부동산학연구 제15집 제3호, 2009. 12, pp. 111～125

Journal of the Korea Real Estate Analysts Association
Vol.15, No.3, 2009. 12, pp. 111～125

특성가격모형을 활용한 아파트 실거래가격지수 산정방법에 관한 연구*

A Study on the Construction of the Transaction-Based Real Estate Price Index using Hedonic Price Model

박 헌 수 (Park, Heonsoo)**

< Abstract >

A real estate price index tracks the price of a standard unit of housing over time. The index typically is constructed from observations on the sale price and characteristics of a sample of house sales. This paper constructs the transaction-based price indices from January 2006 to September 2009 by using hedonic price model. Due to the requirement of the stability of the index, the paper recommends the time varying parameter model specification in hedonic price approach. By using 255,792 transaction data in Seoul, the appraisal-based KB index have appraisal smoothing problems in which they have low volatility and time-lag but the transaction-based price index do not show those problems.

주 제 어 : 실거래가격지수, 특성가격모형, 변동모수모형, 피셔지수, 연쇄지수
Keywords : transaction price, hedonic price model, time varying parameter model, Fisher index, chain index

* 이 논문은 2009학년도 중앙대학교 학술연구비 지원에 의한 것임.

** 중앙대학교 도시및지역계획학과 교수, heonsoo@cau.ac.kr

I. 서론

주택가격지수[1]는 부동산관련 정책을 결정할 때 중요한 자료로 사용되고 있다. 예를 들어 투기과열지구나 주택거래신고지역을 지정할 때와 지정된 지역을 해제할 때 주택가격지수가 사용되고 있다. 개인 투자가에 있어서도 주택가격지수는 투자에 대한 의사결정을 하는데 있어 중요한 역할을 한다. 최근 주택가격지수와 연동된 파생금융과 보험 상품의 개발이 증가하면서 주택가격지수에 대한 활용도가 높아지고 있다.

주택가격지수는 지수를 작성시 사용되는 자료유형에 따라 '평가가격[2]'에 기초한 지수(appraisal-based index)와 '실거래가격'에 기초한 지수(transaction-based index)로 나눌 수 있다(이용만·이상한, 2008). 평가가격에 기초한 지수의 예로는 국민은행에서 발표하고 있는 주택가격지수가 있다. 이는 표본주택에 대해 국민은행에 회원으로 가입된 공인중개사들로 부터 평가된 가격을 이용하여 라스페이레스(Laspeyres) 방식으로 각 표본주택에 대한 지수를 작성한 다음, 가중치를 사용하여 지역별 또는 주택규모별로 통합한 지수이다. 이용만·이상한(2008)은 평가가격에 기초한 국민은행의 주택가격지수는 평가자들의 과거의 평가가격에 일정부분 의존함으로써 평가가격의 변동성이 실제 시장가격의 변동성보다 작고, 평가가격 변화가 시장가격 변화보다 후행하는 평활화(smoothing) 문제를 지적하였다. 지수의 평활화 현상은 부동산 투자의 위험을 낮게 평가함으로써 부동산투자자의 과다한 투자를 불러올 수 있으며, 부동산 정책에 대한 시장상황을 제대로 판단할 수 없는 문제점이 있다.

일반적으로 평가가격은 부동산을 판매하고자 하는 공급자가 시장에서 받고자 하는 가격으로 실거래가격과는 차이가 있다. 실거래가격은 부동산 수요자가 제시하는 가격이 공급자가 받고자 하는 가격보다 높은 경우에 시장에서 거래되기 때문에 단순히 공급자가 받고자 하는 평가가격을 실거래가격으로 보기는 어렵다.

2005년 7월에 제정된 「공인중개사의 업무 및 부동산 거래신고에 관한 법률」에 따라 2006년 1월부터 모든 주택거래자는 실거래가격을 신고하도록 의무화되었다. 신고된 실거래 가격 자료는 인터넷 사이트(rt.mltm.go.kr)를 통해 공개하고 있다.

실거래가격을 기초로 지수를 작성하는 경우 표본을 통한 라스파이레스 방식의 실거래 주택가격지수를 작성할 수 없다. 라스파이레스 가격지수를 작성하기 위해서는 매 시점마다 표본주택의 가격정보가 필요하지만, 표본주택이 매 시점 거래가 되지 않기 때문이다. 평가가격에 기초한 라스파이레스 방식의 주택가격지수 작성방법은 매기 동일한 주택의 가격을 반영하기 때문에 주택의 특성차이에 따른 가격차이가 크지 않다. 그러나 실제로 거래된 주택들은 특성들이 모두 다르기 때문에, 주택가격의 변화가 주택이 지니고 있는 특성변화에 기인한 것이지 시장상황의 변화에 따른 것인지를 구별할 필요가 있다. 따라서 실거래 가격에 기초한 주택가격지수를 작성하기 위해서는 부동산이 지니는 특성들의 차이에 따른 가격변화를 통제할 수 있는 지수작성방법이 필요하다.

1) 본 논문에서는 아파트 실거래 가격지수를 특별한 언급이 없는 한 간단히 "주택가격지수"라 부르기로 한다.
2) 여기서 평가가격이란 감정평가사, 공인중개사 혹은 부동산 산업에 종사하는 전문가 등이 평가한 '거래가능한 가격"을 의미한다.

실거래 주택가격지수를 작성하는데 있어 또 다른 문제는 지수의 안전성(stability) 문제이다. 새로 실거래가격이 추가될 경우에 이전에 작성한 주택가격지수가 변하는 문제가 있다. 주택가격지수와 연동되어 파생금융 상품이나 보험 상품을 개발하는 경우, 새로운 실거래가격 정보가 추가될 때마다 주택가격지수가 변화한다면, 이에 연동된 파생금융상품이나 보험상품을 판매할 수가 없게 된다. 또한 주택가격지수를 이용하여 정부정책에 대한 의사결정이 이루어질 경우 이미 발표된 주택가격지수가 수시로 변한다면 정부정책은 신뢰를 얻을 수 없다.

