

## 상태공간모형을 활용한 부동산실거래가격지수 추정에 관한 연구 :거래빈도가 낮은 지역을 중심으로\*

A Study on the Construction of the Transaction-based Real Estate Price Indices Using  
State Space Model  
:Focusing on Thin Markets

박 헌 수 (Park, Heonsoo)\*\*

유 은 영 (Yoo, Eunyoung)\*\*\*

### < Abstract >

This paper estimates the transaction-based real estate price indices for thin markets that have few transactions by using a state space model. The thinness of the market does have a marked effect on the precision of the price index estimate. Since the volatility of the price indices for thin markets estimated by the hedonic price model or repeated-sales model tends to be high, the precision of them is question of our interest. In this paper, we suggest an alternative approach to make stable price indices even when the number of transactions are small or even does not exist. We have developed the transaction-based price indices for the apartment of Gangnam-Gu and Jongno-Gu, Seoul, by using state space models which are estimated by the Kalman filtering and EM(Expectation and Maximization) algorithm. In order to consider the thin markets, we divide housing market by its size: small-sized and medium and large sized apartment. We find that our suggested apartment price indices have lower volatility and much more accurate than the traditional repeated sales indices.

주 제 어 : 실거래가격지수, 상태공간모형, 칼만필터링, 부동산 하위시장

key word : Transaction-based Price Indices, State Space Model, Kalman-filtering, Thin Market

\* 본 논문은 2013년 국토교통부, 한국감정원, 한국감정평가협회와 한국부동산분석학회가 공동 주관한 “제 1회 가격공시제도 및 감정평가산업 발전 모색을 위한 우수논문 공모전”에서 우수상을 수상한 논문임.

\*\* 중앙대학교 도시계획부동산학과 교수, [heonsoo@cau.ac.kr](mailto:heonsoo@cau.ac.kr)(주저자)

\*\*\* 중앙대학교 도시계획부동산학과 박사수료, [eunci71@naver.com](mailto:eunci71@naver.com)(교신저자)

## I. 서론

우리나라에서 부동산은 국가경제에서 차지하는 비중이 크다. 2013년 공동주택과 단독주택 공시가격 총액은 약 2,200조원이다. 통계청(2013)에 따르면 2013년 3월말 현재 가계자산 중 부동산 등의 실물자산이 차지하는 비중은 73.3%로 금융자산 26.7%의 3배에 달하고 있다. 가계의 부동산 투자는 주식, 채권, 개인연금 등 포트폴리오 구성에 있어 중요하다.

부동산가격지수<sup>1)</sup>는 부동산 가격의 시간에 따른 변화를 나타낸 것이다. 부동산가격지수는 주택연금이나 주택담보대출에서 부동산의 시장가치를 평가하는데 사용된다. 또한 투기과열지구나 주택거래신고지역을 지정할 때, 또는 지정된 지역을 해제하는 등 주택관련 정책을 결정할 때 중요한 판단근거로 사용되고 있다.

그동안 우리나라의 부동산시장은 주기적으로 가격 급등과 급락, 안정을 반복해 왔다. 특히, '90년대 말 외환위기 이후 부동산 가격이 급등하면서 투기, 서민 주거불안 및 거품 논란 등의 문제가 이슈화되면서 10.29대책, 8.31대책, 3.30 조치, 11.15 대책 등이 발표되었다. 그러나 이러한 정부의 시장개입에도 불구하고 여전히 부동산시장에는 불안요인이 존재하고 있다.

부동산 시장의 불안요인을 줄이고 건전한 부동산시장의 육성을 위해 정부는 2005년 7월 「공인중개사의 업무 및 부동산 거래신고에 관한 법률」을 제정을 통해 2006년 1월부터 모든 주택거래는 실거래가격을 신고하도록 의무화하였다. 하지만, 실거래가격이 수집되었다고 하여 주택시

장 상황을 정확하게 반영하는 주택가격지수를 작성할 수 있는 것은 아니다. 실제로 시장에서 거래된 주택들은 특성들이 모두 다르기 때문에 가격변화가 특성변화에 기인한 것이 아니면 시장 상황에 따른 것인지를 구별할 필요가 있다. 실거래가격에 기초한 주택가격지수를 작성하기 위해서는 주택의 특성차이에 따른 가격 차이를 통제할 수 있는 지수작성방법이 필요하다.

