

거래빈도가 낮은 시장에서의  
실거래 부동산 가격지수 작성에 관한 연구\*  
- 강남구를 대상으로 -

A Study on the Construction of a Transaction-Based Real Estate Price Index  
for Thin Markets in Gangnam-Gu, Seoul

박 헌 수 (Park, Heon Soo)\*\*

< Abstract >

In this article, I examine both the hedonic and time-series approaches for constructing real estate price indexes for markets that have few transactions. The thinness of the market does have a marked effect on the precision of the index estimate. The index from the hedonic model by using time dummy variables loses precision rapidly when the number of transactions drops. The imprecision is manifest in a much more volatile index. At the ten transactions per period level of market thinness, the imprecision is manifest. At two transactions per period, the index becomes wholly unreliable. On the other hand, the time-series price index shows much better in thin markets. It is more parsimonious than the hedonic price index and also potentially more accurate and less prone to outliers. It remains close to the true index when there are small transactions per period, and it does not have the volatility of the hedonic index. When there are two transactions per period, the index itself is stable, although it is prone to sample selection bias. Nevertheless, the index retains reasonable precision, and the estimate of the fundamental volatility of real estate prices is good. It achieves this by linking current transactions to preceding transactions, thereby increasing the set of comparable transactions on which to base the index.

주 제 어 : 실거래가격지수, 상태공간모형, 칼만필터링

Keyword : Transaction-based Price Index, State-Space Model, Kalman-Filtering

\* 본 연구는 한국부동산분석학회 2007년도 춘계학술대회에서 발표한 논문을 근거로 하였음.

\*\* 중앙대학교 도시 및 지역계획학과 부교수, heonsoo@cau.ac.kr

## I. 서론

주택가격지수는 주택관련 정책을 결정할 때 중요한 판단기준이 되고 있다. 예를 들어 투기과열지구나 주택거래신고지역을 지정할 때, 또는 지정된 지역을 해제할 때 주택가격지수가 사용되고 있다. 그러나 주택가격지수가 주택시장의 가격동향을 정확하게 반영하지 못하고 있다. 대부분의 주택가격지수는 실제 거래가격이 아니라 호가(판매자가 희망하는 가격)에 기초하여 지수가 작성되고 있다. 실제 거래가격은 주택구입자가 제시하는 가격이 판매자가 받고자 하는 가격보다 높은 경우에만 시장에서 거래되기 때문에 호가를 실제 가격이라고 보고 지수를 작성하는 것은 문제가 있다(한국감정원, 2007).

우리나라는 2006년 1월부터 거래된 모든 주택은 실거래가격을 의무적으로 신고하도록 되어 있다. 이러한 실거래가격을 기초로 지수를 작성하는 경우, 기존의 표본에 의한 라스파이레스(Laspeyres)식 지수작성방법을 사용할 수 없다. 라스파이레스 방식의 지수를 작성하기 위해서는 매 시점마다 표본주택의 가격정보가 필요하나, 표본주택이 매 시점 거래되는 것은 아니기 때문이다. 따라서 실거래가격을 기초로 주택가격지수를 작성하기 위해서는 새로운 접근이 필요하다(이용만, 2007).

우리나라는 현재 국민은행과 일부 민간 부동산정보업체에서 주택가격지수를 발표하고 있다. 이들 기관들은 표본주택을 선정한 후, 표본주택의 가격정보를 등록된 중개사로부터 주기적으로 제공받아 라스파이레스 방법으로 지수를 산출하고 있다. 이와 같은 호가 기준의 라스파이레스식 지수작성방법은 매기 동일한 주택의 가격을 추적

하기 때문에 주택의 특성차이에 따른 가격차이 가능성이 크지 않다.

그러나 실제로 거래된 주택들은 특성들이 모두 다르기 때문에 가격변화가 특성변화에 기인한 것인지 아니면 시장상황에 따른 것인지를 구별할 필요가 있다. 따라서 실거래가격에 기초한 주택가격지수를 작성하기 위해서는 주택의 특성차이에 따른 가격 차이를 통제할 수 있는 지수작성방법이 필요하다.

그동안 국내·외적으로 주택가격지수를 작성 문제를 해결하기 위한 노력들이 많이 이루어져 왔다. 이들 연구들은 크게 두 가지로 나눌 수 있다. 첫 번째는 시간이 경과하더라도 주택의 구조적 특성들이 변화하지 않는다는 가정 하에서 주택가격을 시차(lag)에 대하여 회귀시켜 가격지수를 산출하는 반복매매모형(repeat sales model)으로 Bailey, Muth 및 Nourse(1963)와 Palmquist(1982) 연구가 대표적이다. 두 번째는 주택가격을 주택특성들에 회귀시켜 가격지수를 산출하는 특성가격모형(hedonic price model)이다. 특성가격모형은 Lancaster(1966)와 Rosen(1974)의 연구에서 그 이론적 틀이 마련되었다. 주택가격지수 작성에 대한 연구는 이 두 모형 간의 비교와 각 모형을 정교화시킴으로서 발전해 왔다. 그러나 대부분의 연구는 실거래 자료가 방대한 지역별로 집계한 자료를 이용하여 부동산가격지수를 작성하는 연구들이 대부분이다(Bailey, Muth and Nourse, 1963; Case and Shiller, 1989; Case and Quigley, 1991; Case, Pollakowski and Wachter, 1991; Meese and Wallace, 1991b; Clapp and Giaccotto, 1992; Goetzmann, 1992).

그동안 하부지역 시장에서 가격지수에 대한 관심은 상대적으로 적었다(Gatzlaff and Ling, 1994;

Can and Megbolugbe, 1997; Goetzmann and Spiegel, 1997). 그 이유는 하부지역으로 세분화하여 가격지수를 작성하는 경우에 실거래 자료가 적어 지수작성이 불가능하기 때문이다.