본 논문의 연구목적은 실거래가격에 기초한 주택가격지수를 개발하는데 있다. 앞서 제기된 문제 즉, 부동산 특성의 차이에 따른 주택가격의 차이를 통제하면서, 새로운 실거래가격 정보가 추가될 때 가격지수의 안정성을 확보할 수 있는 방안으로서 특성가격모형(hedonic price model)에 의한 실거래 주택가격지수를 개발하고자 한다.

본 논문과 유사한 대표적인 선행연구로 이용만(2007)을 들 수 있다. 이용만(2007)의 논문에서는 1999년 9월부터 2003년 6월까지 강남구를 대상으로 586개 실거래자료를 사용하여 변동모수 방법에 의한 지수를 개발하였다. 본 논문과 이용만(2007)과의 차별성은 첫째, 본 논문에서는 서울시 25개 구 전체에 대해 255,792건의 자료를 이용하여 추정하였다. 25개 개별 구에 대한 주택가격지수를 산정한 후에 각 지역별 거래량을 기준으로 가중평균방법을 통해 서울시 전체, 강남지역과 강북지역 등 다양한 지역별 주택가격지수를 작성한 측면에서 다르다. 둘째, 변동모수모형을 통해 각 시기별로 특성가격모형을 추정할 때 특성가격들의 통계적 유의성을 고려하여 통계적 유의성이 없는 변수를 모형에 통제함으로써 지수작성의 오류를 줄이고자 하였다. 셋째, 각 지수의 오차를 부트스트래핑(bootstrapping) 방법을 이용하여 추정방법을 제시하였다.

본 논문의 구성은 2절에서는 주택가격지수에 대한 문헌을 고찰하고, 3절에서는 헤도닉가격모형을 이용한 실거래가격지수를 작성하는 방법들을 제시한다. 제4절에서는 분석에 사용한 자료에 대해 기술하고 제5절에서는 실거래가격지수에 대한 추정결과를 비교분석하며, 마지막 6절에서 결론을 맺는다.

II. 문헌고찰

주택가격지수의 기본적인 목적은 관심대상 혹은 표준 주택에 대해 가격변화를 하나의 대표치로 표현하는데 있다. 그러나 주택이 지니는 이질성과 낮은 거래빈도로 인해 매 기간 표준주택의 실거래 가격을 추정하는 것이 쉽지 않다.

주택 특성들은 시간이 지남에 따라서 변화한다. 예를 들어 투입자재의 가격, 소비자의 소득수준, 기호, 주택건설기술 등은 시간이 지남에 따라 변화한다. 주택가격지수를 산정하기 위해서는 이들 특성들이 시간이 지남에 따라 변화는 경우 이들이 주택가격에 미치는 영향을 감안하여 지수를 작성하여야 한다(박헌수, 2004).

주택가격지수를 작성하는 방법으로는 라스페이레스 지수, 중간값지수, 특성가격모형과 반복매매모형에 의한 지수 등이 있다.

라스페이레스 지수의 사례로는 국민은행에서 작성하고 있는 주택가격지수를 들 수 있다. 국민은행은 전국적으로 선정된 표본주택의 평가가격

에 기초하여 라스페이레스 방식을 적용하여 지수를 작성하고 있다. 이 방법은 상대적으로 계산이 간편하고 하부시장의 지수를 다양하게 산출할 수 있다는 장점이 있으나 실거래가격이 아닌 표본주택에 대한 평가가격에 기초하여 지수를 작성하기 때문에 지수의 평활화문제가 있다.

중간값(median)에 의한 가격지수작성방법은 각 시점에서 실거래된 가격의 중간값을 이용하여 지수로 작성한 것으로 지수작성이 간편하다는 장점이 있으나 중간값에 해당하는 주택의 특성들이 차이를 지수작성에서 반영하지 못하는 단점이 있다.

특성가격모형(hedonic price model)에 의한 주택가격지수 산정방법은 주택가격에 영향을 미치는 요인들을 접근성(accessibility), 주변지역의 특성(neighborhood characteristics), 주택의 구조적 특성(structural characteristics) 등 다양한 특성을 지닌 복합적인 재화라고 보고, 주택가격을 이들 개별 특성들에 대한 가치의 합이라 본다. 주택특성들이 가지는 개별적인 시장가치는 시장에서 관측될 수 없기 때문에 특성가격모형을 이용하여 특성변수들의 내재적 시장가치(implicit market price)를 파악하여 주택가격지수를 작성하는 방법이다(박헌수, 2001, 2004; 이용만 2007; 한국감정원 2007; 허세림・곽승준, 1997; Haurin and Henderschoff 1991; Meese and Wallace 1995; Wallace 1996). Meese and Wallace(1991)와 Wallace (1996)는 비모수적 방법인 국지가중회귀(locally weighted regression) 방법으로 주택가격지수를 추정한 결과 지역별로 가격지수에 있어 상당한 차이가 있는 것으로 나타나 여러 지역의 자료들을 통합(pooling)하여 도시 전체의 주택가격지수를 산정하는 방법에 신중을 기할 필요가 있다고 주장하고 있다.

반복매매모형(repeat sales model)은 특성가격모형의 변형으로 시간의 경과에 따른 주택의 특성변수들이 변화하지 않고 고정되어 있다는 가정하에 주택가격지수를 산출하는 모형이다(Bailey, Muth and Nourse 1963; Case and Shiller 1987, 1989; 이창무・김병욱・이현 2002; 이창무・김진유・이상영 2005). 이 방법은 주택특성변수들을 모형으로부터 제거함으로써 모형설정과정에서의 오류를 줄이는 점에서 장점이지만 표본을 선택하는 과정에서 반복적으로 또는 빈번하게 매매되는 주택은 불량주택일 가능성이 많으므로 주택시장의 모집단을 올바르게 대표할 수 없는 문제가 있다(Clapp and Giaccotto 1992). 또한 주택특성들의 잠재가격이 시간에 따라 변화할 수 있다는 점이 간과되고 있다(Case, Pollakowsi and Wachter 1991).