그동안 주택가격지수 개발에 대한 많은 연구들이 이루어져 왔다. 가장 대표적인 방법은 시간이 경과하더라도 주택의 구조적 특성들이 변화하지 않는다는 가정 하에서 주택가격을 시차(lag)에 대하여 회귀시켜 가격지수를 산출하는 반복매매 모형(repeat sales model)과 주택가격을 주택특성들에 회귀시켜 가격지수를 산출하는 특성가격모형(hedonic price model)이다. 주택가격지수 작성에 대한 연구는 이 두 모형 간의 비교와 각 모형을 정교화시킴으로써 발전해 왔다(Bailey et al. 1963; Case and Shiller, 1989; Clapp and Giaccotto, 1992; Goetzmann, 1992).

정부는 부동산 실거래가를 기준으로 한 실거래가지수를 매월 발표하고 있다. 국내에서 사용된 지수작성방법은 동일한 주택에 대해 두 시점 이상 거래된 자료를 이용한 반복매매방법이다. 일반적으로 반복매매방법은 방대한 실거래자료가 구축되어 있을 경우에 유용하다. 2006년부터 실거래자료를 공개하고 있는 우리나라에서 이 방법을 사용한 이유는 아파트의 특성 때문이다. 아파트의 경우 비록 동일한 주택은 아니더라도 표준화된 주택으로 보아 동일주택 쌍으로 구성하여 지수를 작성하여도 큰 문제가 없다고 판단하였기 때문이다.

1) 본 논문에서 부동산가격지수와 주택가격지수는 같은 의미로 사용하기로 한다.

최근 빠른 속도로 고령화와 베이비붐 세대의 은퇴가 이루어지면서 노후대책의 한 수단으로서 주택연금에 대한 수요가 높아지고 있다. 하지만 많은 주택연금에 대한 실수요자는 아파트가 아닌 단독주택, 다세대/다가구 주택에 거주하는 사람이 많아 아파트 이외 다양한 주택유형별 가격지수가 필요하다. 그러나 아파트 이외 다른 주택유형들에 대해 아파트에 적용한 표준주택기준을 사용하여 실거래가격지수를 작성하는 것은 문제가 있다. 특히, 단독주택의 경우 아파트에 비해 하위 시장이 크게 발달되어 있어 세부적인 공간단위에 대한 가격지수가 필요하다.

하부지역으로 세분화하여 가격지수를 작성하는 경우 실거래 자료가 없어 지수를 작성할 수 없거나, 실거래가 있더라도 거래빈도가 많지 않아 지수의 변동성이 커지는 문제가 있다. 이러한 문제 때문에, 아파트에 대한 실거래가 지수의 경우에도 실제로 발표되는 지수는 서울의 경우 크게 4개의 권역(도심권, 동북권, 동남권, 서북권, 서남권)으로 구분하여 발표되고 있다.

거래빈도가 낮을 경우 커지는 지수의 변동성은 부동산 시장에 미치는 영향이 크다. 가격지수의 변동성은 투자자의 입장에서 위험(risk)이기 때문에 중요한 지표이다. 하지만 낮은 거래빈도로 인한 높은 변동성은 실제 시장의 위험보다는 지수작성의 문제로 발생하는 경우이므로 정확한 시장정보를 제공하기 위해서는 이를 보완할 필요가 있다. 그러나 거래빈도가 낮은 문제를 해결하기 위해 지역을 통합하여 지수를 발표하는 경우 지수를 평활화(smoothing) 시킴으로서 투자의 위험을 지나치게 낮게 평가하여 부동산 투자의 왜곡을 불러올 수 있다.

현재의 반복매매방법에 의한 가격지수는 시간

효과를 더미변수로서 추정한 후 지수를 산정하기 때문에 장래 주택가격 수준을 예측할 수 없다. 정부가 주택시장에 효율적으로 대처하기 위해서는 사후적인 수습책보다는 사전적인 예방책으로 가격지수를 활용하여야 하며 이를 위해 예측이 가능한 지수개발이 필요하다.