본 연구에서는 집계자료가 아닌 하부 지역에 대한 개별자료에 의한 가격지수를 작성하는 방법에 대해서 다룬다. 특히, 거래빈도가 낮은 시장에서 부동산 가격지수를 작성하는 방법에 대해서 다룬다.

지역 주택시장은 전형적으로 거래빈도가 낮은 시장이다. 이러한 시장에서 가격의 움직임들을 정확하게 측정하는 것은 중요하다. 특히, 거래빈도가 낮은 지역 시장의 가격지수에 대한 관심은 시장분석가에게 높을 수 있다. 예를 들어, 주거용 부동산 개발업자들에게 있어 특정 지역에 대한 신규주택에 가격지수에 대한 정보가 중요하다.

거래빈도가 낮은 지역에 대해 가격지수를 추정하는데 있어 중요한 문제는 자유도(degree of freedom)이다. 실거래 자료에 기초한 전통적인 헤도닉 가격지수는 지수작성 기간 동안 주택특성에 대한 특성가격이 모두 일정하다는 제약적인 가정하에서 기간에 해당하는 시간더미 변수들과 설명 변수들로 헤도닉 가격모형을 구성한 뒤 시간더미 변수에 대한 추정계수를 이용하여 지수를 작성한다. 시간더미변수에 대한 추정모수는 해당 시점에 대해 주택가격의 평균적인 상승 혹은 하락의 정도를 나타낸다. 만일 특정시기에 있어 거래건수가 작을 경우, 특정 주택가격의 영향이 지수작성과정에서 크게 작용하게 되어 지수가 불안정하게 된다. 이러한 자유도 문제를 해결하기 위해서는 모형에서 추정하고자 하는 설명변수를 줄이는 것이 손쉬운 해결책이다.

또 다른 해결방법은 시간을 통합하여 지수를

작성하는 것이다. 예를 들어 분기별 지수를 반년(semi-annual) 지수나 연간(annual) 지수로 작성하는 방안이다. 이 경우 지수를 시간대별로 평활(smoothing)시켜 변동성을 낮게 추정함으로써 포트폴리오 구성에 왜곡된 정보를 제공하는 문제가 있다(Geltner, 1991, 1993a, 1993b). 또한 시간별로 집계하는 것은 시장 감시(monitoring) 도구로서 용도를 상실하는 것이 된다.

본 논문에서는 거래빈도가 낮은 시장에서의 가격지수를 추정하기 위한 방법론을 제시하고자 한다. 본 논문에서 제시하는 방법은 과거 거래된 실거래 자료와 현재의 실거래자료를 통합하는 방안으로서 시계열 분석방법에 기초한다. 즉, 헤도닉 가격모형에서 여러 개의 시간더미변수를 몇 개의 모수만을 가지는 시계열 함수로 대체하는 방법으로 Quan and Quigley(1971)에 의해 처음 제안되고 Schwann(1998)에 의해 발전되었다. Quan and Quigley에 따르면 거래 자료들은 실제 가격에 오차가 추가되어 있으며 시장가격에 대한 최적 갱신된 값은 과거와 현재 시장정보에 대한 가중치로 나타낼 수 있다고 보았다. Schwann에 따르면 시계열에 기초한 가격지수는 과거에 거래된 자료와 현재 거래된 자료를 연결시킴으로서 가격지수의 기반이 되는 비교거래 자료의 집합을 늘릴 수 있다고 보았다. 이와 같이 시계열 분석방법에 의한 지수작성방법은 기존 방법들과 비해 모형에서 추정하여야하는 모수들이 적기 때문에 거래빈도가 낮은 시장에서 유용하다.

## II. 부동산가격지수의 시계열 구조

일반적으로 지수들은 다음 식의 형태를 가진

다(Schwann, 1998).

$$\ln I_t = \ln I_{t-1} + \Delta \ln I_t \quad (1)$$

예를 들어, 헤도닉 가격지수를 생각해보자. 헤도닉 가격지수는 다음 식에 기초하고 있다.

$$p_i = \alpha_0 + \sum_{k=1}^K \alpha_k x_{ki} + \sum_{t=2}^T \beta_t d_{it} + \epsilon_i, \quad i = 1, \dots, N \quad (2)$$

여기서  $p_i$ 는  $i$  번째 부동산의 가격이며,  $x_{ki}$ ,  $k=1, \dots, K$ 는 그 부동산이 가지고 있는 특성들이다.  $d_{it}$ ,  $t=2, \dots, T$ 는 더미변수로서  $i$  번째 부동산이  $t$  시점에 거래가 이루어지면 1, 기타는 0인 값을 갖는다.  $\epsilon_i$ 는 교란항(disturbance terms)이다. 본 논문에서  $p_i$ 와  $x_{ki}$ 는 로그 값으로 측정되었다고 가정한다. 이 경우, 계수  $\beta_t$ 는 시기 1에서 시기  $t$  사이의 부동산의 시장가치의 증가율이 되며  $I_t = \exp(\beta_t)$ 는  $t$  시기에 대한 부동산 가격지수를 나타낸다. 따라서 이 식의 형식은 식(1)의 따르지는 않는다.

(2)식을 다음 식으로 나타내면

$$p_i = \alpha_0 + \sum_{k=1}^K \alpha_k x_{ki} + \sum_{t=2}^T \delta_t f_{it} + \epsilon_i, \quad i = 1, \dots, N \quad (3)$$

여기서 더미변수  $f_{it}$ 를  $t$  시점 또는 그 이전에 거래된 경우 1 기타 0의 값을 가지는 것으로 정의하면 (3)식에서 계수  $\delta_t$ 는 시기  $t-1$ 과  $t$  사이에 부동산의 시장가치의 증가율을 나타내며 가격지수는 다음 식으로 나타낼 수 있다.

$$I_t = \exp\left(\sum_{s=1}^t \delta_s\right) = I_{t-1} \exp(\delta_t) \quad (4)$$

(4) 식은 시기  $t$ 에 대한 부동산 가격지수를 나타낸다. 이 식을 양변에 로그를 취하면 (1)식의 형태를 가지는 것을 알 수 있다. (2)식과 (3)식은 같은 정보를 가지는 추정식으로 더미변수  $f_{it}$ 는  $d_{it}$ 의 선형결합으로 나타낼 수 있다. 따라서 이들 식으로부터 유도되는 가격지수는 동일하다.

