 적용하여 가격을 결정, 공시하는 토지[엄밀한 의미에서 나지(裸地), vacant lot]와 단독주택 중 자료(거래 사례)가 풍부한 단독주택을 대상으로 분석하였다.⁶⁾

단독주택 거래 사례의 수집 기간은 2012년 1년으로 하였으며, 이 기간 중 신고된 실거래 건수는 총 299건이었다. 그러나 <표 1>에서 보는 바와 같이 부적정한 신고건 46건(약 15%)이 제거되어 분석에 사용된 최종 건수는 253건이었다.

<표 2>와 <그림 2>는 기본적인 선별 과정을 거친 실거래 신고 253건의 속성(attributes) 및 공간(space) 측면에서의 분포 현황을 보여준다. 토지면적 약 140m²(약 40평) 규모의 7억 원 내외 거래금액이 광진구의 전형적인 단독주택임을 알 수 있다.

가격 수준의 공간적 분포 현황을 보면 구의동 중심부 및 화양동 서측(지하철 2호선 구의역 및 건대입구역 인근)의 실거래 가격 수준이 높고, 구의동 북측

에서 중곡동으로 이어지는 지역(용마산 하단경사지대)의 실거래 가격 수준이 낮은 것을 확인할 수 있다.

2. 최소제곱법(Ordinary Least Squares: OLS) 모형을 통한 주요 설명변수의 선별

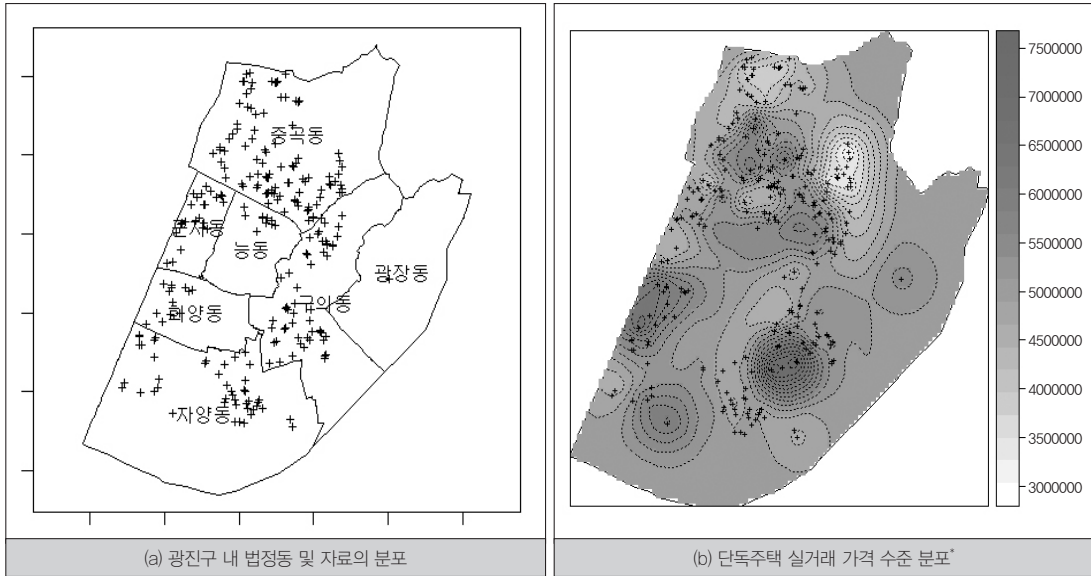
단독주택의 가격 형성에 영향을 미칠 것으로 추정되는 변수를 선정하기 위해 현행 비준표 항목에 포함된 변수들(용도지역, 도로접면, 건물구조 등)을 중심으로 살펴 보았으며, <표 3>은 반복적 적합 과정을 통해 최종 선별된 8개 설명변수(용도지역, 토지면적, 경사도, 형상, 향, 도로접면, 건물구조, 건물층수) 및 회귀계수를 보여준다.⁷⁾

용도지역의 경우 기준 범주인 1종 일반주거지역 대비 2종 일반주거 → 3종 일반주거 → 준주거지역으로 갈수록 주택가격을 높이는 방향으로, 개발제한구역의 경우에는 낮추는 방향으로 나타나 일반적인 기

6) 서울과 같은 대도시의 경우 나지만의 사례는 희소한 반면, 토지와 건물이 일체로 거래되는 단독주택의 거래 사례는 상대적으로 풍부한 편임.

7) 종속변수는 거래금액이 아닌 거래단가(토지면적당 가격)를 사용하되, 자연로그를 취하였음. 이는 현행 비준표 모형과 동일하게 종속변수 형태를 구성하기 위한 목적도 있지만, 실제로 거래단가에 자연로그를 취한 값이 자료의 정규성, 모형 적합도(R² 기준) 등에서 보다 우수하게 나타났기 때문임.

그림 2_자료의 공간적 분포



주: *정규크리깅(Ordinary Kriging)으로 추정한 가격분포 현황.

대와 일치한다. 토지면적의 경우 양(+)의 값으로 나타났는데, 종속변수가 거래총액이 아닌 거래단가이므로 (미소하나마) 규모의 증가에 따른 가격 증가(價格 增價) 현상을 나타내는 것으로 해석할 수 있다.⁸⁾ 경사도의 경우 평지보다는 완경사에 위치한 주택의 가격이 낮은 것으로 나타나 통계적 유의성은 떨어지지만 일반적인 직관과 일치한다.⁹⁾ 형상의 경우 기준 범주인 사각형 토지 대비 건물 신축 등의 효율성이 떨어지는 타 범주의 형상계수가 음(-)으로 산출되었다. 향 역시 기준 범주인 남향 대비 다른 향의 계수가 음(-)으로 나타나 합리적이라 할 수 있으며, 북향의 계수가 비교적 유의하게 산출되었다. 도로접면의 경우 기준범주인 세로가¹⁰⁾보다 도로 폭이 넓거나 유리한

세각가, 소로, 중로, 광로는 양(+)의 계수값이, 세로가보다 도로 폭이 좁은 세로불은 음(-)의 값이 산출되어 일반적인 기대와 역시 일치한다.

이상은 토지와 관련된 설명변수였으며, 건물과 관련된 설명변수를 살펴보면 다음과 같다. 우선 건물구조는 기준 범주인 벽돌조 대비 블록조는 주택의 가격을 낮추는 방향으로, 철근콘크리트조는 높이는 방향으로 산출되어 합리적인 것으로 판단된다. 건물층수 또한 1층 주택보다는 2, 3층 주택의 가격이 높은 것으로 해석되므로 일반적인 직관에 부합한다.¹¹⁾ 조정 결정계수 값은 0.43으로 산출되었다.

종속변수가 자연로그를 취한 값이므로 범주형 설명변수의 계수값을 지수화(exponentiate)할 경우 현황

8) 감정평가 실무에서는 이를 '광평수 증가(廣坪數 增價)'라 함.

9) 설명변수가 통계적으로 유의하지 않더라도 부호(±)가 통상적인 기대와 일치한다면 모형에 포함시키는 것이 바람직할 수 있음. 모형의 예측력을 현저하게 향상시키지는 못하겠지만, 그렇다고 약화시키는 것도 아님(Gelman and Hill, 2007: 69).

10) 세로불 → 세로가 → 세각가 → 소로 → 중로 → 광로일수록 접한 도로의 폭이 넓어짐을 의미함. 세로가(폭 8m 미만의 자동차 통행이 가능한 도로)의 빈도가 가장 높아(253건 중 128건, 약 51%) 해석의 편의를 위해 기준 범주로 정하였음.

11) 층수와 유사하게 건물의 규모를 나타내는 건물면적은 토지면적과 높은 다중공선성이 있는 것으로 나타나 설명변수에서 제외하였음.

비준표와 동일한 형태의 가격배율표가 산출되며, <표 4>는 <표 3>을 이러한 가격배율표로 나타낸 것이다.