본 연구의 목적은 거래빈도가 낮거나 또는 거래가 없는 경우에 대해 실거래가격지수를 작성하는 방법을 개발하는데 있다. 구체적으로 거래쌍이 낮은 하위시장을 대상으로 상태공간모형을 활용한 실거래가격지수를 개발하고, 이를 기존의 반복매매방법에 의한 실거래가격지수와 비교분석하고자 한다. 이를 위해 2006년 1월부터 2013년 8월까지 거래된 실거래가 자료와 한국감정원이 구축한 개별부동산에 대한 특성자료들을 사용하여 반복매매 쌍을 구축한 다음 칼만필터(Kalman filter)와 EM(Expectation and Maximization) 알고리즘을 이용하여 실거래가격지수를 작성한다.

본 논문의 구성은 다음과 같다. 2장에서는 선행연구 검토를 하고, 3장에서는 분석을 위해 사용한 모형에 대해 설명한다. 4장에서는 실증결과를 해석하고 마지막장에서 결론 및 한계점을 기술한다.

## II. 선행연구 고찰

부동산가격지수 작성방법은 헤도닉가격지수, 반복매매지수, 비모수(nonparametric) 가격지수 등 다양하다(Bailey, Muth and Nourse, 1963; Case and Schiller, 1989; Case and Quigley, 1991; Meese and Wallace, 1991; Clapp and Giaccotto, 1992; 한국감정원, 2007). 본 논문에서는 하위시장

(sub-market), 특히, 거래빈도가 낮은 시장에서의 부동산 가격지수 작성방법에 대한 선행연구를 중점적으로 고찰한다.

하위시장은 전형적으로 거래빈도가 낮은 시장이다. 이러한 시장에서 가격의 움직임들을 어떻게 측정할 것인가는 매우 중요하다. 특히, 거래빈도가 낮은 지역 시장의 가격지수에 대한 관심은 시장분석가에게 높을 수 있다. 예를 들어, 특정 지역에 대한 신규주택 개발의 경우 가격지수에 대한 정보는 주거용 부동산 개발업자들에게 중요하다.

기존의 실거래 자료에 기초한 가격지수 작성 방법은 시간더미변수를 사용하여 지수를 작성한다. 시간더미변수들의 추정치 값은 해당 시점에서의 가격지수를 나타낸다. 따라서 최소한 각 시기마다 하나이상의 실거래자료가 필요하다. 거래건수가 적고 거래된 자료의 이상치(outlier)는 가격지수에 크게 영향을 미쳐 지수의 변동성을 증가시킨다. 이러한 문제를 해결하기 위한 방법은 지수작성 기간 혹은 지역을 통합하여 지수를 작성하는 것이다. 예를 들어 월별 지수를 분기별(quarterly) 또는 연간(annual) 지수로 작성하는 방안이다. 그러나 시기를 통합하거나 지역을 통합하여 작성된 지수는 변동성을 낮게 추정함으로써 포트폴리오 구성에 왜곡된 정보를 제공할 수 있다(Geltner, 1991, 1993).

시간더미변수에 의한 지수작성 문제에 대해 Schwann(1998)은 과거 거래된 실거래 자료와 현재의 실거래자료를 통합하여 가격지수의 기반이 되는 비교거래 자료의 집합을 늘릴 수 있다고 보았다. 시계열 자료의 특성을 이용하여 거래가 없거나 낮은 경우에도 안정적인 지수를 작성할 수 있음을 실증적으로 보였다. Quan and Quigley(1971)

는 실거래 자료들에 실제 가격과 오차가 추가되어 있기 때문에 시장가격은 과거와 현재 시장정보에 대한 가중치로 나타났다. 시계열 분석 방법에 의한 가격지수 작성은 기존 더미변수에 의한 가격지수 작성에 비해 모형에서 추정하여야 하는 모수들이 적어 시장을 세분화하여도 지수를 추정할 수 있는 장점이 있다.