식(1)의 시계열 특성을 살펴보면 본질적으로 로그가격지수는 불안정성(nonstationary)을 전제로 하고 있다. 부동산 가격지수의 불안정성에 대한 검증으로 Kuo(1997), Case and Shiller (1989), Meese and Wallace (1991a), 경기개발연구원(2007) 등에서 연구된 바 있으나 결과에 대해서는 일치된 것이 없다. 다음으로 가격증가율  $\Delta \ln I_t$ 에 대해, Case and Shiller(1989), Meese and Wallace (1991a)는 주택가격의 증가율은 단위근(unit root)이 존재하지 않는 안정성을 가진다고 보았지만 다른 연구에서는 주택가격은 계열상관(serial correlation)이 존재하는 불안정한 시계열이라고 보았다(Gatzlaff and Tirtiroglu, 1995).

자기상관은 연구하는 시장에 따라 다른 것으로 나타났다. Gunterman and Smith(1987)은 시차가 적은 경우 계열 상관의 존재는 일정하지 않지만 3년 이상의 시차에 대해서는 부(-)의 자기상관을 갖는다고 보고 있다. 반면 Rayburn, Devaney and Evans (1987)는 자기상관의 크기와 패턴은 지역마다 다르다고 주장하였으며, Gatzlaff(1994)는 주택가격증가율에 있어 자기상관은 분석하는 시기에 따라 달라진다고 보았다.

선행연구에서 부동산가격 증가율에 대해서는 일치된 결론이 없기 때문에 본 논문에서는

$\Delta \ln I_t$ 는 안정적(stationary)인 자기회귀(autoregressive; AR) 과정을 따른다고 가정하고 AR의 차수에 대해서는 실증분석에서 명시적으로 검정하고자 한다. 따라서 본 논문에서는 다음 모형을 기초로 분석을 한다.

$$p_{it} = \alpha_0 + \sum_{k=1}^K \alpha_k x_{kit} + \ln I_t + \epsilon_{it},$$

$$i = 1, \dots, N_i; t = 1, \dots, T \quad (6)$$

여기서

$$\ln I_t = \ln I_{t-1} + \sum_{r=1}^s \psi_r \Delta \ln I_{t-r} + \xi_t \quad t = 1, \dots, T \quad (7)$$

그리고  $I_0 = 1$ 이다. 식(6)은 헤도닉 모형식 (2) 또는 (3)식에 더미변수에 대해서 지수의 로그 값으로 대체하여 다시 쓴 것이다. 그렇지만 수식의 표현에 있어 약간의 변화가 있다.  $i$  번째 주택이  $t$  시점에 거래된 가격을  $p_{it}$ 로 나타내었으며  $x_{kit}$ 는 이들 주택의  $k$  번째 속성을 나타낸다.  $\epsilon_{it}$ 는 오차항으로 평균이 0이고 분산이  $\sigma_\epsilon^2$ 인 분포를 가진다. 수식을 이와 같이 나타내면 아래에서 보는바와 같이 로그우도함수를 이해하는데 도움이 된다. 식 (7)은 지수는 AR 증분(increments)을 가지는 단위근(unit root)을 가지는 확률과정임을 나타낸다. 교란항  $\xi_t$ 는 가격지수에서의 시간충격(temporal shock)을 나타내며 평균이 0이고 분산이 일정한  $\sigma_\xi^2$ 인 분포를 한다. 정의에 의해 충격은  $t$  시점에 거래되는 모든 부동산들에게 동일하다. 비율  $\sigma_\xi^2/\sigma_\epsilon^2$ 은 시간충격에 대한 오차 비율(signal-to-noise)을 나타낸다. 마지막으로 초기조건  $I_0 = 1$ 은 가격지수를 초기화하는데 필요하다.

본 논문에서 주택가격 함수식은 헤도닉가격함수(hedonic price function)를 사용한다. 반복매매도 가능하지만 우리나라에서 실거래가격이 발표된 것은 기간이 짧아 반복매매 쌍을 확보하는 것이 매우 적기 때문에 헤도닉가격함수를 사용하기로 한다. 헤도닉가격 함수를 이용하여 지수를 추정할 경우 지수가 헤도닉 함수식에 의존하고 주택에 있어 질적 차이들을 통제하기 위해 사용한 설명변수에 의존한다는 것이다. 이러한 결점들 때문에 헤도닉 모형보다 반복매매모형이 타당성이 있다고 종종 인용된다. 이에 대해 Dombrow, Knight and Sirmans (1997, p.76)은 “반복매매 모형은 헤도닉 모형으로부터 유도된 것이기 때문에 헤도닉 모형이 갖는 약점들이 반복매매모형에도 존재하지만 단지 눈에 보이지 않을 뿐이다.”라고 주장하고 있다. 그들은 집계와 누락변수에 대한 편기(bias) 문제가 헤도닉 모형에서 뿐만 아니라 반복매매 모형에서도 발생할 수 있다고 입증하였다.

식 (6)은 특성가격은 시간에 따라 변하지 않는다는 제약적인 가정이다. Dombrow, Knight and Sirmans(1997)와 Meese and Wallace(1997)는 거래빈도가 낮은 시장에서는 제약이 없는 헤도닉 함수식을 정확하게 추정할 만큼 자료가 충분하지 않기 때문에 추정된 모형을 사용하여 지수를 집계하게 되면 편기(bias)가 발생할 수 있다고 주장한다.