Knight, Dobrow and Sirmans(1995)는 반복매매모형과 특성가격모형의 약점을 보완한 변동모수모형 (varying parameter model)을 제시하였다. 반복매매모형과 특성가격모형에서는 주택 특성변수들의 잠재가격이 시간의 경과에 따라 일정하다고 가정하고 있는 반면, 변동모수모형은 잠재가격의 변화를 반영하고 있다. 특성가격모형은 매매시점에 대한 더미변수를 포함하고 있는 반면, 변동모수모형은 시간의 경과에 따른 특성가격의 변화를 감안하고 있다. 하지만 변동모수모형 역시 함수형태는 특성가격모형의 변형이기 때문에 모형설정 오류로 인한 모수의 추정치의 편향 가능성을 여전히 가지고 있다.

특성가격모형에 의한 실거래 주택가격지수의 적용사례는 미국 상무부의 신규 단독주택에 대한 가격지수와 MIT 부동산연구소의 TBI(Transactions-Based Index)가 있다. 반복매매가격지수 모형을

사용하는 주택가격지수로는 OFHEO(Office of Federal Housing Enterprise Oversight)의 HPI(House Price Indexes)와 Freddie Mac의 CMHPI(Conventional Mortgage Home Price Index), S&P/Case-Shiller Home Price Indicies, 이창무·김진유·이상영(2005) 등이 있다. 변동모수모형을 이용한 사례로는 Knight, Dobrow and Sirmans(1995)과 이용만(2007)이 있다[3].

III. 특성가격모형에 의한 실거래 가격지수 작성방법

특성가격모형을 이용한 실거래 주택가격지수 작성방법은 시간더미(time dumy) 방법과 시간변동모수(time varying parameter) 방법으로 나눌 수 있다.

시간더미 방법은 특성가격모형에 시간에 대한 더미변수(dummy variables)들을 추가하여 모수들을 추정한 다음 이들 추정계수를 이용하여 지수를 작성하는 방법이다. 특성가격모형에서 함수형태를 이중로그함수로 가정할 경우 함수형태는 다음과 같다.

$$\ln V_{it} = \sum_{j=1}^{k} \beta_j \ln X_{ijt} + \sum_{t=2}^{T} \alpha_t D_{it} + \epsilon_{it}, \quad i = 1, \cdots, N_t, \ t = 1, \cdots, T \tag{1}$$

여기서 V_{it}는 t시기에 i번째 거래된 주택의 가격이며, X_{it}는 k개의 주택특성벡터이며, D_{it}는 시간 추세를 나타내는 항으로 i번째 주택이 t 시점에 거래된 경우 1, 기타는 0의 값을 가지는 더미변수(dummy variable)이다. ϵ_{it}는 평균이 0이고 분산이 σ^2인 상호 독립적이고 분포가 일정한 분포를 한다고 가정한다.

시간더미 방법은 지수작성 전 기간에 대해 주택특성들이 가지는 특성가격(hedonic price)이 일정하다고 가정한다[4]. 시간에 따라 주택가격 변화는 시간더미변수에 대한 모수추정치인 α_t를 이용하여 주택가격지수를 산정한다. α_t는 기준년도($t=0$)에 비해 t년도의 주택가격의 변화율로 $\alpha_t = \ln \frac{V_t}{V_0}$ 관계가 성립된다. 따라서 주택가격지수는 다음과 같이 구할 수 있다.

$$L_t \equiv \frac{V_t}{V_0} \times 100 = \exp(\alpha_t) \times 100 \tag{2}$$

시간더미방법은 새로운 시점에 대한 실거래가격이 추가되면 회귀식을 다시 추정하기 때문에 이전에 산정한 주택가격지수를 다시 수정하여야 하는 지수의 안정성 문제가 있다.

시간변동모수방법에 의한 지수작성은 매기의 특성가격함수를 추정하여 주택의 잠재가격을 추정하고, 각 기의 추정된 잠재가격에 주택특성을 가중치로 두어 주택가격을 작성하여 기준년도와의 비를 구하여 지수를 작성한다. 각 시점에 대한 특성가격함수는 다음과 같이 가정하며,

3) 이에 대한 자세한 내용은 이용만(2007)과 한국감정원(2007) 참조.

4) 주택특성들의 특성가격을 나타내는 β_j가 모든 지수 작성기간 동안 일정하다는 제약적인 가정이다. 주택가격을 실질가격(real price)로 사용할 경우는 문제가 되지 않으나 명목가격(nominal price)로 나타낼 경우는 문제가 된다.

$$\ln V_{it} = \sum_{j=1}^{k} \beta_{jt} \ln X_{ijt} + \epsilon_{it}, \quad i = 1, \cdots, N_t, \ t = 1, \cdots, T \tag{3}$$

각 변수에 대한 설명은 식(1)과 같다.

시간변동모수 방법은 시간더미변수 방식과는 달리 매 시기마다 특성가격모형을 추정하기 때문에 주택특성들의 특성가격 β_j가 시기마다 다르다. 즉, t 시기의 i 번째 주택특성에 대한 회귀계수 β_{jt}는 t 시기의 i 번째 주택특성에 대한 잠재된(implicit) 시장가치(market value) 또는 특성가격(hedonic price)이다. 주택을 여러 특성들로 구성되어 있는 복합재라고 볼 때 β_{jt}는 개별 특성 한 단위가 가지는 잠재가치를 나타낸다.

한 시점의 주택가격은 이들 개별특성들의 잠재가치들의 합으로 나타난다. 이 경우 개별특성들의 단위가 서로 다르기 때문에 가중치를 사용하게 되는데 이때 가중치로서 사용하는 특성변수의 값을 무엇을 사용하느냐에 따라 지수가 달라질 수 있다. 보통은 특정 시기의 특성변수 평균값을 이용하여 지수를 작성하는데, 특성변수의 값은 가중치 역할을 하게 된다. 기준시점의 특성변수 평균값을 사용하는 방법을 라스파이레스 지수(Laspayres index)[5]라 하며, 비교시점의 특성변수 평균값을 사용하는 방법을 파세지수(Passe index)라 한다. 그리고 두 방법으로 작성된 지수의 증가율을 기하평균한 것을 피셔지수(Fischer index)라 한다.