주택가격지수에 대한 국내 선행연구로, 남수현·이장우(2009)는 2006년 1월부터 2007년 6월 기간의 서울 개포동 아파트 실거래가 자료를 활용하여 Case-Shiller 지수와 Laspeyres 지수를 작성하고 다른 지수와 비교하여 부동산 선물거래를 위한 가격지수를 개발하였다. 유정석(2011)은 적시성 및 신뢰성 있는 주택가격지수를 활용한 파생금융상품 거래를 도입하기 위해서는 국토해양부가 사용하는 Baily et al. (1963)의 동일가중 반복매매모형보다는 Shiller(1991)가 제안한 총자산의 가격 변동률을 가정한 가치가중 반복매매모형으로 지수를 산정하는 것이 더 적합하다고 주장하였다.

### III. 모형

Kuo(1997)가 제시한 일반적인 반복매매모형을 다음 식으로 표현하자.

$$y_{it} = \mu_i + (l_{it} - 1)\gamma + \beta_t + \alpha_{it} + \epsilon_{it} \quad (1)$$

$$\alpha_{i,t+1} = \rho\alpha_{it} + \nu_{it} \quad (2)$$

여기서  $i = 1, \dots, M$  이고  $t = 1, \dots, T$ 이다.  $y_{it}$ 는  $t$ 시점에서  $i$ 주택의 매매가격의 로그값이며,  $\mu_i$

는 시간에 대해 불변인 주택  $i$ 의 효과를 나타낸다.  $l_{it}$ 는  $t$ 시점까지 거래된 횟수를 의미하며,  $\gamma$ 와  $\beta_t$ 는 각각 주택투자 수익률과 가격지수를 나타낸다. 오차항  $\epsilon_{it}$ 와  $\nu_{it}$ 는 각각  $N(0, \sigma^2)$ 와  $N(0, q, \sigma^2)$  분포를 하고 이들은 서로 독립적이라고 가정한다. 오차항은 지속성(persistence), 또는 추세(trend)를 가지는  $\alpha_{it}$ 와 지속성이 없는  $\epsilon_{it}$ 로 나누며 통상적으로  $\alpha_{it}$ 는 (2)와 같이 자기회귀(AR) 과정을 가정한다. 만일  $\rho$ 가 0인 경우 식 (1)은 다음과 같이 쓸 수 있다.

$$y_{it} = \mu_i + (l_{it} - 1)\gamma + \beta_t + \epsilon_{it} \quad (3)$$

위 식 (3)을 동일한 주택에 대한 거래쌍을 차분의 형태로 쓰면 다음과 같다.

$$y_{it} - y_{is} = \gamma + \beta_t - \beta_s + \epsilon_{it} - \epsilon_{is} \quad (4)$$

이를 행렬로 표현하면,

$$\tilde{y} = \iota\gamma + \tilde{X}\beta + \tilde{\epsilon} \equiv A\delta + \tilde{\epsilon} \quad (5)$$

여기서  $\tilde{y}$ 는  $N-M$  차원 벡터( $N$ 은 총 거래건수,  $M$ 은 총 주택수)의 수익률에 대한 로그값이며,  $\iota$ 는  $N-M$ 개의 1로 이루어진 벡터를 말한다.  $\tilde{X}$ 는  $(N-M) \times (T-1)$  행렬로  $-1, 1, 0$ 의 값을 원소로 갖는다. 쌍으로 이루어진 거래에 대해 처음 거래된 월은  $-1$ , 마지막으로 거래된 월은 1의 값을 갖으며, 그 이외의 월은 0의 값을 갖는다. 상수항과의 다중공선성을 피하기 위해 기준이 되는 월(본 연구의 경우 2006년 1월)은 모형에서 제외하였다. 행렬  $A$ 는  $[\iota \tilde{X}]$ 이며,  $\delta$ 는

$A$ 에 대응하는  $T \times 1$ 인 모수벡터, 즉  $\delta = (\gamma \beta')'$ 이다. 지수작성을 위해  $\beta_1$ 은 0으로 설정하였다. 오차항  $\tilde{\epsilon} \sim N(0, \sigma^2\Omega)$ 라 가정한다. 개별 주택들 간에는 상관관계가 없다고 가정할 경우 오차의 공분산 행렬  $\Omega$ 은 블록대각행렬(block diagonal matrix) 형태를 갖는다.