<표 4>를 보면 용도지역, 경사, 건물구조, 향 등은 일반적인 직관과 부합하게 산출되었으나 도로접면

과 형상은 그렇지 않은 것으로 보인다. 먼저 도로접면의 경우 도로폭이 좁은 세로불에서 시작하여 중로까지는 일반적인 기대와 일치하나, 광로(廣路)의 경우 오히려 가격배율이 중로보다 하락하고 있다. 또한

표 3_ 최소제곱법(OLS) 적합 결과

구분	Estimate	Std. Error	t-value	p-value	비고
상수항	15.0535	0.07	202.84	0.00	
2종 일반주거지역	0.130	0.05	2.56	0.01	기준범주: 1종일반주거지역
3종 일반주거지역	0.138	0.07	2.02	0.04	
준주거지역	0.457	0.07	6.10	0.00	
개발제한구역	-0.783	0.14	-5.55	0.00	
토지면적	0.001	0.00	2.40	0.02	
완경사	-0.065	0.06	-1.12	0.26	기준범주: 평지
부정형	-0.017	0.07	-0.24	0.81	기준범주: 사각형
사다리형	-0.093	0.04	-2.47	0.01	
자루형	-0.082	0.05	-1.51	0.13	
동향	-0.037	0.06	-0.64	0.52	기준범주: 남향
북향	-0.055	0.03	-1.77	0.08	
서향	-0.021	0.06	-0.37	0.71	
광로	0.067	0.10	0.66	0.51	기준범주: 세로가
세각가	0.071	0.04	1.80	0.07	
세로불	-0.058	0.04	-1.41	0.16	
소로	0.093	0.06	1.60	0.11	
중로	0.133	0.08	1.68	0.09	
블록조	-0.121	0.14	-0.87	0.38	기준범주: 벽돌조
철근콘크리트조	0.100	0.05	1.93	0.05	
건물층수	0.070	0.03	2.78	0.01	

표 4_ 최소제곱법(OLS)에 기초한 가격배율표

용도지역		도로접면		형상	
1종 일반주거	1.00	광로	1.07	사각형	1.00
2종 일반주거	1.14	중로	1.14	부정형	0.98
3종 일반주거	1.15	소로	1.10	사다리형	0.91
준주거	1.58	세각가	1.07	자루형	0.92
개발제한	0.46	세로가	1.00	-	-
-	-	세로불	0.94	-	-
향		건물구조		경사	
남향	1.00	블록조	0.89	평지	1.00
동향	0.96	벽돌조	1.00	완경사	0.94
북향	0.95	철근콘크리트조	1.10	-	-
서향	0.98	-	-	-	-

형상의 경우에도 평가실무에서 사각형 토지 대비 자루형 토지(도로로부터의 주출입구가 매우 좁은 통로 형태로 되어 있는 토지)의 가격배율을 가장 낮게 정하는 것이 일반적이나 OLS 결과는 자루형 토지의 가격배율이 사다리형 토지보다 높은 것으로 나타났다. 이는 253건의 한정된 거래 자료만으로는 광로에 접한 토지 또는 자루형 토지의 가격 형성 차별성을 데이터가 구분해낼 수 없음을 의미한다.

3. 베이지안 모형의 적용

베이지안 추론을 실무에 비교적 용이하게 적용할 수 있게 된 것은 MCMC(Markov Chain Monte Carlo) 방법의 개발에 힘입은 바 크다. MCMC는 모수 θ 를 사전 확률분포 $p(\theta)$ 로부터 추출하고, 이를 반복적으로 수정하여 중국에는 의도한 사후 확률분포 $p(\theta|y)$ 로 근사화시키는 일반적 방법이다(Gelman et al. 2004). MCMC는 사전 확률분포 및 우도함수가 복잡한 형태를 갖게 되어 해석적 해(analytic solution)의 도출이 불가능한 경우에도 모형의 계수를 추정할 수 있는 매우 유연한 분석도구다. MCMC는 일종의 시뮬레이션 도구로 일련의 표본을 추출하여 체인(chain)을 구성하되, 시뮬레이션 횟수가 매우 커질 경우 이러한 체인은 안정적 분포(stationary distribution)로 수렴하게 된다. 일단 수렴 상태에 이르렀다면 이후 추출되는 표본은 그러한 안정적 분포, 즉 연구자가 의도한 사후 확률분포로부터 도출되었다고 해석한다.

MCMC 시뮬레이션 기법 중 가장 폭넓게 활용되는 방법은 Gibbs 샘플링(Gibbs sampling)이며 본 논문도 Gibbs 샘플링에 기반하여 분석을 수행하였다. Gibbs 샘플링은 다차원 구조(multi-dimensional structure)를 갖는 복잡한 데이터를 한 번에 해결하여

해(solution)를 찾는 대신, 이러한 데이터를 저차원 구조(low-dimensional structure)로 분해하여 각 부분에 대해 하나씩 순차적으로 해를 찾는 방법이다. 예를 들어, 관찰된 데이터를 y , 추정할 모수 벡터를 θ 라 하고, θ 가 k 개의 구성요소로 이루어졌다면(즉, $\theta=(\theta_1, \theta_2, \dots, \theta_k)$) k 개의 모수를 다음과 같은 절차를 통해 순차적으로 찾아낸다.

- ① 초기값 선택: $\theta_1^{(0)}, \theta_2^{(0)}, \dots, \theta_k^{(0)}$
- ② 첫 번째 모수 $\theta_1^{(1)}$ 을 조건부 분포 $p(\theta_1 | \theta_2^{(0)}, \theta_3^{(0)}, \dots, \theta_k^{(0)}, y)$ 로부터 추출
두 번째 모수 $\theta_2^{(1)}$ 을 조건부 분포 $p(\theta_2 | \theta_1^{(1)}, \theta_3^{(0)}, \dots, \theta_k^{(0)}, y)$ 로부터 추출
...
- k 번째 모수 $\theta_k^{(1)}$ 을 조건부 분포 $p(\theta_k | \theta_1^{(1)}, \theta_2^{(1)}, \dots, \theta_{k-1}^{(1)}, y)$ 로부터 추출

- ③ 단계 ②를 충분히 반복 수행(1,000번 등)하여 사후 확률분포 $p(\theta|y)$ 를 구한 후, 이러한 사후 확률분포로부터 표본을 추출

단계 ②를 Gibbs 샘플링의 업데이트 또는 반복 과정이라 하며, 수렴 상태에 이르게 되면(사후 확률분포의 형태가 정해지면) 해당 분포로부터 추출된 표본에는 k 개의 모수 추정치가 포함된다. 이러한 표본을 충분히 많이 추출하여 대표값(평균 또는 중위수)을 모수의 최적 추정치로 사용할 수 있게 된다.¹²⁾

본 논문의 종속변수는 주택가격(연속변수)이므로 비교적 단순한 조건부 정규분포(conditional normal distribution) 모형을 구성하였으며 다음과 같은 식으로 표현할 수 있다.

12) 보다 자세한 내용은 Gelman et al.(2004) 및 Geman and Geman(1984) 참조.

$$y \sim N(\mu, \sigma^2)$$

$$\mu = \beta_0 + \beta_1 X_1 + \dots + \beta_{p-1} X_{p-1} \quad \langle \text{식 6} \rangle$$

위 식에서 y (주택가격)는 평균 μ , 분산 σ^2 인 정규 분포에서 추출된 관찰치임을 나타낸다. 평균 μ 는 도시 용도지역, 도로접면 등 다양한 X (설명변수)와 선형결합 형태로 표현될 수 있다.

본 논문에서는 모수 θ 에 대해 아무런 정보도 가정하지 않는 모호 사전분포(vague prior) 대신 주관적 사전 확률(subjective prior)을 명시적으로 적용하였다. 이재우 외(2010)는 감정평가업무의 신뢰성을 높이기 위한 일환으로 '개별요인 비교의 객관화를 위한 연구' 보고서를 발간한 바 있다. 동 보고서에는 도로접면, 경사도, 형상 등 현행 비준표 항목의 가격배율에 대해 평가 실무자를 대상으로 실시한 설문조사 결과가 포함되어 있다.¹³⁾ <표 5>는 비준표 항목별 설문조사 결과를 보여주며, 본 논문에서는 이러한 설문조사 결과를 인용하여 사전 확률분포를 정하였다.

<표 5>의 OLS 분석 과정에서 일반적인 직관과 부합하지 않았던 도로접면과 형상을 살펴보면 다음과 같다. 먼저 도로접면의 경우 평가 실무자들은 세로가 대비 중로의 가격배율을 1.15, 이보다 도로 폭이 넓은 광로의 가격배율을 1.22 정도로 생각하고 있는 것으로 보인다. 또한 형상의 경우에도 사각형 토지 대비 자루형 토지의 가격배율을 0.84로 가장 낮게 판단하고 있다.