(6)식과 (7)식을 통합하면

$$p_{it} = \alpha_0 + \sum_{k=1}^K \alpha_k x_{kit} + \ln I_{t-1} + \sum_{r=1}^s \psi_r \Delta \ln I_{t-r} + (\xi_t + \epsilon_{it}), \quad (8)$$

추정식에서 오차는 개별 부동산에 대한 오차와 가격지수에 대한 교란항의 합이다. Case and Shiller(1989), Geltner(1997), Goetzmann and Spiegel (1995, 1997), Kuo(1997)의 연구에서도 유사하게 오차항을 규정하였다.

(8)식으로부터 관측치에 대한 로그-우도는 예측오차분해(prediction error decomposition)를 통해 유도된다. 즉,

$$\ln L(\mathbf{p}_1, \dots, \mathbf{p}_T) = \sum_{t=1}^T \ln L(\mathbf{p}_t | P_{t-1}), \quad (9)$$

여기서  $\mathbf{p}_t = (\mathbf{p}_{1t}, \dots, \mathbf{p}_{N_t, t})$ 는  $t$  시기에 거래된 부동산들의 가격들이며,  $P_{t-1} = \{\mathbf{p}_{t-1}, \mathbf{p}_{t-2}, \dots, \mathbf{p}_1\}$ 는  $t-1$  시기까지의 정보 집합을 나타낸다. 그러므로  $L(\mathbf{p}_t | P_{t-1})$ 는 모든 이전의 거래가격에 대한 정보들이 주어졌을 때  $t$  시기 가격의 조건부 분포이다.  $\varepsilon_{it}$ 와  $\xi_t$ 는 상관되어있지 않다는 가정을 한다. 이는 부동산 가격에서의 측정오차들은 지수를 발생시키는 확률과정과는 상관되어 있지 않다는 것이다. 또한 이들 분포들은 모두 독립적이고 분포가 일정한 정규분포를 한다고 가정한다.  $\sigma_\xi^2$ 는 부동산가격에 있어 구조적인 변동성을 측정하는 것이며  $\sigma_\xi^2 / (1 - \sum_{r=1}^s \psi_r)$ 는 가격에서 변동성의 장기효과를 나타낸다.

$\mathbf{v}_{it} = \varepsilon_{it} + \xi_t$ 라 하고 합성회귀오차라 하고  $\mathbf{v}_t = (\mathbf{v}_{1t}, \dots, \mathbf{v}_{N_t, t})'$ 는 시기  $t$ 에 거래된 부동산들의 가격에서의 오차 벡터라 하면

$$E(\mathbf{v}_t) = 0 \quad (10)$$

그리고

$$\Omega_t = E(\mathbf{v}_t \mathbf{v}_t') = \begin{bmatrix} \sigma_\varepsilon^2 + \sigma_\xi^2 & \sigma_\xi^2 & \dots & \sigma_\xi^2 \\ \sigma_\xi^2 & \sigma_\varepsilon^2 + \sigma_\xi^2 & \dots & \sigma_\xi^2 \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_\xi^2 & \sigma_\xi^2 & \dots & \sigma_\varepsilon^2 + \sigma_\xi^2 \end{bmatrix} \quad (11)$$

공분산행렬의 차수는 시기에 얼마나 많은 거래가 있느냐에 달려 있다. 이 공분산 행렬은  $\sigma_\varepsilon^2 > 0$ 인 경우 즉, 부동산가격에 있어 개별적인 차이가 존재하는 한 비특이(nonsingular)행렬이며 양정(positive definite) 행렬이다.

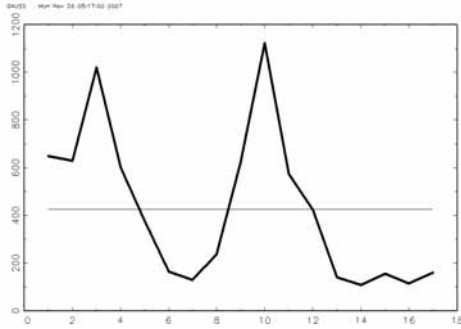
공분산 행렬 구조는 동일한 시기에 거래가 이루어진 주택들의 가격증가율에 대한 상관관계를 나타낸다. 이 상관계수가 양수이고 모든 주택에 있어  $\sigma_\varepsilon^2 / (\sigma_\varepsilon^2 + \sigma_\xi^2)$ 이다. Goetzmann and Spiegel (1995)는 유사하지만 다소 제약이 완화된 가정을 하고 있다. 비록 그들의 모형은 보다 일반적인 상관계수 구조를 가지고 있지만 소규모 표본에만 직접적으로 추정이 가능하다. 주택의 가격변동률은 시간뿐만이 아니라 공간적으로도 상관되어 있을 수 있다. 이러한 아이디어는 Can and Megboulougbe(1997), Case and Mayer (1996), Goetzmann and Spiegel (1997) 연구에서 다루고 있다.

위 가정 하에서 로그-우도는 다음 식으로 나타낼 수 있다.

$$\ln L = \frac{N}{2} \ln 2\pi - \frac{1}{2} \sum_{t=1}^T \ln |\Omega_t| - \frac{1}{2} \sum_{t=1}^T \mathbf{v}_t' \Omega_t^{-1} \mathbf{v}_t. \quad (12)$$

모수 추정은 (12)식을  $\alpha_0, \alpha_1, \dots, \alpha_K, \psi_1, \dots, \psi_s, \sigma_\varepsilon^2$  그리고  $\sigma_\xi^2$ 에 대해서 (12)식을 최대화하도록 모수들을 추정한다. 로그-우도를 평가하기 위하여 칼만필터(Kalman-filter)를 사용한다. 모형이 불안정한 시계열이기 때문에 de Jong (1989, 1991)이

〈그림 1〉 서울 강남구 아파트 실거래 건수  
(2006년 1월-2007년 5월)



제시한 확산(diffuse) 칼만필터 알고리즘을 사용한다.