식 (5) 오차항의 분산-공분산 행렬이 개별 관측치마다 다른 이분산성(heteroschedasticity)을 갖기 때문에 일반화된 최소자승(Generalized Least Squares, GLS) 추정방법에 의해  $\hat{\delta}_{GLS}$ 와  $\hat{\sigma}_{GLS}^2$ 를 각각 다음과 같이 추정한다.

$$\hat{\delta}_{GLS} = (A' \Omega^{-1} A)^{-1} A' \Omega^{-1} \tilde{y} \quad (6)$$

$$\hat{\sigma}_{GLS}^2 = \frac{(\tilde{y} - A\hat{\delta})' \Omega^{-1} (\tilde{y} - A\hat{\delta})}{N - M - T} \quad (7)$$

개별주택효과  $\mu_i$ 는 측정된  $\hat{\gamma}$ 와  $\hat{\beta}_t$ 를 사용하여 식(3)을 이용하여 구한다.

$\alpha_{it}$ 가 AR(1) 확률과정을 따르는 경우 위의 식 (4)는 다음과 같이 쓸 수 있다.

$$y_{it} - y_{is} = \gamma + \beta_t - \beta_s + \alpha_{it} - \alpha_{is} + \epsilon_{it} - \epsilon_{is} \quad (8)$$

위 식을 행렬로 표시하면,

$$\tilde{y} = \iota\gamma + \tilde{X}\beta + \tilde{\alpha} + \tilde{\epsilon} \quad (9)$$

이며, 오차항은  $\tilde{\alpha} + \tilde{\epsilon} \sim N(0, \sigma^2\Omega(q, \rho))$  분포를 따른다고 가정한다.

서로 다른 주택 간에는 상관관계가 존재하지

않기 때문에 이 경우에도 공분산행렬  $\Omega$ 은 블록 대각행렬이다<sup>2)</sup>.

식 (9)의 우도함수(likelihood function)는 다음과 같다.

$$\ell(\tilde{y}) = -\frac{N-M}{2} \log(2\pi\sigma^2) - \frac{1}{2} \log|\Omega| - \frac{1}{2\sigma^2} (\tilde{y} - A\delta)' \Omega^{-1} (\tilde{y} - A\delta) \quad (10)$$

위 우도함수를 최대화하도록 최우추정량을 구한다.

반복매매모형에 의한 가격지수는 시간에 대한 더미변수에 대한  $\beta_t$ 를 추정하는 다음 작성하는 방법으로 다음과 같은 문제점이 있다. 첫째, 거래가 없는 시기에는 지수를 산정할 수 없다. 둘째, 비록 거래가 발생할지라도 거래건수가 적을 경우 작성된 지수의 변동성이 커진다. 이는 주택시장의 특성상 거래가 빈번하게 이루어지지 못하며, 높은 가격에 거래되기 때문이다<sup>3)</sup>. 주택시장을 하위시장으로 세분하거나 지수공표주기를 짧게 잡을수록 거래건수는 작아진다. 세 번째 문제는 더미변수를 통해 지수를 작성할 경우 장래 가격지수를 예측할 수 없다. 정부가 더미변수에 의한 지수에 의존하여 정책을 발표할 경우 주택문제를 사전에 판단하여 대응하는 것이 아니라 문제가 발생한 후 사후 처방적으로 주택시장에 개입할 수 밖에 없게 된다.

Goetzmann (1992)는 반복매매모형에서 시간효과를 기존의 더미변수 접근방법이 아닌 특정 확

률과정(stochastic process)을 따른다고 보았다. 즉, 시간에 따라 주택가격 수준(level)과 증가율(slope)은 동시에 변동한다고 보고 다음과 같이 국지선형추세(local linear trend, LLT) 모형을 제시하였다. 즉,

$$\beta_{t+1} = \beta_t + \kappa_t + \zeta_t \quad (11)$$

$$\kappa_{t+1} = \kappa_t + \xi_t \quad (12)$$

여기서,  $\zeta_t \sim N(0, q_\zeta\sigma^2)$ ,  $\xi_t \sim N(0, q_\xi\sigma^2)$  분포를 따른다고 가정하고 서로 독립적이라 보았다.