<표 5>의 ① 도로접면에서 세로가 대비 광로의 가격 격차율은 평균 1.22(약 22% 우세), 표준편차는 0.07인 것으로 제시되어 있다. 따라서 도로접면 중 광로 항목의 사전 확률분포는 <식 7>과 같이 지정하였고, 다른 회귀계수의 경우에도 이와 동일하게 사전 확률분포를 처리하였다.¹⁴⁾

$$P(\beta_{\text{광로}}) \sim N(1.22, 0.07^2) \quad \langle \text{식 7} \rangle$$

반면 설문조사 항목에 포함되지 않은 용도지역,

표 5_비준표 항목에 대한 설문조사 결과

① 도로접면	맹지	세로불	세로가	소로	중로	광로
평균	0.77	0.89	1.00	1.08	1.15	1.22
표준편차	0.06	0.02	0.00	0.02	0.05	0.07
② 경사도	저지		평지	완경사		고지
평균	0.90		1.00	0.96		0.90
표준편차	0.04		0.00	0.02		0.05
③ 형상	사각형		부정형	사다리형		자루형
평균	1.00		0.90	0.95		0.84
표준편차	0.00		0.05	0.05		0.07
④ 향	남향		동향	북향		서향
평균	1.00		0.97	0.91		0.95
표준편차	0.00		0.01	0.06		0.03

출처: 이재우 외, 2010.

13) 3년 이상의 경력을 가진 감정평가사 180명을 대상으로 2010.3.27~2010.4.16까지 21일간 설문을 실시하였으며, 최종 150명이 설문응답하였음(이재우 외, 2010: 159).

건물구조 그리고 도로접면 중 '세각가' 등에 대해서는 다음과 같은 모호 사전분포를 적용하였다.

$$P(\beta) \sim N(0.00, 100^2) \quad \langle \text{식 8} \rangle$$

〈표 6〉은 이러한 MCMC 시뮬레이션 결과를 보여준다. 사후 확률분포의 수렴 여부는 Rhat(Gelman-Rubin convergence statistic)과 유효 시뮬레이션 횟수

(effective number of simulation draws)를 통해 확인하였다. 먼저 Rhat은 체인 간 변동과 체인 내에서의 변동을 비교한 수치로 시뮬레이션 초기에는 그 수치가 매우 크지만 일단 수렴 상태에 도달하면 1.0에 가까워지며, 1.1 이하의 Rhat 값을 보이면 의도한 사후 확률분포에 수렴하였다고 판단할 수 있다(Gelman and Hill, 2007: 358). 유효 시뮬레이션 횟수 역시 최소 100개 이상이 되면 추론상의 큰 문제는 없는 것으

표 6_MCMC 시뮬레이션 결과

구분	평균	표준편차	2.5%	25%	50%(중위수)	75%	97.5%	Rhat	n,eff
상수항	15.048	0.065	14.920	15.000	15.050	15.090	15.180	1.001	2000
2중 일반주거	0.145	0.041	0.060	0.118	0.145	0.172	0.227	1.001	2000
3중 일반주거	0.157	0.058	0.042	0.121	0.158	0.197	0.268	1.001	2000
준주거	0.487	0.066	0.360	0.442	0.487	0.531	0.615	1.001	2000
개발제한	-0.787	0.136	-1.054	-0.878	-0.783	-0.697	-0.511	1.000	2000
토지면적	0.001	0.000	0.000	0.000	0.001	0.001	0.001	1.002	1600
완경사	-0.043	0.020	-0.083	-0.056	-0.044	-0.030	-0.005	1.004	610
부정형	-0.069	0.041	-0.153	-0.097	-0.067	-0.042	0.012	1.004	490
사다리형	-0.084	0.031	-0.146	-0.105	-0.084	-0.062	-0.024	1.002	1300
자루형	-0.112	0.043	-0.196	-0.142	-0.112	-0.083	-0.025	1.003	1400
동향	-0.031	0.010	-0.052	-0.038	-0.031	-0.024	-0.012	1.001	2000
북향	-0.062	0.025	-0.111	-0.079	-0.062	-0.044	-0.013	1.002	2000
서향	-0.045	0.027	-0.101	-0.064	-0.044	-0.027	0.007	1.000	2000
광로	0.161	0.051	0.060	0.127	0.161	0.197	0.260	1.001	2000
세각가	0.061	0.038	-0.015	0.035	0.061	0.086	0.136	1.000	2000
세로블	-0.102	0.020	-0.141	-0.115	-0.102	-0.089	-0.064	1.003	800
소로	0.078	0.018	0.043	0.066	0.077	0.090	0.112	1.001	1900
중로	0.137	0.038	0.060	0.112	0.138	0.164	0.209	1.003	2000
블록조	-0.119	0.133	-0.373	-0.209	-0.121	-0.029	0.140	1.001	2000
철근콘크리트조	0.103	0.052	0.001	0.068	0.102	0.138	0.204	1.000	2000
건물층수	0.062	0.025	0.014	0.045	0.062	0.078	0.111	1.000	2000

주: * 3 chains, 4,000 iterations with first 2,000 discarded, n.thin = 3.

14) 모형에 포함된 각 회귀계수에 대해서는 설문조사 결과상의 평균과 표준편차를 적용한 정규분포(normal distribution) 형태의 사전 확률분포를 부여하였음. 오차항 분산(σ^2)에 대해서는 균일분포(uniform distribution) 형태의 사전 확률분포를 구성하였음. 오차항 분산의 경우, 오랫동안 역감마(inverse gamma) 분포를 전형적인 사전 확률분포로 사용하였으나 최근에는 해당 모수의 표준편차 범위를 감안하여 충분히 넓은 범위의 균일분포를 적용하는 경향이 강함(Gelman, 2006). 선행연구의 경우에도 균일분포를 지정하기도 하고(Kéry, 2010: 105), 감마분포(gamma distribution)를 지정한 사례도 있음(Levy and Crawford, 2009). 본 논문에서는 오차항 분산에 대해 어떠한 사전 확률분포를 적용하든 적합 결과는 거의 동일하였음. 회귀계수와 오차항 분산에 대한 다른 형태의 확률분포 지정은 Gelman et al.(2004) 참조.

로 본다(Plant, 2012: 469). <표 6>을 보면 이러한 수렴 조건을 모두 충족하여('Rhat', 'n.eff') <표 6>의 결과에 기초한 추론에 문제가 없는 것으로 판단된다.

<그림 3>은 사후 확률분포 수렴의 예시로 용도지역 계수 중 2종 일반주거지역(기준범주: 1종 일반주거지역) 계수의 수렴 결과를 시각적으로 보여주며, 여타 계수의 경우에도 이와 유사한 패턴을 보였다(부록 참조).

<표 7>은 베이지안 모형을 통해 산출된 계수값(평균)을 가격배율표 형태로 제시한 것이다. 대부분의 항목에서 OLS 적합 결과와 유사하나 도로접면 및 형상에서는 차이가 뚜렷하다.

먼저 도로접면의 경우 OLS 결과에서는 광로

(1.07)가 중로(1.14)보다 가격배율이 낮아 일반적인 직관에 반하였으나, 베이지안 적합 결과에서는 광로 1.17 > 중로 1.15로 나타나 일반적인 기대와 일치한다. 또한 세로볼의 경우 OLS 결과에서는 0.94로 산출되어 세로가 토지 대비 6% 정도 열세한 것으로 나타났으나 베이지안 적합 결과에서는 0.90으로 산출되어 세로가 토지와 격차율이 더욱 커졌다(약 10% 열세). 이러한 결과는 평가 실무자들이 일반적으로 가지고 있는 경향과 일치하는 것으로 해석할 수 있다.

형상의 경우 OLS 결과에서는 자루형(0.92) > 사다리형(0.91)이었으나 베이지안 결과에서는 사다리형(0.92) > 자루형(0.89)으로 토지 활용의 정도가 가장 낮은 자루형이 가장 열세하게 나타나 평가 실무자들

그림 3_ MCMC 시뮬레이션 결과의 수렴 예(2종 일반주거지역 회귀계수)

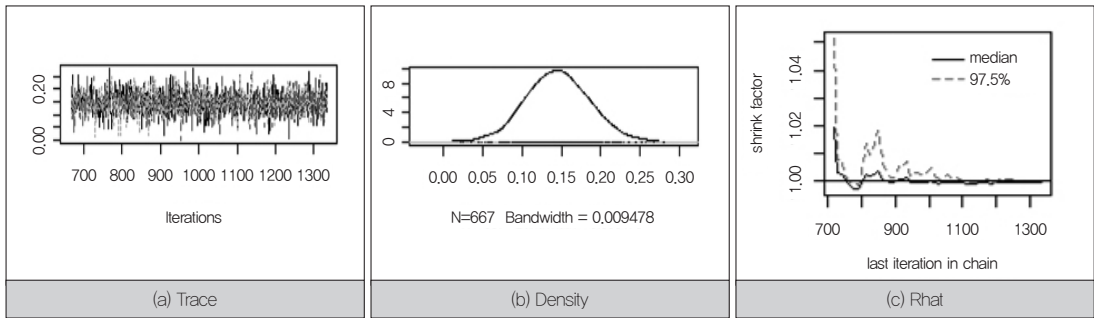


표 7_ MCMC 시뮬레이션 결과에 기초한 가격배율표

용도지역		도로접면*		형상*	
1종 일반주거	1.00	광로	1.17	사각형	1.00
2종 일반주거	1.16	중로	1.15	부정형	0.93
3종 일반주거	1.17	소로	1.08	사다리형	0.92
준주거	1.63	세각가	1.06	자루형	0.89
개발제한	0.46	세로가	1.00	-	-
-	-	세로볼	0.90	-	-
향*		건물구조		경사*	
남향	1.00	블록조	0.89	평지	1.00
동향	0.97	벽돌조	1.00	완경사	0.96
북향	0.94	철근콘크리트조	1.11	-	-
서향	0.96	-	-	-	-

주: *주관적 사전 확률분포 부여.