### III. 자료

본 연구에서는 서울 강남구 실거래자료를 사용하였다. 분석에 사용한 자료는 2006년 1월부터 2007년 5월 사이에 거래된 실거래자료로서 주요 속성변수로 전용면적( $m^2$ ), 방수, 욕실수, 층수, 단지규모, 출입문구조, 난방방식, 건물연수, 거래시기이다. 실거래자료는 부동산실거래가(rt.mocet.go.kr)

〈표 1〉 헤도닉 모형 추정 결과

변수	모형 1 (실거래 전체)		모형 2 (실거래 10개)		모형 3 (실거래 2개)	
	계수	t값	계수	t값	계수	t값
상수항	4.1215	(114.755)	3.5567	(14.035)	2.8340	(4.135)
전용면적	1.3997	(188.017)	1.5131	(28.523)	1.5117	(15.077)
건물연수	0.1148	(23.523)	0.0887	(2.468)	0.1896	(2.600)
단지규모	0.2455	(71.165)	0.2623	(10.401)	0.4065	(7.652)
계단식	0.2076	(33.297)	0.1960	(4.375)	0.0281	(0.295)
고층아파트	0.1271	(12.561)	0.0438	(0.629)	0.4301	(2.664)
200602	0.0246	(1.727)	0.1638	(1.385)	0.1549	(0.719)
200603	0.0920	(7.183)	0.1160	(0.972)	0.3782	(2.032)
200604	0.1240	(8.593)	0.2138	(1.808)	0.2860	(1.433)
200605	0.0771	(4.651)	0.2596	(2.173)	0.3708	(1.555)
200606	0.0676	(3.019)	0.1854	(1.542)	0.3560	(1.967)
200607	0.0939	(3.823)	0.2904	(2.472)	0.9069	(3.986)
200608	0.0637	(3.295)	0.2444	(2.085)	0.4286	(2.307)
200609	0.1176	(8.231)	0.2483	(2.062)	0.5391	(2.458)
200610	0.1901	(15.127)	0.3201	(2.725)	0.6482	(2.545)
200611	0.2574	(17.560)	0.3987	(3.411)	0.9749	(4.129)
200612	0.2509	(15.739)	0.3065	(2.579)	0.8444	(3.924)
200701	0.2058	(8.634)	0.5781	(4.885)	0.4172	(1.808)
200702	0.2749	(10.324)	0.3357	(2.813)	0.7543	(3.779)
200703	0.2531	(11.087)	0.4484	(3.688)	0.9058	(4.713)
200702	0.2253	(8.703)	0.3110	(2.609)	0.5171	(2.419)
200703	0.2267	(10.029)	0.4369	(3.648)	0.5349	(2.835)
관측치수	7,223		170		34	
결정계수	0.8667		0.8554		0.8572	
로그우도값	-350.59		-11,5944		12,4565	
RMSE	0.2540		0.2590		0.1677	

주: ( )은 t 값임

사이트에서 7,223개 실거래 자료에 대해 부동산 114(www.r114.co.kr)와 부동산테크(www.tech.co.kr) 사이트를 통해 아파트 단지 속성을 구하였다. <그림 1>은 강남구 실거래 자료건수에 대한 그래프이다. 가는 실선은 평균 거래건수 425건을 나타내고 있다. 2006년 10월에 1,123로 가장 많이 거래되었으며, 그 다음으로 2006년 3월 1,020건이 거래되었다. 반면 2007년 2월 거래건수가 107건으로 가장 적으며, 대체적으로 2007년 들어서 거래건수가 급감하여 150건 이내에서 거래가 이루어지고 있다.

#### IV. 모형 추정 및 지수 비교

<표 1>은 헤도닉 가격모형 추정결과를 보여주고 있다. 모형 1은 실거래 전체 자료를 사용한 것이다. 모형 2는 거래빈도가 낮은 시장을 가상적으로 나타낸 모형으로 매월 거래된 자료 가운데 임의적으로 10개의 실거래 자료를 추출한 것

이다. 모형 3은 거래빈도가 아주 낮은 시장으로 매월 거래된 실거래 자료 가운데 임의적으로 2개를 추출한 자료를 가지고 추정한 것이다.

모형 추정 결과 거래빈도가 낮아질수록 주택 특성들의 특성가격, 즉 주택가격 탄력성이 매우 크게 변하는 것을 알 수 있다. 예를 들어 계단식 더미변수의 경우 모형1에서는 복도식 아파트에 비해 계단식 아파트의 가격이 평균 20% 높은 것으로 나타났으나 모형3에서는 2.8% 높은 것으로 그 차이가 매우 크게 나타난다. 고층아파트 더미변수의 경우 모형1에서는 12.7%로 나타난 반면 모형3에서는 43.0%로 특성가격의 차가 매우 크게 나타난다.