예를 들어 첫 번째 시점의 특성변수 평균값을 $X_{m1} = (X_{m11}, X_{m21}, ..., X_{mk1})'$이라 하면 라스페이레스지수는 다음과 같이 구할 수 있다.

$$L_t = \frac{\sum_{j=1}^{k} \beta_{jt} \ln X_{mj1}}{\sum_{j=1}^{k} \beta_{j1} \ln X_{mj1}} \times 100 \tag{4}$$

마지막 시점의 특성변수 평균값을 $X_{mT} = (X_{m1T}, X_{m2T}, \cdots, X_{mkT})'$라 하면 파세지수는 다음과 같이 구할 수 있다.

$$P_t = \frac{\sum_{j=1}^{k} \beta_{jt} \ln X_{mjT}}{\sum_{j=1}^{k} \beta_{j1} \ln X_{mjT}} \times 100 \tag{5}$$

피셔지수는 라스페이레스지수와 파세지수의 기하평균이다. 즉,

$$F_t = \sqrt{L_t \times P_t} \tag{6}$$

연쇄지수(Chain Index)방법은 기준년도와 비교연도를 바꾸어가면서 연결한 것이다. 즉,

$$C_t = \frac{F_t^* \times F_{t-1}^*}{100} \tag{7}$$

여기서,

$$F_t^* = \sqrt{L_t^* \times P_t^*} \tag{8}$$

$$L_t^* = \frac{\sum_{j=1}^{k} \beta_{jt} \ln X_{m,j,t-1}}{\sum_{j=1}^{k} \beta_{j,t-1} \ln X_{m,j,t-1}} \times 100 \tag{9}$$

5) 여기서의 라스페이레스지수는 실거래가격에 기초한 지수로 표본에 의한 라스페이레스지수와는 다르다. 이후 라스페이레스 지수는 특별한 언급이 없는 한 실거래가격에 기초한 지수를 의미한다.

$$P_t^* = \frac{\sum_{j=1}^{k} \beta_{jt} \ln X_{mjt}}{\sum_{j=1}^{k} \beta_{j,t-1} \ln X_{mjt}} \times 100 \quad (10)$$

IV. 자료

본 논문에서 사용된 자료는 서울시 25개 자치구에 대해 2006년 1월부터 2009년 9월까지 신고된 아파트 실거래 가격이다. 실거래가격에 대한 정보는 국토해양부 아파트실거래가 웹사이트(rt.mltm.go.kr)를 통해 아파트 단지명, 전용면적, 거래된 시기, 거래금액, 층수에 대한 정보를 공개하고 있다.

특성가격모형을 이용하여 가격지수를 작성하기 위해서는 아파트에 대한 개별 특성과 아파트 단지에 대한 특성자료, 입지에 대한 특성자료들이 필요하다. 아파트에 대한 특성변수로 전용면적, 방수, 욕실수, 출입문구조, 아파트 층수 등을 사용하였으며, 아파트 단지에 대한 특성변수로는 단지 내 총세대수, 난방방식, 사용연료, 재건축여부, 건물의 경과년수 등을 사용하였다. 이들 특성자료들은 부동산테크(www.ret.or.kr)를 이용하였다.

방수와 욕실수는 전용면적(SIZE)과 상관관계가 높기 때문에 최종모형에서는 제외하였다. 특성가격모형에서 종속변수로 전용면적당 매매가격에 대해 자연대수를 취한 것을 사용하였으며 설명변수에 전용면적에 대한 자연대수값을 취한 변수를 추가적으로 사용함으로써 전용면적이 클수록 주택가격에 프리미엄(premium)이 존재하는가에 대한 여부를 파악하고자 하였다.

건물특성변수로 주택연수(DAGE)는 건축년월과 실거래된 시점과의 경과년수를 계산한 다음 자연대수값을 취하였다. 경과년수는 6개월을 기준으로 반올림하였다[6]. 건물의 노후도가 주택가격에 미치는 영향은 재건축대상 아파트 여부에 따라 다를 수 있지만 일반적으로 영향의 정도는 비선형일 가능성이 있다. 서울의 아파트들은 건물연수가 많아지면 일반적으로 가격이 하락하지만 일정 연수가 경과하고 나면 재건축 가능성이 높아지면서 다시 가격이 오르는 현상을 보인다. 이런 현상을 반영하기 위해 주택연수를 제곱한 DAGESQ 변수를 사용하였다. 두 변수의 추정계수에 대한 부호는 사전적으로 알 수는 없지만 주택연수가 많을수록 노후화에 의해 주택가격이 하락하다가 일정 기간이 지나면 재개발·재건축에 대한 기대 때문에 노후화가 진행되더라도 주택가격이 상승하게 될 경우는 DAGE의 부호는 음(-)의 값, DAGESQ의 부호는 양(+)의 값을 가질 것으로 기대된다.

재건축 여부(REBUILD)는 중요한 변수이지만 모든 실거래 자료에 대해 이들 정보를 정확하게 파악하기에는 현실적으로 불가능하다. 본 논문에서는 대안적으로 DAGE가 20년 이상인 아파트를 재건축 대상 아파트로 보았다. 이는 건축년도가 20년 이상인 아파트들은 비록 재건축조합이 설립되지 않았더라도 재건축에 대한 기대가 형성되어 있을 것으로 판단하였다.

단지내 총세대수(THHD)는 규모의 경제에 대한 외부효과를 나타낸다. 단지규모가 클수록 주택가격에 양(+)의 영향을 미칠 것으로 기대된다.

아파트의 층별로 효용가치가 다른 것을 감안하여 본 논문에서는 저층(4층이하) 아파트에 대

6) 예를들어 경과년수가 2년 6개월인 경우는 3년으로 반올림하였다.

해서는 BASE 더미변수를, 고층(15층 이상) 아파트에 대해서는 HIGH 더미변수를 사용하였다.