이 갖는 통상의 인식과 일치한다.

4. 가격 예측력 비교

부동산 대량평가 분야에서 평가모형의 성능은 주로 COD(Coefficient Of Dispersion: 분산계수)를 이용하여 측정하며,¹⁵⁾ 아래와 같은 산식을 통하여 계산한다.

$$COD = \frac{\left[\frac{\sum | \text{개별 비율} - \text{비율들의 중위수} |}{\text{비율들의 개수}} \right]}{[\text{비율들의 중위수}]} \times 100$$

〈식 9〉

〈식 9〉에서 비율은 실제 가격(actual sales price) 대비 모형을 통해 산출된 추정가격(estimated price)의 비율을 말한다. COD가 작을수록 가격 균형성이 우수한 것으로 해석할 수 있으며, 통상 20.0 이하일 경우 과세표준 산정 등 행정업무에서 받아들일 수 있는 수준으로 간주한다(International Association of Assessing Officers, 2010: 17).¹⁶⁾

253건의 전체 데이터를 모두 활용하여 모형을 적합시킨 경우 OLS 모형과 베이지안 모형의 COD 값은 유사한 수준이었다(OLS: 15.9, 베이지안: 16.4). 또는 해석하는 관점에 따라 베이지안 모형의 적합 결과가 보다 미흡하다고 볼 수 있다. 그러나 이것은 최소제곱법(OLS) 또는 최우추정법의 모형 적합 논리를 살펴보면 당연한 결과라 할 수 있다. 최우추정법과 같은 전통적 접근법의 경우 표본과 '가장 잘 일치'할 수 있는 회귀계수 값이 산출되기 때문이다(즉, 실제 관찰된 데이터를 얻을 수 있는 확률이 최대인 모수 추

정치를 제공한다).

따라서 보다 현실적인 측면에서 모형의 성능을 비교하려면 모형 적합에 사용되지 않은 미래 발생자료(future samples), 또는 모집단 전체(population)를 대상으로 모형의 가격 예측력을 비교하여야 한다. 다시 말해, 모형 적합 과정에 포함되지 않은 새로운 표본을 대상으로 가격 예측력을 검토할 경우[교차검증(cross validation)이라 한다], 전통적 접근법은 모형의 성능이 언제나 악화될 수밖에 없다(COD 증가 또는 결정계수 R²의 하락). 그러나 베이지안 접근법에서는 모형 적합 과정에 포함된 최초 표본을 대상으로 하든, 새로운 표본을 대상으로 하든 가격 예측력은 여전히 최초의 수준을 유지할 가능성이 높다.

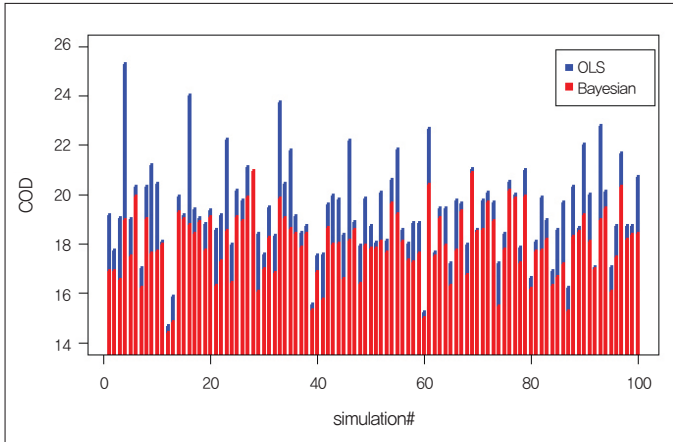
〈그림 4〉는 253건의 거래 사례 중 무작위 추출(random sampling)한 126건(50%)을 모형 적합자료(calibration dataset)로, 나머지 127건을 모형 검증자료(validation dataset)로 활용하여 COD를 측정할 결과다. 즉 무작위로 추출한 126건에 대해 OLS 및 베이지안 모형을 적합시킨 후, 나머지 127건을 대상으로 가격을 추정하고 실제 거래가격과 비교하여 COD를 계산하였다. 100번의 시뮬레이션 결과에서 단 한번만 OLS가 베이지안 모형의 COD보다 그 값이 작게 산출되었다.

두 모형 간 COD 차이가 가장 큰 경우는 총 100번 중 4번째 시뮬레이션에서 나타났으며(OLS: 25.31, 베이지안: 19.01, 차이 약 6.3) 〈그림 5(b)〉는 이때의 자료 분포를 보여준다. 〈그림 5(a)〉는 모형 적합자료와 검증자료의 일반적인 무작위 추출 현황으로 두 종류의 자료가 광진구 전체에 걸쳐 고르게 추출된 편이다. 그러나 〈그림 5(b)〉의 경우 적합자료는 광진구의

15) 공시지가 등 국내 가격공시제도에서 가격 적정성은 COD를 기준으로 검토하고 있으며 이는 해외의 경우도 마찬가지임.

16) IAAO(International Association of Assessing Officers) 기준의 경우 아파트 등 비교적 표준화된 부동산은 COD 10.0 이하, 단독주택은 15.0 이하, 상업용은 25.0 이하 등으로 상한선을 정하고 있음. 주로 미국, 캐나다 등 북미의 지역 사정에 맞는 기준으로 비교적 좁은 공간에 다수의 이질적 부동산이 분포하는 우리나라의 경우에는 상한선을 이보다 높여 적용하여야 할 것으로 판단됨.

그림 4_ OLS 및 베이지안 모형의 COD 값 비교



북쪽(중곡동 일대) 및 남쪽(자양동 일대)에서 집중 추출된 반면, 검증자료는 광진구 가운데 부분(화양동, 능동, 군자동)에 몰려 있다. 즉 4번째 시물레이션에서 화양동, 능동 및 군자동에 소재한 주택 가격은 중곡동 및 자양동에 소재한 주택 특성 및 가격에 기초하여 추정된 결과가 된다. 즉, 표본이 중곡동 및 자양동 일대에 공간적으로 편중된 경우 OLS 모형의 성능은 크게 약화되었으나(COD 25.31), 베이지안 모형은 OLS 만큼 약화되지 않았으며(COD 19.01), 이는

전문가의 주관적 사전 확률분포가 공간적 편중으로 인한 표본의 모집단 대표성 상실을 일부 보완하였기 때문인 것으로 풀이된다.

비단 공간적 분포의 편중뿐 아니라 속성 항목(용도지역, 도로접면 등)이 편중되어 있다 하더라도 베이지안 모형은 이러한 자료의 제약을 주관적 사전 확률분포로 보완할 수 있으므로 보다 강건한(robust) 모형이라 할 수 있다.