<표 2>는 시계열 모형 추정결과이다. 헤도닉 모형은 16개의 시간 더미변수를 사용하여 매월 가격의 평균적인 변동을 통제하여 추정한 것이고 시계열 모형의 추정결과는 식 (7)의 확률과정(stochastic process)을 가정하여 추정한 것이다. 확률과정은 단위근 과정과 주택가격 변화율을 차수가 1인 자기회귀과정 AR(1)으로 설정하여 확

<표 2> 시계열 모형 추정 결과

변수	모형 1 (실거래 전체)		모형 2 (실거래 10개)		모형 3 (실거래 2개)	
	계수	t값	계수	t값	계수	t값
상수항	4.0930	(12.526)	3.6989	(10.765)	3.6378	(4.242)
전용면적	1.4024	(3.094)	1.4957	(21.147)	1.4390	(8.187)
건물연수	0.1108	(2.196)	0.0984	(1.991)	0.1521	(1.267)
세대수	0.2454	(2.472)	0.2518	(8.664)	0.2848	(3.015)
계단식	0.2129	(2.693)	0.2077	(3.797)	0.2376	(1.296)
고층아파트	0.1276	(0.764)	0.0592	(0.886)	0.2629	(1.279)
AR(1)	1.1145	(2.412)	0.9229	(2.715)	0.9284	(1.866)
$\sigma_\epsilon$	3.3626	(0.403)	0.7045	(4.077)	0.2827	(0.485)
$\sigma_\xi$	1.1197	(0.088)	0.0920	(0.358)	0.0321	(0.106)

주: ( )은 t 값임



틀과정의 오차항의 분산인  $\sigma_{\epsilon}^2$ 과 자기회귀 모수  $\phi$  두 개를 추정하여 얻은 결과이다. 시계열 모형은 헤도닉 모형에 비해 추정하여야 하는 모수의 개수가 16개에서 2개로 14개가 줄어들어 거래빈도가 낮은 시장에서도 자유도의 문제를 크게 해결할 수 있는 것을 알 수 있다.

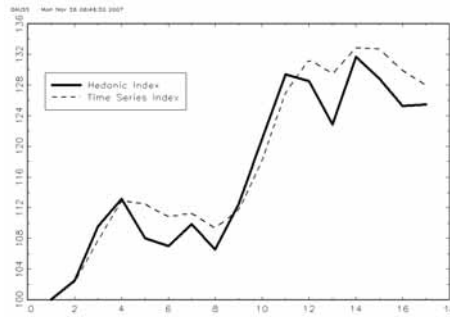
또한 모수의 추정결과를 보면 모형1의 경우 시계열 모형과 헤도닉 가격모형에 큰 차이가 없음을 알 수 있다. 모형2와 모형3의 거래빈도가 낮은 경우에도 모형1과 특성가격의 추정치의 변화가 크지 않음을 알 수 있다. 예를 들어 계단식 아파트의 경우 복도식 아파트에 비해 주택가격이 24.54% (모형1) 높은 것에 대해 모형2는 20.77%, 모형3은 23.76%로 비교적 큰 차이를 보이지 않는다. 반면 헤도닉 모형의 경우 모형1은 20.76%에서 모형2는 19.60%, 모형3은 2.81%로 변화가 모형 1과 모형3의 변화가 상당히 큰 것을 알 수 있다. 다른 변수에 대해서도 비슷한 패턴을 보이고 있다.

이는 헤도닉 모형의 경우 실거래자료가 2개만 존재할 경우 이상치(outlier)가 존재하는 경우 이의 영향이 지수작성에 매우 크게 반영되는 반면 시계열 모형은 비록 실거래 자료가 적은 시장에서도 과거의 실거래자료와 현재의 실거래자료를 통합하여 가격지수에 대한 확률과정을 통해 지수를 작성하기 때문에 실거래 자료가 이상치일 경우라도 이것이 지수작성에 영향을 미치는 것이 크지 않기 때문이다.

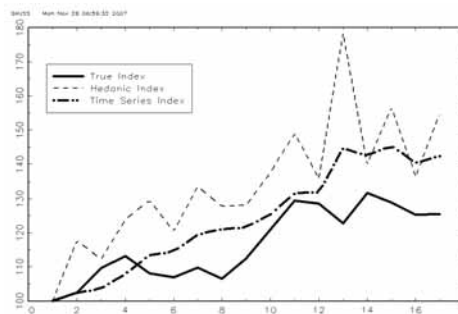
헤도닉 모형과 시계열 모형으로부터 지수를 작성하여 이를 비교하여 보자. 비교의 편의상 전체 실거래 자료를 이용한 헤도닉 가격지수를 “실제지수”라고 부르고 거래빈도가 낮은 경우의 지수들을 비교한다.

<그림 2>에서 실제지수는 굵은 실선이다. 2006년 1월부터 4월까지 지수가 상승하다 4월이후부터 8월까지 하락하는 모습을 보이고 있다. 2006년 8월 이후부터 연말까지 가격이 급등하였다. 2007년에 들어와서는 급락을 반복하는 패턴을 보이고 있다. 시계열 지수는 실제지수를 잘 반영하는 것으로 나타나고 있다. 단지 거래량이 적은 시기에서 차이가 크게 나타나고 있으나 대체적으로 헤도닉 가격지수에 비해 변동성이 낮은 것이 특징이다.

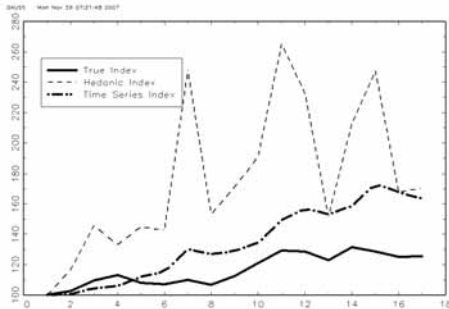
<그림 3>과 <그림 4>는 모형 2와 모형 3에 대해 <그림 2> 헤도닉 가격지수와 시계열 가격지수 비교 (모형 1)



<그림 3> 거래빈도가 낮은 시장에서의 지수비교 (모형 2 경우)



〈그림 4〉 거래빈도가 낮은 시장에서의 지수비교  
(모형 3 경우)



한 헤도닉 가격지수와 시계열 지수를 실제지수와 비교하여 나타낸 것이다. 헤도닉 가격지수의 경우 거래빈도가 작은 경우 지수의 불안정성이 큰 것을 알 수 있으며 실거래자료수가 2개와 같이 극단적인 경우는 변동성이 더욱 큰 것을 알 수 있다. 지수의 변동성은 포트폴리오 구성에 있어 커다란 영향을 미치게 되기 때문에 지수작성에 주의가 필요하다.