기타 난방방식, 사용연료, 가구당 주차대수 등은 실거래자료에서 다수의 자료들이 누락되어 있어 최종모형에서 제외하였다.

<표 1> 사용변수 설명

사용변수	변수명	단위	변수설명
전용면적	SIZE	m^2	전용면적 로그값
주택연수	DAGE	년	입주년월일에서 거래시점까지 경과년수 로그값
주택연수 제곱	DAGESQ	년	DAGE×DAGE
재건축 아파트	REBUILD	더미변수	건물경과연수가 20년 이상=1, 기타=0
총세대수	THHD	100세대	아파트 단지 내 총세대수 로그값
출입문 구조	STAIR	더미변수	계단식=1, 기타=0
저층 아파트	BASE	더미변수	아파트가 4층이하인 경우=1, 기타=0
고층 아파트	HIGH	더미변수	아파트가 15층 이상인 경우=1, 기타=0

최종모형에서 사용한 변수는 <표 1>과 같다. 2006년 1월부터 2009년 9월 까지 서울시에서 실거래된 아파트는 총 290,956건 이다. 이중에서 주택특성들이 파악된 아파트는 전체의 87.91%인 255,792건이다.

V. 실거래가격지수 산정

특성가격모형은 주택특성변수들을 설명변수로 하고 주택가격을 종속변수로 한 회귀모형이다. 종속변수는 단위면적당(m2) 가격(만원)이며 자연대수를 취한 값을 사용하였다. 설명변수들과 종속변수 간의 관계를 나타내는 함수형태에 따라 특성가격모형은 선형모형(linear model), 반로그모형(semi-log model), 이중로그모형(double log model), 초월대수모형(translog model), 박스-콕스모형(Box-Cox model) 등으로 나눌 수 있다.

이 중 어떤 모형을 사용할 것인가에 대한 일반적인 이론은 없으나, 초월대수모형(translog model)이나 박스-콕스모형(Box-Cox model)은 독립변수와 종속변수 간의 관계를 좀 더 유연하게 보여주지만, 시간변동모수방법으로 지수를 작성할 경우 추정된 모형 형태가 매 시기마다 변할 수 있어 지수를 작성하기에는 바람직하지 않다. 본 논문에서는 전용면적, 건물연수, 총세대수는 설명변수에 로그값을 취한 반면, 더미변수들은 로그값을 취하지 않는 반로그모형과 이중로그모형을 함께 사용하였다. 시간더미변수 모형에서는 2006년 2월부터 2009년 9월까지 매월 더미변수를 사용하였다.

모형추정은 통상최소자승(Ordinary Least Squares, OLS) 방법을 사용하여 추정하였다. 시간변동모수방법에서는 특성가격모형을 추정할 때 주택특성들의 변화가 없거나 아주 작은 경우 모수추정에 어려움이 생긴다. 예를들어 특정시점에 거래된 모두 아파트가 모두 계단식 아파트의 경우 특성가격을 추정할 수가 없다[7]. 본 연구에서는 사용한 변수들의 표준편차가 0.01보다 작을 경우는 해당 변수를 모형에서 제외하였다. 또한 매기마다 회귀식을 추정할 때 변수들의 통계적 유의성이 없는 경우가 발생할 수 있다. 통계적 유의성이 없는 변수들을 사용하여 가격지수를 산정할

7) 이는 STAIR의 값이 모두 1이 되어 상수항과 완전하게 다중공선성 문제가 발생하기 때문이다.

때 지수의 변동성이 커지게 되는 문제가 있다. 따라서 본 연구에서는 모든 변수들을 모형에 추가한 다음 AIC(Akaike Information Criteria) 통계량을 이용하여 유의성이 낮은 변수들을 제거하였다. 추정하여야 할 모형의 수는 시간더미 모형에서 25개(25개 구), 변동모수모형은 각 구별로 45개로 총 1,150개이다. 지면의 제약으로 추정결과를 강남구에 대해서만 제시하고 나머지에 대해서는 생략하기로 한다.

〈표 2〉 강남구 시간더미방법 추정결과
(결정계수:0.7048, 관측치수: 15,470)

변수	추정치	변수	추정치
상수항	5.671 (0.033)	200708	0.198 (0.021)
SIZE	0.222 (0.005)	200709	0.271 (0.022)
DAGE	-0.265 (0.021)	200710	0.237 (0.018)
DAGESQ	0.058 (0.005)	200711	0.247 (0.019)
THHD	0.177 (0.002)	200712	0.235 (0.019)
STAIR	0.188 (0.004)	200801	0.288 (0.020)
REBUILD	0.274 (0.008)	200802	0.259 (0.018)
BASE	0.011 (0.004)	200803	0.259 (0.016)
HIGH	0.052 (0.007)	200804	0.297 (0.017)
200602	0.033 (0.014)	200805	0.259 (0.019)
200603	0.084 (0.012)	200806	0.267 (0.022)
200604	0.115 (0.014)	200807	0.264 (0.025)
200605	0.076 (0.016)	200808	0.241 (0.028)
200606	0.070 (0.020)	200809	0.232 (0.029)
200607	0.076 (0.021)	200810	0.185 (0.037)
200608	0.049 (0.017)	200811	0.069 (0.034)
200609	0.099 (0.013)	200812	0.014 (0.022)
200610	0.186 (0.012)	200901	0.100 (0.015)
200611	0.231 (0.013)	200902	0.160 (0.015)
200612	0.235 (0.015)	200903	0.160 (0.014)
200701	0.190 (0.021)	200904	0.199 (0.013)
200702	0.207 (0.023)	200905	0.211 (0.014)
200703	0.203 (0.020)	200906	0.248 (0.013)
200704	0.175 (0.023)	200907	0.246 (0.013)
200705	0.176 (0.020)	200908	0.285 (0.013)
200706	0.224 (0.018)	200909	0.290 (0.015)
200707	0.199 (0.018)	200909	0.278 (0.015)

주: ( )안은 표준오차임

<표 2>는 강남구에 대한 시간더미변수를 사용한 특성가격모형 추정결과를 보여주고 있다. 전용면적이 큰 아파트일수록 주택가격이 높아 대형아파트일수록 가격프리미엄이 존재하는 것으로 나타났다. 주택연수의 경우 처음에는 주택연수가 많을수록 주택가격이 낮아지다 일정 연수 이상이 되면 주택가격이 높아지고 있다. 나머지 단지내 총세대수(THHD)가 클수록 주택가격이 높다. 20층이상 재건축(REBUILD) 대상 아파트는 27.4% 높은 것으로 나타났다. 계단식 아파트는 복도식보다 18.8% 가격이 높다. 5층-14층 아파트에 비해 4층 이하 저층 아파트는 1.1%, 15층 이상 고층아파트는 5.2% 가격이 높은 것으로 나타났다.