베이지안 모형의 강건성은 <그림 6>을 통해서도 확인할 수 있다. <그림 6>은 두 모형의 COD 값을 작은 순서대로 나열하여 표현한 것으로, 베이지안 모형의 COD 값이 OLS 모형의 COD 값보다 일관되게 낮는데, 특히 COD 값의 격차는 모형 성능이 전반적으로 악화될 때 더욱 크게 벌어짐을 알 수 있다. 예를 들어, 두 모형의 COD 값이 16.0 내외인 경우에는 두 모형 간 COD 값의 차이가 크지 않으나, 20.0 내외로 모형 성능이 악화될 경우 COD 격차가 커지고 있다. 이는 자료의 공간적 편중 등이 발생

그림 5_ 모형 적합자료 및 검증자료의 분포

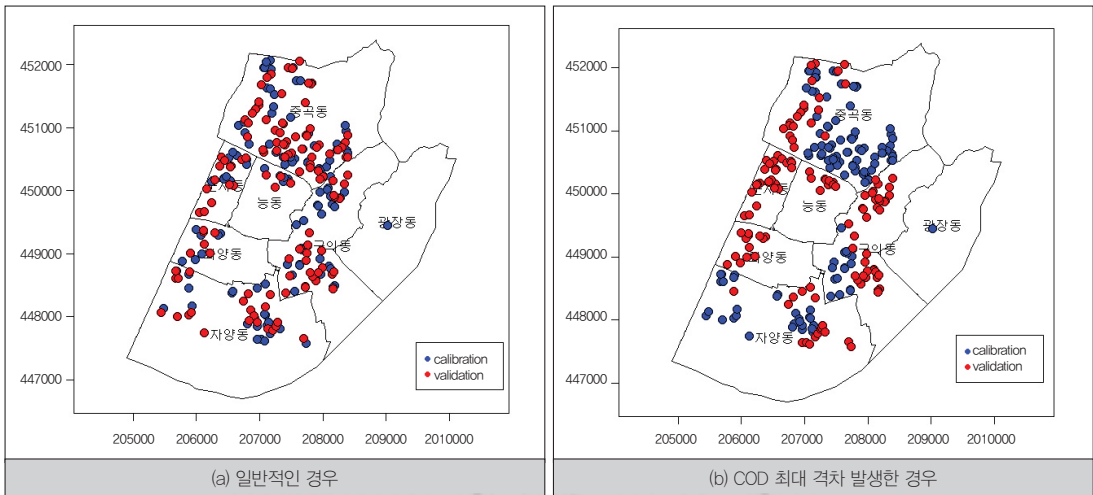
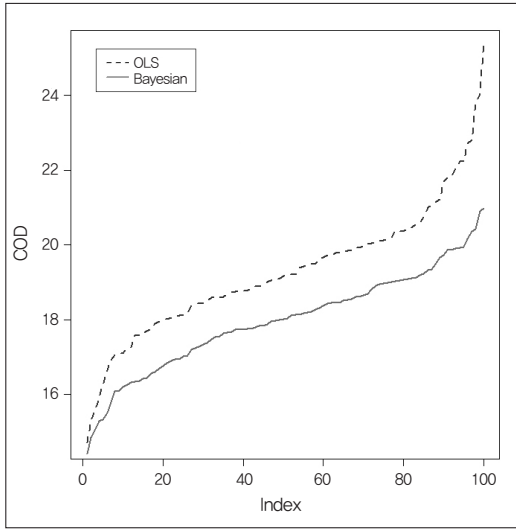


그림 6_ 베이지안 모형의 강건성



할 경우 OLS 모형의 성능은 급격히 악화되나 베이지안 모형은 그렇지 않음을 보여주는 것이다.

IV. 부동산 가격산정업무에의 활용

베이지안 접근법은 사실 기존의 통계적 접근법보다 먼저 제시(18세기)되었으나 이를 실행할 수 있는 소프트웨어 등 전산기술의 미비로 적극적으로 활용되지 못하였다. 그러나 현재는 MCMC와 같은 매우 효율적인 시뮬레이션 기법의 등장으로 실행의 문제는 거의 없다고 할 수 있다. 본 연구에서는 평가 실무자가 생각하는 가격배율을 명시적으로 포함시켜 가격 추정 모형을 구축하였고, 이를 토대로 보다 정확한 주택가격 예측이 가능함을 보였다.

이러한 베이지안 접근법은 수리적 모형을 구축하여 다수의 부동산 가격을 일시에 산정하는 대량평가(mass appraisal) 분야뿐 아니라 은행에서의 담보물 가치평가, 공익사업 수행자의 보상평가 등 정밀평가

(single-property appraisal) 분야에서 활용 잠재성이 매우 높다고 할 수 있다.

1. 대량평가

우리나라의 대표적 대량평가모형인 토지가격 비준표는 1989년 자가공시제도 도입 이후 20년 넘게 사용되어오고 있다. 이처럼 장기간 계속하여 적용하는 대량평가모형의 가장 큰 어려움은 기존 모형의 갱신(update)이라고 할 수 있다(Jensen, 1985). 모형의 갱신을 통해 변하는 시장 상황을 적시에 반영하되, 가치 추정치가 급격하게 변하지 않아야 하며 연도별로 일관성 있게 변동해야 모형 갱신이 합리적으로 이루어진 것이다. 그러나 대부분 최초의 모형 적합 결과를 이후 연도에는 사용할 수 없게 되거나, 사용하더라도 가치 추정치가 급격하게 변하는 것이 일반적이다. 이러한 문제는 시간 차원의 외삽(外挿, extrapolation in time)으로 인한 필연적 결과로서, 최초 구축한 모형을 이후 연도에 적용할 경우 대부분 모형의 성능은 악화될 수밖에 없으며, 끊임없이 변하는 부동산 시장 상황에서 그 원인을 찾을 수 있다.¹⁷⁾

베이지안 접근법은 전문가의 누적된 지식을 활용함으로써 이러한 시간 차원의 외삽문제를 일정 부분 극복할 수 있다.¹⁸⁾ 또한 과거의 누적된 정보(비준표의 경우 과거 표준지 평가정보)를 금년의 모형 적합 과정에 사전 확률분포로 활용함으로써 납세자가 받아들이기 어려운 가치 변동(예를 들어, 인접한 유사 부동산과 가격이 크게 차이나는 경우, 지가가 비교적 안정적인 지역임에도 해당 필지의 가격이 급격하게 상승 또는 하락하는 경우 등)을 줄일 수 있다.

향후 비준표를 표준지 평가 결과가 아니라 실거

17) 매수자 선호도 변화(단독주택보다 아파트 선호, 남향집보다 북향집 선호, 순수 주거용보다 주상용 선호 등), 금리 및 대출조건의 변화, 에너지 효율화 정책 및 환경 규제의 변화, 취득세율 등 부동산 세제의 변화, 용도지역, 용도지구 등 공적 규제 변화, 유흥시설의 입지 등 주변 환경의 변화, 인근지역의 개발, 쇠퇴와 같은 지역의 발전 주기 등.

래 가격에 기초하여 작성할 경우 이러한 베이지안 접근법의 유용성은 더욱 명확해진다. 실거래 자료는 정부에 의해 매년 안정적으로 관리되는 표준지(50만 필지) 자료보다 그 공간적 분포나 유형에 따른 발생의 편중이 매우 심하다. 따라서 실거래 가격에 기초한 비준표 항목의 가격 배율은 전년과 비교하여 매우 현격하게 변할 수밖에 없다. 또한 비준표 항목(설명변수) 자체가 매년 추가 또는 삭제되는 등 모형의 일관성이 급격하게 악화될 수 있다. 이러한 측면에서 실거래 가격에 기초한 비준표 작성은 매우 신중하게 접근하되, 베이지안 접근법이 제공하는 장점을 심분 활용할 필요가 있다.

2. 정밀평가

담보평가, 보상평가와 같이 한 건의 부동산을 1인 또는 2인의 전문가가 현장을 실지 답사하고 물적 특성을 확인한 후, 인근지역의 가격자료를 수집, 분석하여 최종 가격을 결정하는 경우에도 베이지안 접근법은 활용의 여지가 크다.

예를 들어, 담보평가의 경우 평가 대상 부동산과 유사한 특성을 가진 거래 사례가 인근지역에 존재하지 않아 원거리의 거래 사례를 기초로 가격을 추정하였다면 해당 평가의 결과는 그 산출 근거가 매우 미약해질 수 있다. 다시 말해 평가 대상 부동산으로부터 근거리의 위치한 거래 사례를 선택하여야 평가 과정에서 일일이 고려할 수 없는 다양한 비교요인들(지역의 향후 개발 가능성, 학군과 같은 교육의 질, 사회복지서비스의 수준 등)이 '동일'하다고 가정할 수 있게 된다. 그러나 사례가 부족하여 원거리에 위치한

거래 사례를 선택하였다면 이러한 비교요인들을 동일한 것으로 가정하기 어렵다. 이는 공간 차원의 외삽(extrapolation in space) 문제로서, 전문가가 생각하는 지역별 가격수준의 격차 정도(예를 들어 광진구 대비 강남구의 주택가격 수준은 2.5배 정도 우세 등)를 사전 확률분포의 형태로 모형에 반영함으로써 원거리 거래 사례에 기초한 평가 과정의 산출 근거를 보완할 수 있다.