〈표 3〉은 헤도닉 가격지수와 시계열 가격지수에 대한 지수의 평균오차와 실제지수와 평균자승오차근(RMSE)를 비교하여 보았다. 거래빈도가

〈표 3〉 시계열 모형 추정 결과

변수	모형 1 (실거래 전체)	모형 2 (실거래 10개)	모형 3 (실거래 2개)
헤도닉 가격지수			
평균오차	2,1287	16,3778	39,4779
RMSE	-	22,1924	74,4774
시계열 가격지수			
평균오차	28,3437	11,9280	29,8896
RMSE	3,1465	11,0477	24,4754

낮은 시장에서는 평균오차가 헤도닉 가격지수에 비해 시계열 가격지수가 대체적으로 낮은 것을 알 수 있으며, 실제지수와 RMSE도 시계열 가격지수가 낮은 것을 확인할 수 있다.

## V. 결론

실거래 자료가 작은 시장에서 헤도닉 방법에 의한 가격지수는 지수의 변동성이 크게 나타나는 문제와 지수의 정확성이 떨어지는 문제가 있다. 변동성이 커지게 되면 포트폴리오 구성에 있어 부동산의 비중을 낮추게 하는 문제가 있다. 또한 지수의 부정확성은 지수를 기초로 시장에 대한 분석과 정책적인 판단에 오류로 이어지기 때문에 이에 대한 해결방안이 필요하다.

본 논문에서는 기존의 전통적인 헤도닉 가격지수 방법에 대한적인 방법으로 시계열 분석 방법에 의한 지수작성을 제시하였다. 헤도닉 모형은 16개의 시간 더미변수를 사용하여 매월 가격의 평균적인 변동을 통제하여 추정된 것이고 시계열 모형의 추정결과는 식 (7)의 확률과정(stochastic process)을 가정하여 추정된 것이다. 확률과정은 단위근 과정과 주택가격 변화율을 차수가 1인 자기회귀과정 AR(1)으로 설정하여 확률과정의 오차항의 분산인  $\sigma_{\epsilon}^2$  과 자기회귀 모수  $\phi$  두 개를 추정하여 얻은 결과이다. 시계열 모형은 헤도닉 모형에 비해 추정하여야 하는 모수의 개수가 16개에서 2개로 14개가 줄어들어 거래빈도가 낮은 시장에서도 자유도의 문제를 크게 해결할 수 있는 것을 알 수 있다.

시계열 모형에 의한 모수추정은 전체 실거래 자료를 이용한 헤도닉 가격모형의 추정결과와 큰

차이가 없는 반면 헤도닉 가격모형의 경우 실거래 자료가 적을 경우 모수 추정이 매우 불안정한 것을 알 수 있다. 모수 추정결과를 바탕으로 지수를 작성하는 것을 감안할 때 헤도닉 가격지수는 함수 모형의 설정문제와 표본선택(sample selection)의 문제에 대해 거래빈도가 낮은 지역의 시장에서는 표본선택의 문제가 매우 심각한 것을 알 수 있다.

특히 실거래자료가 2개만 존재하며 이상치(outlier)가 존재하는 경우 이의 영향이 지수작성에 매우 크게 반영된다. 또한 극단적으로 특정 기간에 실거래가 이루어지지 않을 경우는 지수자체의 작성이 불가능하다. 반면 시계열 모형은 비록 실거래 자료가 적은 시장에서도 과거의 실거래자료와 현재의 실거래자료를 통합하여 가격지수에 대한 확률과정을 통해 지수를 작성하기 때문에 실거래 자료가 이상치일 경우라도 이것이 지수작성에 영향을 미치는 것이 크지 않으며, 실거래 자료가 존재하지 않을 경우에도 지수작성이 가능하다.

결론적으로 거래빈도가 낮은 경우, 헤도닉 지수는 변동성이 크고, 이상치의 영향이 큰 반면, 시계열 지수는 과거의 실거래자료들을 현재 실거래자료에 대한 정보로 사용하며 이상치의 영향력을 낮출 수 있어, 비교적 안정적인 지수를 얻을 수 있다. 실거래자료를 기초로 전국 모든 지역에 대한 실거래가격지수를 작성하고자 할 때, 거래빈도가 적은 지역에 대해 실거래 가격지수를 작성할 수 있는 방법을 제시하였다는 점에서 나름대로 의의가 있다고 본다. 하지만 본 논문에서 가정된 시계열에 대한 확률과정은 간단한 과정이기 때문에 실제 적용을 위해서는 확률과정에 대한 추가적인 연구가 필요하다.