식 (9)를 상태공간모형으로 쓰면 다음과 같다.

$$y_t = Z_t x_t + w_t \quad (13)$$

$$x_{t+1} = \Phi x_t + R\eta_t \quad (14)$$

식 (13)은 관측치와 상태와의 관계를 나타내는 관측방정식, 또는 측정방정식이라 부르며, 식 (14)는 상태(state)의 움직임을 나타내는 전이방정식 또는 상태방정식이라 한다. 위의 상태공간형태로 나타난 모형에 대해서 칼만필터(Kalman filter)와 EM(Estimation-Maximization) 알고리즘을 사용하여 모형을 추정한다<sup>4)</sup>.

2) 실제 거래된 주택들 간에는 공간적으로 자기상관되어 있을 수 있다. 이에 대해서는 Hwang and Quigley (2011) 참조.  
 3) Schwann (1998)은 이러한 특성을 가지는 시장을 thin market으로 정의하며, 주식시장은 이와 대조적인 hot market으로 보았다.  
 4) 칼만필터와 EM에 대해서는 Shumway and Stoffer (1982)과 Watson and Stock(1983) 참조.

#### IV. 실증분석

본 논문에서는 서울시 종로구와 강남구를 대상으로 2006년 1월부터 2013년 9월까지의 실거래가 자료를 사용하여 가격지수를 작성하였다. 종로구를 선택한 것은 아파트 거래건수가 적어 거래가 발생하지 않는 시기가 존재하기 때문에 거래빈도가 적을 경우 지수의 변동성 문제를 보일 수 있기 때문이다. 강남구를 선택한 것은 그동안 각종 부동산 정책을 수립할 때 가장 많은 관심을 두었던 지역일 뿐만 아니라 강남구의 가격변화가 주변지역 주택시장에 미치는 영향이 크기 때문이다.

가격지수는 전체 가격지수와 건물면적 85㎡를 기준으로 소형과 중대형으로 구분하여 지수를 작성하였다. 면적 기준으로 시장을 세분한 것은 거래건수가 적어서 생기는 문제를 예시적으로 보이고자 함이다.

지수작성은 반복매매쌍을 기준으로 하였다. 이는 실거래가 공개시 입주연도, 층수, 면적에 대한 정보만 공개하고 있어 특성이격을 기초로 한 지수작성은 어렵기 때문이다. 반복매매쌍을 만들기 위해서는 최소 2번이상 반복적인 거래가 관측되어야 한다. 동일한 주택의 반복매매쌍을 구하기 위해서는 장기간의 자료가 요구된다. 우리나라의 실거래자료의 공개가 2006년에 시작된 점을 고려하면 장기간의 자료는 현재로서는 어렵다. 본 논문에서는 국내 아파트의 경우, 단독주택과 달리 동일 단지 동일 평형의 경우 표준화된 주택특성을 유지하고 있다고 보고 동일주택 가정을 통해 반복매매쌍을 구하였다.

동일주택의 가정은 소재지(단지), 단지내 동 위치, 평형, 층을 기준으로 하였다. <표 1>은 동

일주택 가정에 따른 분류를 보여주고 있으며, <표 2>는 각 기준별 주택수와 거래쌍을 보여준다.

원자료 검토 과정에서 지분거래와 중복거래, 이상치(outlier) 등을 삭제시켰다. 동일주택가정은 거래빈도가 낮은 경우의 지수작성에 비중을 두었기 때문에 기준 3을 적용시켰다. 분석에 사용된 자료는 종로구는 총 1,752개, 강남구는 13,033개이다.