이러한 활용 예는 반드시 공간적 차원에 국한되지 않으며, 평가하려는 대상 부동산이 매우 특수한 유형의 부동산(예를 들어 스키장)이어서 인근지역 범위를 전국으로 확대하여도 적절한 거래 사례를 찾기 힘든 경우 매매가 빈번하게 발생하는 다른 유형의 거래 사례(주택이나 농지)에서 획득한 정보를 사전 확률분포로 활용하여 거래 사례의 부족문제를 해결할 수 있다. 예를 들어 스키장이 간선도로로부터 멀리 떨어져 있어 접근성이 떨어지는 경우, 간선도로로부터 1km 멀어질 때마다 어느 정도의 가격 하락이 발생하는지 추정하기 어려울 수 있다. 그러나 거래가 빈번한 주택으로부터 추정한 간선도로와의 거리에 따른 가격 하락률을 사전 확률분포로 활용할 경우, 접근성 하락에 따른 스키장의 가치 감가를 계산할 수 있을 것이다.

V. 결론

본 논문은 지금까지 거의 활용하지 않은, 또는 '자의성' 개입의 여지가 커 활용을 기피하였던 전문가의 지식을 비준표 모형 적합 과정에 최초로 포함시켰다는 데 의의가 있다. 전문가의 지식을 모형 구축 과정에 반영함으로써 모형의 성능이 개선된 사실은 대량평

18) 예를 들어, 해당 연도에는 '남향' 주택이 집중적으로 거래되었고, 남향 주택이 대부분인 자료에 기초하여 구축된 금년 모형은 북향 주택의 가격 차이를 정확하게 포착하기 어려울 것임. 특히 이후 연도에 매수자 선호도가 바뀌어 '북향' 주택이 집중적으로 거래된 경우 이러한 문제는 더욱 심각해짐. 그러나 전문가가 생각하는 남향과 북향 주택의 가격 차이 정도를 사전 확률분포로 모형에 반영할 경우 이러한 문제는 어느 정도 완화될 수 있음.

가 실무에서 일반적으로 사용하는 성과측정 지표인 COD 값의 하락을 통해 확인하였다.

이와 같이 베이시안 접근법은 관찰된 자료 외에 사전 확률분포라는 정보를 추가적으로 이용할 수 있다는 탁월한 장점이 있다. 따라서 대량평가 및 정밀 평가에 있어 거래 사례, 임대 사례 등 가격자료가 특정 지역에 국한되어 발생하거나, 특정 유형에 편중된 경우, 또는 부동산 경기변동으로 특정 연도에만 집중된 경우 누적된 사전지식을 보충적으로 활용함으로써 이러한 제약을 극복하는 데 도움이 될 것으로 판단된다.

그러나 사전 확률분포를 활용한다고 하여 모든 분석 결과가 향상될 것이라 단언하기는 어렵다. 전문가의 생각을 담으려는 여러 선행연구 중 Haining and Law(2007)의 연구는 이러한 측면에서 시사하는 바가 크다. 이 연구는 우범지역 파악을 위해 해당 지역 경찰관들이 생각하는 우범지역의 범위를 모형에 반영하였으나, 범죄 발생 자료만 사용한 경우와 경찰관들의 지식을 추가로 반영한 결과 사이에 큰 차이가 없었다.

전문가는 자신이 활동하는 분야에 풍부한 지식을 보유하고 있지만 연구자가 관심을 갖는 사안[예를 들어 구축된 모형의 형태나 모수(parameter) 값] 등에 대해 직접적인 지식을 가지고 있는 것은 아니다. 따라서 전문가로부터 사전지식을 추출할 때에는 추출 절차를 신중하게 설계하여야 하고, 추출된 결과에 대해서도 세밀한 검증이 필요하다.

전문가 지식의 활용은 추출된 지식의 타당성 검증 및 이러한 사전 확률분포를 이용한 경우와 그렇지 않은 경우의 차이 분석 등이 선행되어야 그 결과에 대해 비로소 폭 넓은 지지를 받을 수 있을 것이다.

전문가가 보유한 광범위한 지식과 데이터가 가진 객관성의 결합은 부동산 평가에 있어 매력적인 대안이라 할 수 있으며, 본 논문의 결과가 우리나라 공시

가격의 정확성 제고와 감정평가업무의 신뢰성 향상에 일조하길 기대한다.

참고문헌 •••••

- 이재우 외. 2010. 개별요인 비교의 객관화를 위한 연구. 서울 : 한국감정원.
- 지방재정세계국. 2012. 지방재정통계. 서울 : 안전행정부.
- Box, G. E. P. and G. C. Tiao. 1973. *Bayesian Inference in Statistical Analysis*. Menlo Park : Addison-Wesley.
- Brunauer, W., Lang, S. and Umlauf, N. 2013. "Modelling House Prices Using Multilevel Structured Additive Regression". *Statistical Modelling* vol.13, no.2. pp95-123.
- Crome, F. H. J., Thomas, M. R. and Moore, L. A. 1996. "A Novel Bayesian Approach to Assessing Impacts of Rain Forest Logging". *Ecological Applications* vol.6. pp1104-1123.
- Dennis, B. 1996. "Discussion: Should Ecologists Become Bayesian?". *Ecological Applications* vol.6. pp1095-1103
- Gelman, A., Carlin, J. B., Stern, H. S. and Rubin, D. B. 2004. *Bayesian Data Analysis*. Boca Raton : Chapman & Hall/CRC.
- Gelman, A. 2006. "Prior Distributions for Variance Parameters in Hierarchical Models". *Bayesian Anal* vol.1. pp515-534.
- Gelman, A. and Hill, J. 2007. *Data Analysis Using Regression and Multilevel/Hierarchical Models*. Cambridge : Cambridge University Press.
- Geman, S. and Geman, S. 1984. "Stochastic Relaxation, Gibbs Distributions, and the Bayesian Restoration of Images". *Trans IEEE. Pattern Analysis and Machine Intelligence* vol.6. pp721-741
- Gill, J. and Walker, L. D. 2005. "Elicited Priors for Bayesian Model Specifications in Political Science Research". *Journal of Politics* vol.67, no.3. pp841-872.
- Haining R. and Law J. 2007. "Combining Police Perceptions with Police Records of Serious Crime Areas: A Modelling Approach". *Journal of the Royal Statistical Society Series A* vol.170. pp1019-1034.
- International Association of Assessing Officers. 2010. Standard On Ratio Study. Kansas City : IAAO.
- Jensen, D. L. 1985. The Application of Bayesian Regression for a Valuation Model in Computer-Assisted Mass Appraisal. Kansas City : IAAO.
- Kéry, M. 2010. *Introduction to WinBUGS for ecologists: A Bayesian*

- approach to regression, ANOVA, mixed models and related analyses*. Burlington : Academy Press.
- Levy, R. and Crawford, A. V. 2009. "Incorporating Substantive Knowledge into Regression Via a Bayesian Approach to Modeling". *Multiple Linear Regression Viewpoints* vol.35, no.2. pp4-9.
- Martin, T. G., Kuhnert, P. M., Mengersen, K. and Possingham, H. P. 2005. "The Power of Expert Opinion in Ecological Models Using Bayesian Methods: Impact of Grazing on Birds". *Ecological Applications* vol.15. pp266-280.
- McCarthy, M. A. 2007. *Bayesian Methods for Ecology*. Cambridge : Cambridge University Press.
- Parent, O. and Hofe, R. 2013. "Understanding the Impact of Trails on Residential Property Values in the Presence of Spatial Dependence". *The Annals of Regional Science* vol.51. pp355-375.
- Plant, Richard E. 2012. *Spatial data analysis in ecology and agriculture using R*. Boca Raton : CRC press.
- Scholten, L. 2013. "Combining Expert Knowledge and Local Data for Improved Service Life Modeling of Water Supply Networks". *Environmental Modeling & Software* vol.42. pp1-16.
-
- 논문 접수일: 2014. 3. 7
 - 심사 시작일: 2014. 3. 26
 - 심사 완료일: 2014. 4. 29