접 수 일 : 2007년 11 월 30 일

심사완료일 : 2007년 12 월 26 일

## 참고문헌

1. 경기개발연구원 (2007). 「경기도 부동산 데이터베이스 구축실태 파악 및 활용방안에 관한 연구」. 2007.
2. 이용만. “특성가격함수를 이용한 주택가격지수 개발에 관한 연구-시간변동계수모형에 의한 연쇄지수,” 「부동산학연구」, 제13집 제1호, 2007, pp. 103-125.
3. 한국감정원. 「부동산 실거래가격에 기초한 주택가격지수 개발」. 2007.
4. Bailey, M. R. Muth, and H. Nourse. “A Regression Method for Real Estate Price Index Construction,” *Journal of the American Statistical Association* 58, 1963, pp. 933-942.
5. Can, A., and I. Megbolugbe. “Spatial Dependence and House Price Index Construction,” *Journal of Real Estate Finance and Economics* 14(1,2), 1997, pp. 203-222.
6. Case, K., and C. J. Mayer. “Housing Price Dynamics Within a Metropolitan Area,” *Regional Science and Urban Economics* 26, 1996, pp. 387-407.
7. Case, B., and J. M. Quigley. “The Dynamics of Real Estate Prices,” *Review of Economics and Statistics* 73, 1991, pp. 50-58.
8. Case, B., H.O. Pollaskowski and S.M. Wachter. “On Choosing among House Price Index Methodologies,” *Journal of AREUEA*, 19, 1991, pp. 286-307.
9. Case, K., and R. Shiller. “The Efficiency of the Market for Single-Family Homes,” *American Economic Review* 79(1), 1989, pp. 125-137.
10. Clapp, J., and C. Giaccotto. “Estimating Price Indices for Residential Property: A Comparison of Repeat Sales and Assessed Value Methods,” *Journal of Real Estate Finance and Economics* 9(2), 1992, pp. 137-164.
11. de Jong, P. “Smoothing and Interpolation with the State-Space Model,” *Journal of the American Statistical Association* 84(408), 1989, pp. 1085-1088.
12. de Jong, P. “The Diffuse Kalman Filter,” *Annals of Statistics* 19(2), 1991, pp. 1073-1083.
13. Dombrow, J., J. R. Knight, and C. F. Sirmans. “Aggregation Bias in Repeat Sales Indices,” *Journal of Real Estate Finance and Economics* 14(1,2), 1997, pp. 75-88.
14. Gatzlaff, Dean H. “Excess Returns, Inflation and the Efficiency of the Housing Market,” *Journal of the American Real Estate and Urban Economics Association* 22(4), 1994, pp. 553-581.
15. Gatzlaff, Dean H., and Donald R. Haurin. “Sample Selection Bias in Repeat-Sales Index Estimates,” *Journal of Real Estate Finance and Economics* 14(1,2), 1997, pp. 33- 50.

16. Gatzlaff, Dean H., and D. C. Ling. "Measuring Changes in Local House Prices: An Empirical Investigation of Alternate Methodologies," *Journal of Urban Economics* 35, 1994, pp. 221-244.
17. Gatzlaff, Dean H., and Dogan Tirtiroglu. "Real Estate Market Efficiency: Issues and Evidence," *Journal Real Estate Literature*, 3(2), 1995, pp. 157-189.
18. Geltner, D. "Smoothing in Appraisal- Based Returns," *Journal of Real Estate Finance and Economics* 4(3), 1991, pp. 327-345.
19. Geltner, D. "Estimating Market Values from Appraised Values Without Assuming an Efficient Market," *Journal of Real Estate Research* 8, 1993a, pp. 325-346.
20. Geltner, D. "Temporal Aggregation in Real Estate Return Indices," *Journal of the American Real Estate and Urban Economics Association* 21(2), 1993b, pp. 141-166.
21. Geltner, D. "Bias and Precision of Estimates of Housing Investment Risk Based on Repeat-Sales Indexes: A Simulation Analysis," *Journal of Real Estate Finance and Economics* 14(1,2), 1997, pp. 155-172.
22. Goetzmann, W. "The Accuracy of Real Estate Indices: Repeat Sale Estimators," *Journal of Estate Finance and Economics* 5(1), 1992, pp. 5-14.
23. Goetzmann, W., and M. Spiegel. "Non-Temporal Components of Residential Price Appreciation," *Review of Economics and Statistics* 77, 1995, pp. 199-206.
24. Goetzmann, W. and M. Spiegel. "A Spatial Model of Housing Returns and Neighborhood Substitutability," *Journal of Real Estate Finance and Economics* 14(1,2), 1997, pp. 11-32.
25. Gunterman, Karl L., and Richard L. Smith. "Efficiency of the Market for Residential Real Estate," *Land Economics* 63(1), 1987. pp. 34-43.
26. Kuo, C. "A Bayesian Approach to the Construction and Comparison of Alternative House Price Indices," *Journal of Real Estate Finance and Economics* 14(1,2), 1997, pp. 113-132.
27. Lancaster, K. J. "A New Approach to Consumer Theory," *Journal of Political Economy*, 74, 1966, pp. 132-157.
28. Meese, R. and N. E. Wallace. "Determinants of Residential Housing Prices: Effects of Economic Factors or Speculative Bubbles?" Working paper, Center for Real Estate and Urban Economics, University of California at Berkeley, 1991a
29. Meese, R. and N. E. Wallace. "Non-parametric Estimation of Dynamic Hedonic Price Models and the Construction of Residential Housing Price Indexes." *AREUEA Journal* 19, 1991b, pp. 308-332.
30. Meese, R. and N. E. Wallace. "The

- Construction of Residential House Price Indices: A Comparison of Repeat-Sales, Hedonic-Regression, and Hybrid Approaches," *Journal of Real Estate Finance and Economics* 5(1), 1997, pp. 51-74.
31. Palmquist, R. B. "Measuring Environmental Effects on Property Values Without Hedonic Regressions," *Journal of Urban Economics* 11, 1982, pp. 333-347
32. Quan, D., and J. Quigley. "Price Formation and the Appraisal Function in Real Estate Markets," *Journal of Real Estate Finance and Economics* 4, 1991, pp. 175-190.
33. Rayburn, William, Michael Devaney, and Richard Evans. "A Test of Weak-Form Efficiency in Residential Real Estate Returns," *Journal of the American Real Estate and Urban Economics Association* , 1987, pp. 220-233.
34. Rosen, S. "Hedonic Prices and Implicit Markets: Product Differentiation in Pure Competition," *Journal of Political Economy*, 82, 1974, pp. 34-55.
35. Schwann, G. M. "A Real Estate Price Index for Thin Markets." *Journal of Real Estate Finance and Economics* 16(3), 1998, pp. 269-287.
36. Webb, C. "A Probabilistic Model for Price Levels in Discontinuous Markets." In W. Eichorn (ed.), *Measurement in Economics*. Heidelberg: Physic-Verlag. 1988.