<표 1> 주택특성 조합에 따른 동일주택 가정

기준	주택특성 조합
1	소재지 + 평형
2	소재지 + 평형 + 동
3	소재지 + 평형 + 동 + 층

<표 2> 각 기준별 반복매매건수

거래쌍	종로구(4,381)			강남구(35,402)		
	기준1	기준2	기준3	기준1	기준2	기준3
1	184	253	1,193	1,123	1,719	10,170
2	81	88	467	386	601	3,867
3	55	68	198	240	377	1,799
4	37	44	112	155	259	869
5	26	37	65	120	222	484
6	25	29	30	83	180	304
7	20	19	25	73	140	178
8	16	21	16	58	112	118
9	5	11	11	44	115	58
10회 이상	82 (3,129)	96 (2,813)	24 (305)	684 (29,698)	1,061 (25,213)	137 (1,669)
합계	531	666	2,141	2,966	4,786	17,984

주: ( )안은 거래건수임

<표 3>은 각 구별 상태공간모형 추정결과를 보여주고 있다. 추정결과 대부분의 지역에서  $q_{\xi}$ 가 통계적으로 유의성이 없어 주택가격지수는 부

동(浮動) 임의보행(random walk with drift) 특성을 보이고 있다.

<표 3> 상태공간모형 추정결과

구분	전체	소형	중대형	
	추정치	추정치	추정치	
종로구	$\sigma^2$	0.085 (4.419)***	0.079 (5.952)***	0.177 (8.622)***
	$q_{\zeta}$	0.096 (1.397)	0.082 (1.754)*	0.101 (1.647)*
	$q_{\xi}$	0.000 (0.145)	0.000 (0.314)	0.000 (0.043)
	$\gamma$	0.016 (1.907)*	0.020 (2.754)***	-0.008 (0.240)
	우도값	-254.33	-260.89	-137.55
	아파트수	835	655	180
	거래건수	1,752	1,514	238
강남구	$\sigma^2$	0.096 (1.465)	0.123 (3.073)***	0.158 (4.256)***
	$q_{\zeta}$	0.095 (0.657)	0.052 (1.245)	0.068 (1.437)
	$q_{\xi}$	0.000 (0.427)	0.000 (0.524)	0.000 (0.071)
	$\gamma$	0.051 (3.035)***	0.046 (3.261)***	-0.013 (0.508)
	우도값	-263.869	-262.481	-216.246
	아파트수	6,905	4,959	1,946
	거래건수	14,033	11,005	3,028

다음으로, 모형추정 결과를 이용하여  $t-1$  시점까지의 정보, 그리고 실제로 관측된 개별 주택들의 가격정보를 이용하여 실거래 주택가격지수를 산정하였다.

<그림 1>은 종로구의 아파트 매매가격지수를 보여주고 있다. 전체 아파트의 경우 매 시기마다 반복매매된 거래쌍이 존재하기 때문에 지수작성은 문제가 없다. 하지만 중대형 아파트의 경우 전체 분석기간 중에 거래된 쌍이 238건에 불과하

다. 3개의 기간에는 거래된 자료가 존재하지 않아 지수를 작성할 수 없다(그림 1의 우측 하단 참조). 지수 작성된 기간에 대해서도 변동성이 크게 나타났다. 예를 들어, 2011년 5월 주택가격지수는 130에서 2011년 6월에는 192로 상승하였다가 2011년 7월에 다시 136으로 하락하여 현실성이 떨어진다. 본 연구에서 제안한 상태공간모형을 사용할 경우 주택가격지수는 2011년 5월 137에서 2011년 6월에는 152, 2011년 7월에 149로 기존방법에 비해 변동성이 매우 낮은 안정적인 지수를 보였다. 특히, 거래가 이루어지지 않은 2010년 6월과 7월, 2011년 9월의 가격지수는 각각 126.9, 127.4, 149.2로 추정가능하였다.