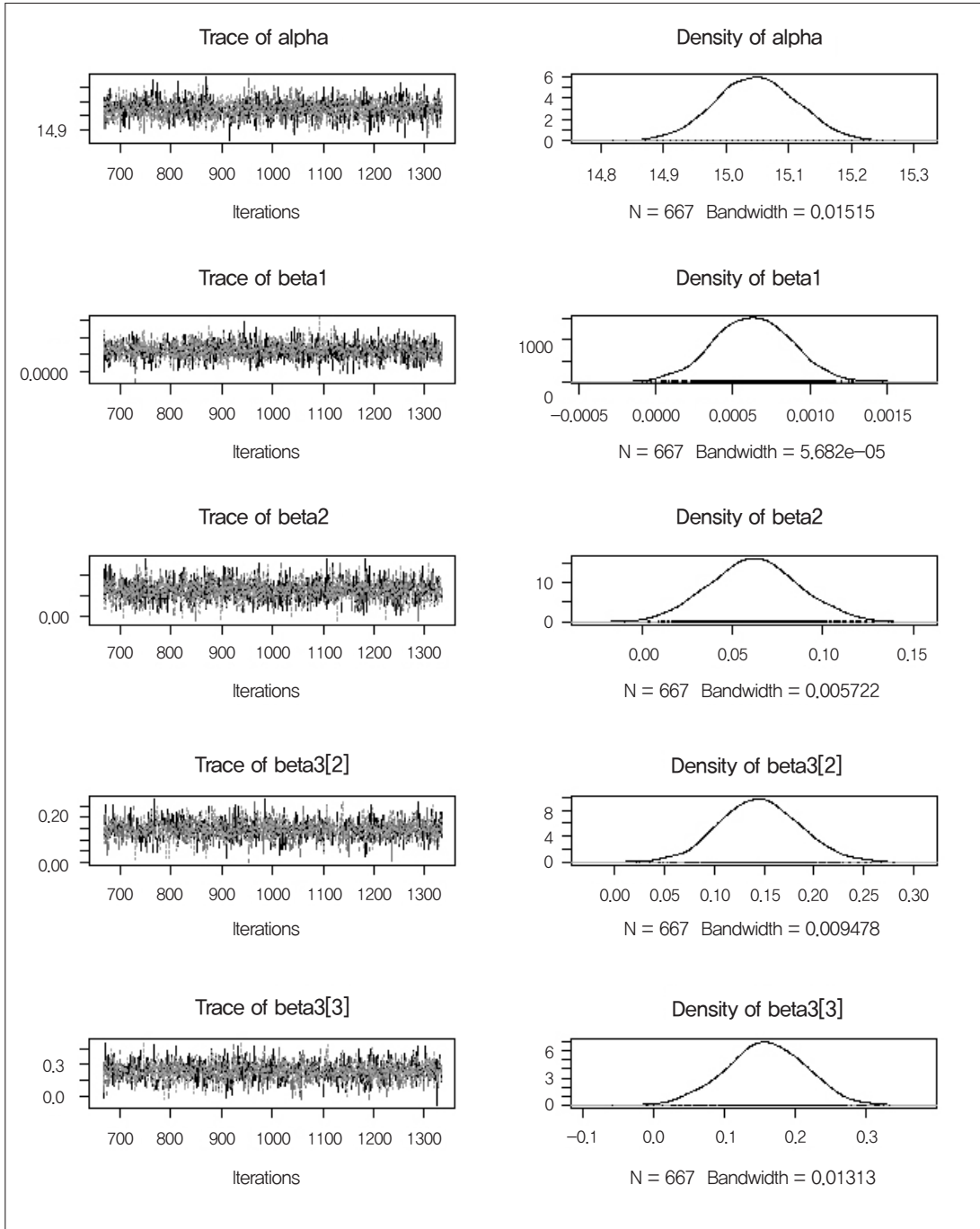
요약

주제어: 부동산 대량평가, 전문가 지식, 주관적 사전확률, 베이지안 추론, 비준표

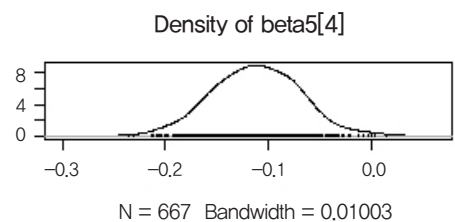
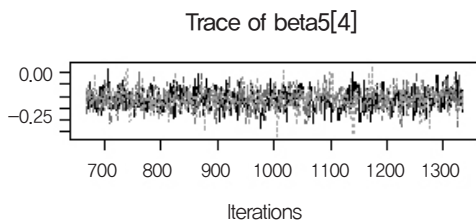
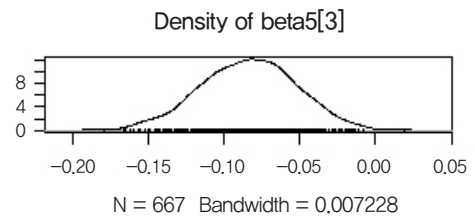
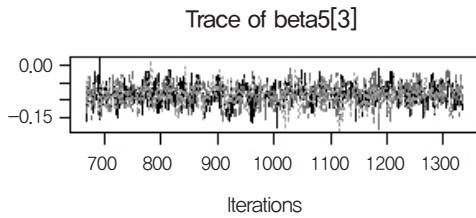
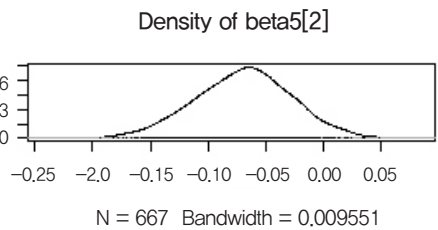
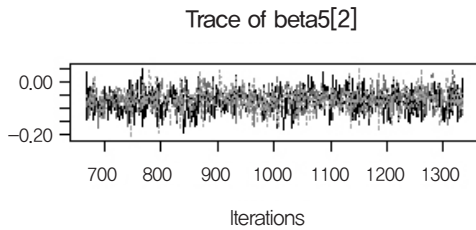
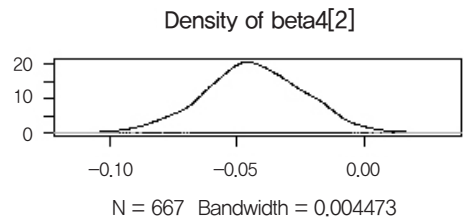
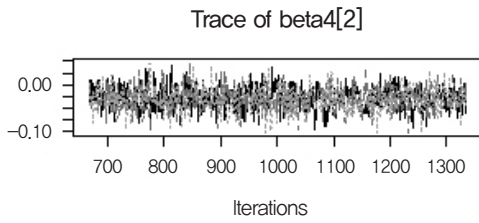
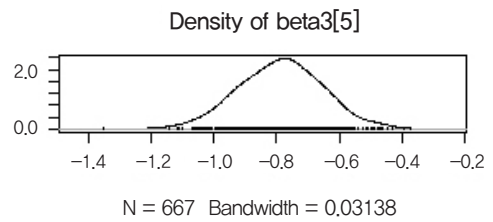
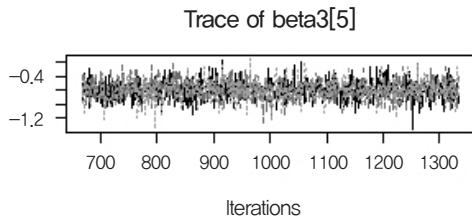
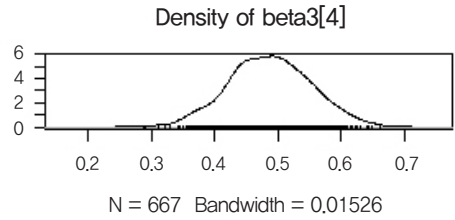
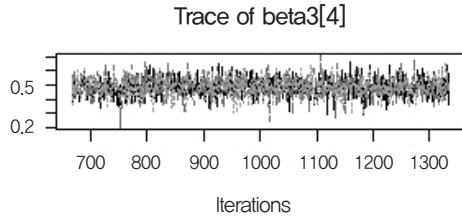
본 논문은 지가 정보의 제공, 과세표준 산정 등 다양한 행정 목적에 활용되는 공시지가(公示地價)의 정확성 제고방안에 대해 검토하였다. 현행 공시지가는 헤도닉 가격모형의 일종인 비준표(比準表)를 통해 가격을 산정하며 이러한 비준표는 고전적 통계접근법(classical or frequentist approach)인 최소제곱법(ordinary least squares)을 통해 작성된다. 그러나 고전적 통계접근법은 모형에 투입한 데이터의 질(quality) 만큼 좋은 결과를 산출할 수밖에 없으며, 데이터의 질이나 양이 충분하지 않은 경우 모형의 성능은 악화될