변동모수방식은 각 시기별로 특성가격모형을 추정하기 때문에 2006년 1월부터 2009년 9월까지 45개의 특성가격모형을 추정하여야 한다. <표 3>은 강남구에 대해 45개 모형에서의 회귀계수에 대한 추정치의 평균과 표준편차를 보여주고 있다. <표 2>의 특성가격 추정치와 추정치의 크기에는 다소 차이가 있지만 부호는 모두 기대한

〈표 3〉 강남구 변동모수방법 추정결과

변수	추정치
상수항	5.635 (0.284)
SIZE	0.974 (0.036)
DAGE	-0.686 (0.198)
DAGESQ	0.186 (0.050)
THHD	0.175 (0.014)
STAIR	0.206 (0.032)
REBUILD	0.298 (0.046)
BASE	0.005 (0.036)
HIGH	0.065 (0.043)

주: ( )안은 표준오차임

것과 같이 나타나고 있다.

회귀계수가 추정이 되면 이를 이용하여 시간더미방법에 의한 지수와 변동모수방법에 의한 피셔지수와 연쇄지수 방법을 사용하여 지수를 산정하였다. 지수는 전체 주택에 대해서 각 구별로 작성한 다음 거래된 주택수를 가중치로 사용하여 서울시, 강남지역과 강북지역으로 나누어서 지역별 실거래 주택가격지수를 작성하였다.

주택가격지수는 특성가격모형을 추정한 후 기준년도와 비교년도의 주택특성들을 반영하여 기준년도 대비 각 시점별로 주택가격의 변화를 비율로 작성한 것이기 때문에 특성가격모형의 추정계수의 표준오차를 가지고 지수의 표준오차를 직접 구할 수 없다. 본 논문에서는 특성가격모형을 추정한 후 부트스트래핑(bootstrapping) 방법[8]을 100회 반복하여 지수를 작성한 후 각 시점에서의 지수의 표준오차를 구하였다.

<표 4>에서 보면 각 지수에 대한 평균자승오차근(root mean square error, RMSE)은 피셔지수가 가장 낮으며, 연쇄지수가 가장 높은 것으로 나타났다.[9] 하지만 피셔지수는 새로운 실거래가격정보가 추가되면 비교년도의 주택특성들이 변화됨에 따라 기존에 작성한 지수를 변경하여야 하는 지수의 안정성 문제가 있다. 반면 연쇄지수는 새로운 실거래가 추가되더라도 기존의 가격지수가 변하지 않기 때문에 지수작성에 있어 안정성이 높다.

따라서 지수작성의 안정성 측면에서 변동모수방식의 기준년도와 비교년도를 매년 바꾸어나가면서 지수들을 연결하는 연쇄지수방법이 적절한

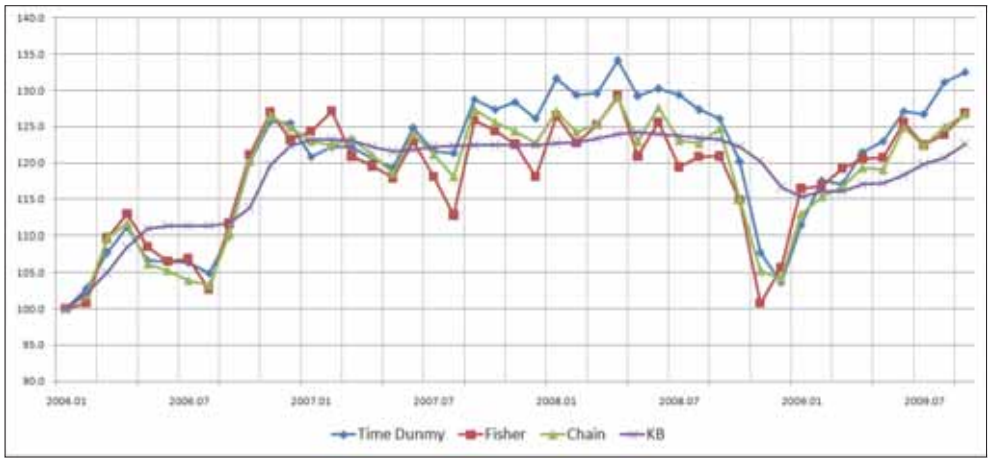
〈표 4〉 지수별 평균자승오차(RMSE) 비교

지역	시간더미	피셔	연쇄
서울시	3.13	2.90	3.62
종로구	7.74	6.15	14.12
중구	4.61	3.99	5.29
용산구	5.71	6.08	7.26
성동구	3.48	2.94	3.56
광진구	3.20	2.88	3.80
동대문구	2.74	2.22	2.52
중랑구	2.93	2.26	2.42
성북구	2.48	2.04	2.37
강북구	3.67	3.20	3.27
도봉구	2.72	2.55	3.01
노원구	1.99	1.73	1.91
은평구	4.07	3.77	3.96
서대문구	3.71	3.24	3.43
마포구	2.69	2.70	3.10
양천구	2.87	3.48	3.67
강서구	2.44	2.20	2.40
구로구	2.61	2.80	2.84
금천구	2.93	2.66	3.00
영등포구	2.39	2.10	2.51
동작구	2.42	2.33	2.90
관악구	2.71	2.52	2.80
서초구	2.21	2.32	2.77
강남구	2.26	2.39	2.98
송파구	1.95	1.86	2.35
강동구	1.83	2.10	2.38

8) 부트스트래핑 방법은 특성가격모형을 추정한 후 구한 잔차(residuals)를 임의적으로 배분하여 새로운 주택가격을 생성시킨 다음 다시 특성가격모형을 추정하여 주택가격지수를 반복적으로 구한다음 주택가격지수에 대한 표준오차를 구한다.

9) 본 논문에서 연쇄지수는 매년 비교년도를 바꿈에 따서 RMSE가 시간더미지수에 비해 1.15배 높게 나타났다.

〈그림 1〉 강남구 아파트 실거래가격지수



것으로 판단된다. 이와 같은 연쇄지수방식은 매기마다 기준년도를 바꾸어나가기 때문에 가중치가 현실과 괴리되는 문제를 최소화할 수 있는 장점이 있다. 연쇄지수방식이 이런 장점이 있음에도 불구하고 각종 물가지수 작성에서 이 방식을 사용하지 않는 이유는 매기 표본을 새로 설정해야 하는데 따른 비용과 시간이 많이 들기 때문이다. 그러나 실거래가격에 기초한 주택가격지수의 경우, 매기 거래가 일어난 표본주택이 바뀌기 때문에 연쇄지수방식을 사용해도 별도의 비용이나 시간이 들어가지 않는다.

<그림 1>은 강남구에 대해 시간더미변수모형에 의한 지수(그림에서 Time Dummy로 범례)와 변동모수모형에 의한 피셔지수(Fisher)와 연쇄지수(Chain), 실거래가격지수와 비교를 위해 평가가격에 기초한 국민은행 가격지수(KB)를 함께 보여주고 있다. 전반적으로 실거래가격에 기초한 가격지수는 지수의 변동성이 크게 나타나는 반면 국민은행 지수는 평활화되어 주택가격 변동률이 낮게 나타나고 있다. 특히, 국민은행의 지수의 평활화 문제는 경기가 하락할 때 실거래가격지수와 큰 차이를 보인다. 예를 들어 2006년 4월 이후 2006년 8월까지, 2008년 9월부터 12월까지 실거래가격은 크게 하락하지만 평가가격에 기초한 국민은행지수는 가격지수의 변화가 거의 없거나 완만하게 감소하는 것으로 나타나고 있어 경기변동을 잘 반영하지 못하는 문제가 있다. 이에 대해서 이용만・이상한(2008)은 평가자들의 평가가격에 과거의 가격에 일정부분 의존하는 지수의 평활화 문제로 보았다. 이러한 지수의 평활화가 가지는 문제는 부동산 투자에 따른 위험을 과소추정함으로써 부동산에 대한 포트폴리오 구성에 왜곡을 가져오는 문제가 있다.[10]

<그림 2>는 연쇄지수를 이용하여 강남구, 서울시, 강남지역과 강북지역에 대한 가격지수를 보여주고 있다. 4개 지역에서 강남구의 실거래가

10) 반대로 투자자의 입장에서 실거래가격에 기초한 가격지수가 수익률을 과다추정함으로써 아파트에 대한 투자를 조장하는 문제도 지적될 수 있다. 실제 국민은행지수의 경우 2008년 12월대비 2009년 9월 강남구의 주택가격지수는 5.1% 상승한 반면 실거래가격지수는 28.9% 상승한 것으로 나타났다.

〈그림 2〉 연쇄지수를 이용한 지역별 실거래가격지수 비교

격지수가 가장 낮게 나타나고 있다. 강남구는 2007년 11월까지는 지수가 가장 높게 형성되었지만 글로벌 경제위기를 겪으면서 실거래가격은 거의 변화가 없거나 하락하는 추세를 보이고 있다. 2009년 12월을 저점으로 최근에는 증가하는 추세를 보이고 있다. 강남지역과 강북지역을 비교할 때 강남지역의 실거래 가격지수는 강북지역에 비해 상승률이 빠르게 증가하면서 두 지역의 지수의 차이는 2008년 6월에 가장 크게 나타났다. 이후 두 지역의 가격지수 상승은 비슷한 패턴으로 증가하면서 두 지역의 지수차이는 일정하게 지속적으로 유지되면서 움직이고 있는 것으로 파악된다[11].

VI. 결론

실거래가격에 기초한 새로운 주택가격지수는 시장상황에 대한 정확한 정보를 제공함으로써 올바른 주택정책을 이끌어낼 수 있는 기초가 된다. 더 나아가 실거래가격에 기초한 새로운 주택가격지수를 이용하여 다양한 파생적인 투자정보를 제공할 수 있다. 주택의 질적 차이에 따른 가격차이를 지수화할 수 있으며, 주택가격지수를 이용하여 투자수익률 지표를 개발할 수 있다.

실거래가격에 기초한 주택가격지수를 작성함에 있어 주택가격지수는 최소한 주택의 특성차이에 따른 가격 차이를 걸러내야 하고, 지수의 안정성을 확보하고 있어야 한다. 실거래가격을 이용할 경우, 매기 표본주택이 실제 거래된다는 보장이 없으므로 표본 방식에 의한 라스파이레스 지수는 작성할 수가 없다. 따라서 실거래가격에 기초한 주택가격지수를 작성하기 위해서는 반복매매가격지수모형 또는 특성가격지수모형을 사용할 수밖에 없다.

본 논문에서는 특성가격지수모형에 대해서 살

11) 강남과 강북지역은 2008년 6월 이후부터는 실거래 가격지수 간에는 공통추세(common trend)를 가지는 것으로 보여진다. 다시말해 두 지역의 가격지수는 서로 공적분(cointegrated)되어 균형상태를 유지하는 것으로 보인다. 공적분검정(cointegration test)에 대한 통계적 검정은 추후 과제로 남긴다.