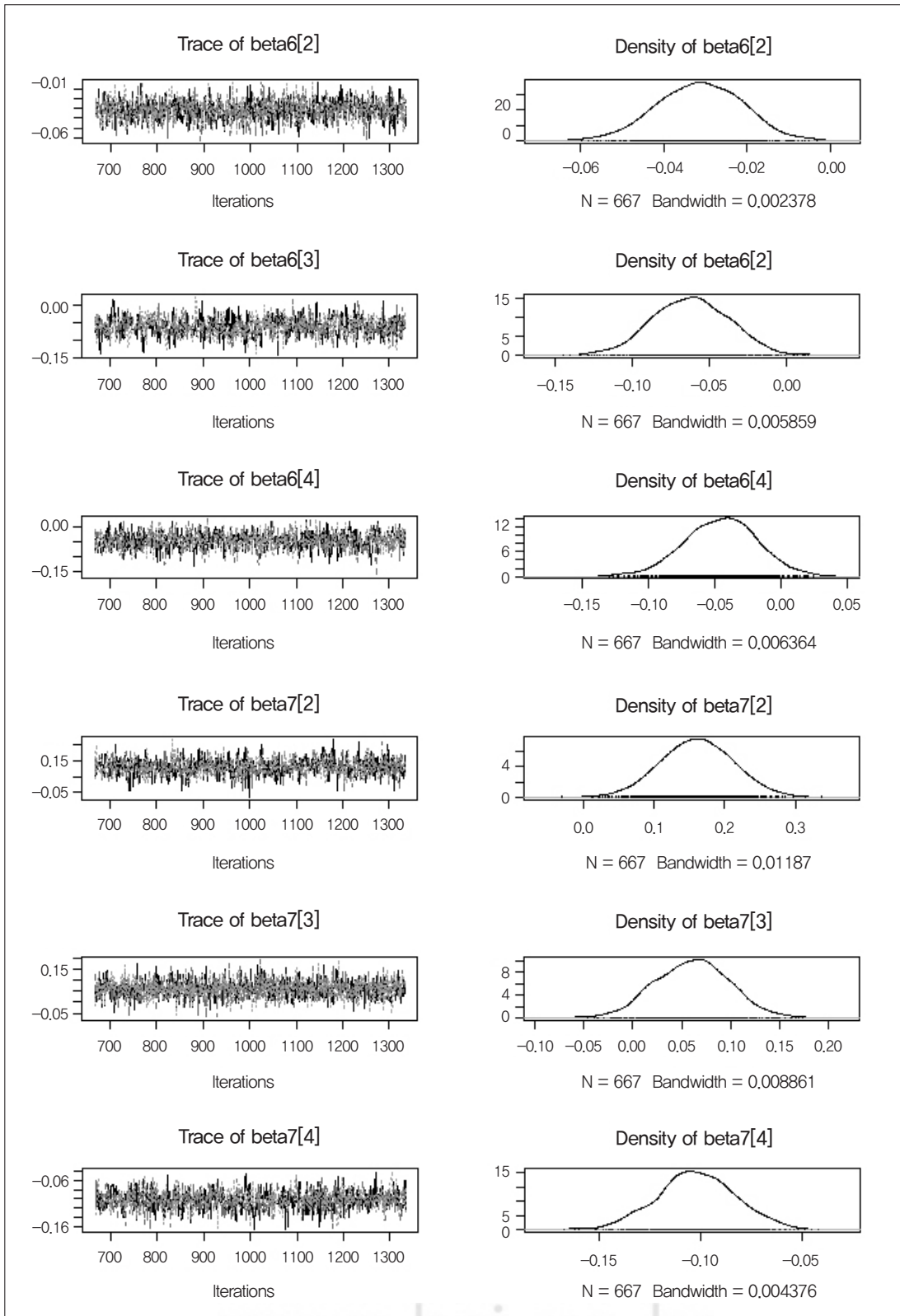
수밖에 없다. 본 논문에서는 부동산 전문가가 보유한 지식, 즉 주관적 사전확률(subjective priors)을 모형에 반영하여 이러한 데이터의 제약을 극복하였다. 모형 적합 결과는 교차검증(cross-validation)에 기반한 시뮬레이션을 통해 개선된 성능을 확인하였다. 전문가의 머리 속에 있는 지식을 '주관적'인 것 또는 '정성적' 판단이라 하여 배척하기보다는 그러한 주관적 지식이 어떻게 모형에 반영되는지 적어도 투명하게 설명할 수 있다는 점에서 베이지안 모형은 매력적인 접근법이라 할 수 있다.

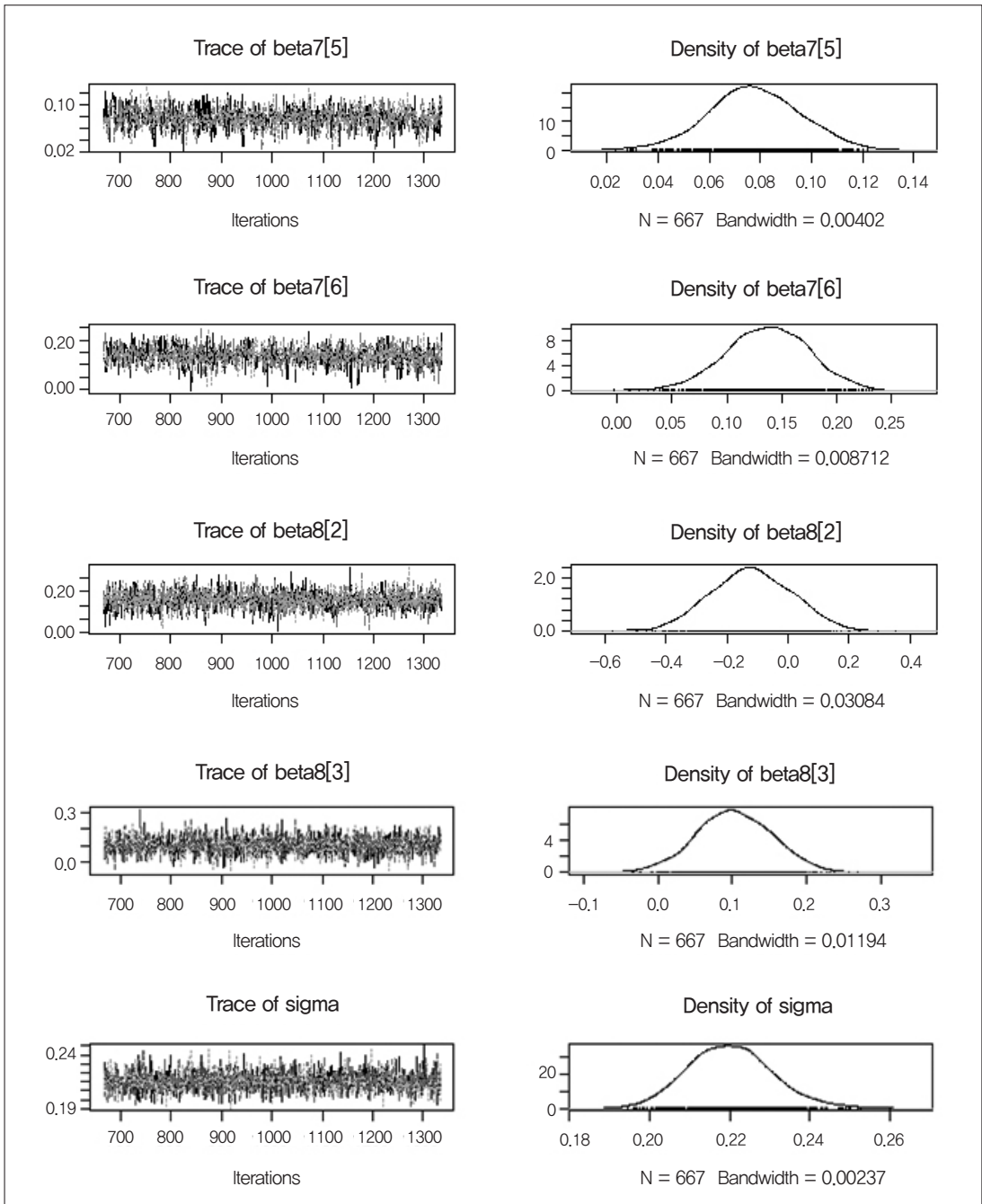
부록 1 _ Trace 및 Density



주: *alpha: 상수항, beta1: 토지면적, beta2: 건물층수, beta3: 용도지역, beta4: 경사, beta5: 형상, beta6: 향, beta7: 도로접면, beta8: 건물구조, sigma: 오차항 분산.







부록 2_ Rhat