펴보았다. 특성가격모형에 의한 지수작성방법으로 시간더미방법에 의한 지수와 변동모수모형에 의한 피셔지수와 연쇄지수를 사용하였다. 지수의 표준오차는 연쇄지수가 시간더미지수와 피셔지수보다 높으나 시간더미 방법을 이용한 지수와 피셔지수는 새로운 실거래자료가 추가될 경우 기존의 지수들을 변경해야 하는 단점이 있는 반면 연쇄지수는 새로운 실거래가격정보가 추가되더라도 기존의 가격지수가 변하지 않기 때문에 지수작성에 있어 안정성이 높다.

따라서 특성가격지수를 이용한 아파트 실거래가격지수는 변동모수모형을 이용하여 기준년도와 비교년도를 매년 바꾸어나가면서 지수들을 연결하는 연쇄지수방법을 사용하는 것이 적절한 것으로 평가되었다.

특성가격함수 모형에 의한 가격지수를 작성할 때 해결해야 할 문제들이 여전히 남아 있다. 첫째, 주택특성함수의 모형설정에 대한 문제이다. 본 연구에서는 일부 정량적인 설명변수는 이중로그 형태로 모형을 규정한 반면 일부 더미변수와 일부 정량적인 변수는 반로그 형태로 모형을 규정하였으나 이에 대한 논리적인 근거가 미약하다. 일부 변수에 대해 로그값을 취하는 것은 변수의 이상치(outlier)에 의한 영향을 지수작성에서 가능한 줄여보자는 의도로 해석할 수 있다. 함수형태에 대해서 보다 일반적인 함수형태를 고려할 수 있는 비모수적(nonparametric) 방법에 대해 추후 연구가 필요하다고 판단된다.

둘째, 실거래자료의 빈도가 낮은 경우 특성가격모형에 의한 지수작성은 매우 불안정한 모습을 보이고 있다. 따라서 거래빈도가 낮은 시장에서의 안정적인 지수작성을 위한 추가적인 연구가 필요하다[12].

셋째, 특성가격모형을 추정할 때 오차항은 모두 독립적이고 분포가 일정한 분포를 한다고 가정하였는데, 많은 경우 분포가 일정하지 않는 이분산을 가지는 분포를 하는 경우가 많다. 또한 실거래 자료는 부동산을 구입하고자 하는 사람이 제시하는 가격이 부동산을 판매하고자 하는 사람의 희망가격보다 높은 경우에만 거래가 이루어지기 때문에 모든 자료가 독립적인 자료라고 보기 어렵다. MIT에서 개발한 TBI 지수와 같이 표본선택의 편기(sample selection bias) 문제를 해결하는 노력이 필요하다.

논문접수일 : 2009년 11월 30일
심사완료일 : 2009년 12월 23일

12) 이에 대해서는 박헌수(2007) 참조.

참고문헌

1. 박헌수, “거래빈도가 낮은 시장에서의 실거래 부동산 가격지수 작성에 관한 연구”, 「부동산학연구」 제13권 제3호, 2007, pp. 187-200.
2. 박헌수, “시공간자기회귀모형을 이용한 서울 아파트가격지수 추정에 관한 연구”, 「국토연구」 제42권, 2004, pp. 125-140.
3. 박헌수, “준모수방법에 의한 주택가격지수 추정에 관한 연구”, 「부동산학연구」 제7권 제1호, 2001a, pp. 1-16.
4. 이용만, “특성가격함수를 이용한 주택가격지수 개발에 관한 연구-시간변동계수모형에 의한 연쇄지수-”, 「부동산학연구」 제 13권 제 1호, 2007, pp. 103-125
5. 이용만・이상한, “국민은행 주택가격지수의 평활화 현상에 관한 연구”,「주택연구」제 16권 제 4호, 2008, pp.27-47
6. 이창무・김병욱・이현, “반복매매모형을 활용한 아파트 매매가격지수”, 「부동산학연구」 제8권 제2호, 2002, pp. 1-19
7. 이창무・김진유・이상영, “공동주택 실거래가지수 산정에 관한 연구”, 「국토계획」제40권 제4호, 2005, pp. 121-134
8. 한국감정원, 「부동산실거래가격에 기초한 주택가격지수 개발」, 2007.
9. 허세림・곽승준, “한국주택시장에서의 주택가격지수 산출방법에 관한 연구”, 「주택연구」 제5권 제1호, 1997, pp. 1-18
10. Bailey, M.J., R.F. Muth and H.O. Nourse, “A Regression Method for Real Estate Price Index Construction,” Journal of the American Statistical Association Vol. 58, 1963, pp. 933-942.
11. Case, B., H.O. Pollakowski and S.M. Wachter, “On Choosing Among House Price Index Methodologies,” Journal of the American Real Estate and Urban Economics Association Vol. 19, 1991, pp. 286-307.
12. Case, K.E. and R.J. Shiller, “Prices of Single-Family Homes Since 1970: New Indexes for Four Cities,” New England Economic Review, 1987, pp. 45-56.
13. Case, K.E. and R.J. Shiller, “The Efficiency the Market for Single Fimily Homes,” American Economic Review Vol. 79, 1989, pp. 125-137.
14. Clapp, J.M. and C. Giaccotto, “Estimating Price Indexes for Residential Property: A Comparison of Repeat Sales and Assessed Value Methods,” Journal of the American Statistical Association Vol. 87, 1992, pp. 300-306.
15. Haurin, D.R. and P.H. Hendershott, “House Price Indexes: Issues and Results”, AREUEA Journal Vol. 19, 1991, pp. 259-269
16. Knight, J.R., J. Dombrow, and C.F. Sirmans, “A Varing Parameters Approach to Constructing House Price Indexes”, Real Estate Economics Vol. 23, 1995, pp. 187-205.
17. Meese, R., and N. Wallace, “Nonparametric Estimation of Dynamic Hedonic Price Models and the Construction of Residential Housin Indices.g Price”, AREUEA Journal Vol. 19, 1991, pp 308-332.
18. Wallace, N., “Hedonic-Based Price Indexes

for Housing: Theory, Estimation, and Index Construction" Federal Reserve Bank of San Francisco Economic Review Vol. 1, 1996, pp. 34-48.

19. http://rt.mltm.go.kr
20. http://www.ret.co.kr