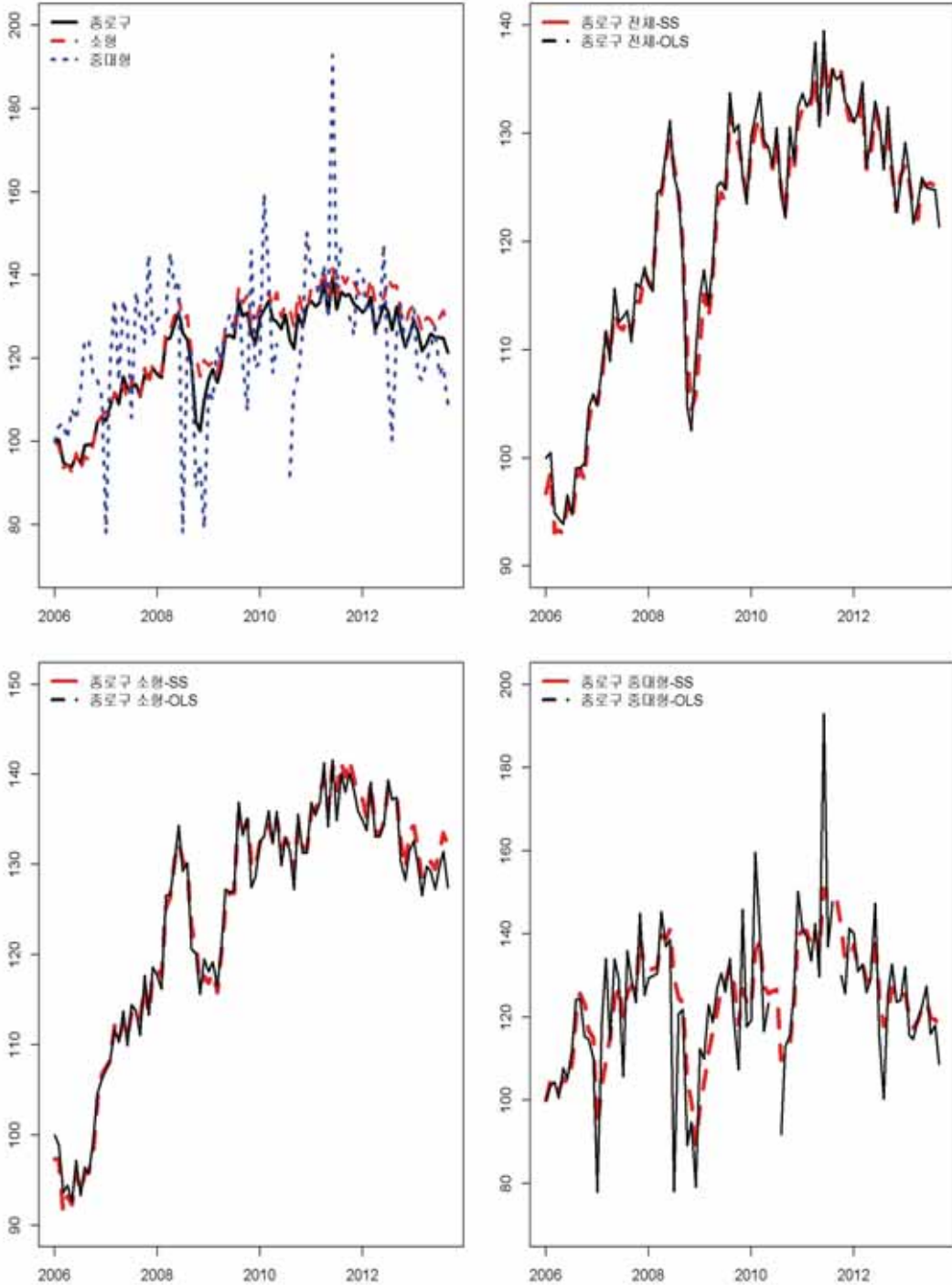
<그림 2>는 강남구에 대한 아파트 매매가격지수를 보여주고 있다. 강남구는 종로구에 비해 아파트 재고량이 많고 상대적으로 거래량이 많아 주택규모별 하위시장으로 세분하여도 지수작성에 문제가 없다. 하지만 지역을 세분화하여 하위시장을 구분할 경우 거래빈도가 낮아 지수작성의 문제가 발생함을 알 수 있다.

<그림 3>은 강남구 내 삼성동에 대한 아파트 실거래 주택가격지수를 작성한 결과 거래빈도가 적어짐에 따라 지수의 변동성이 높아지는 것을 확인할 수 있다.

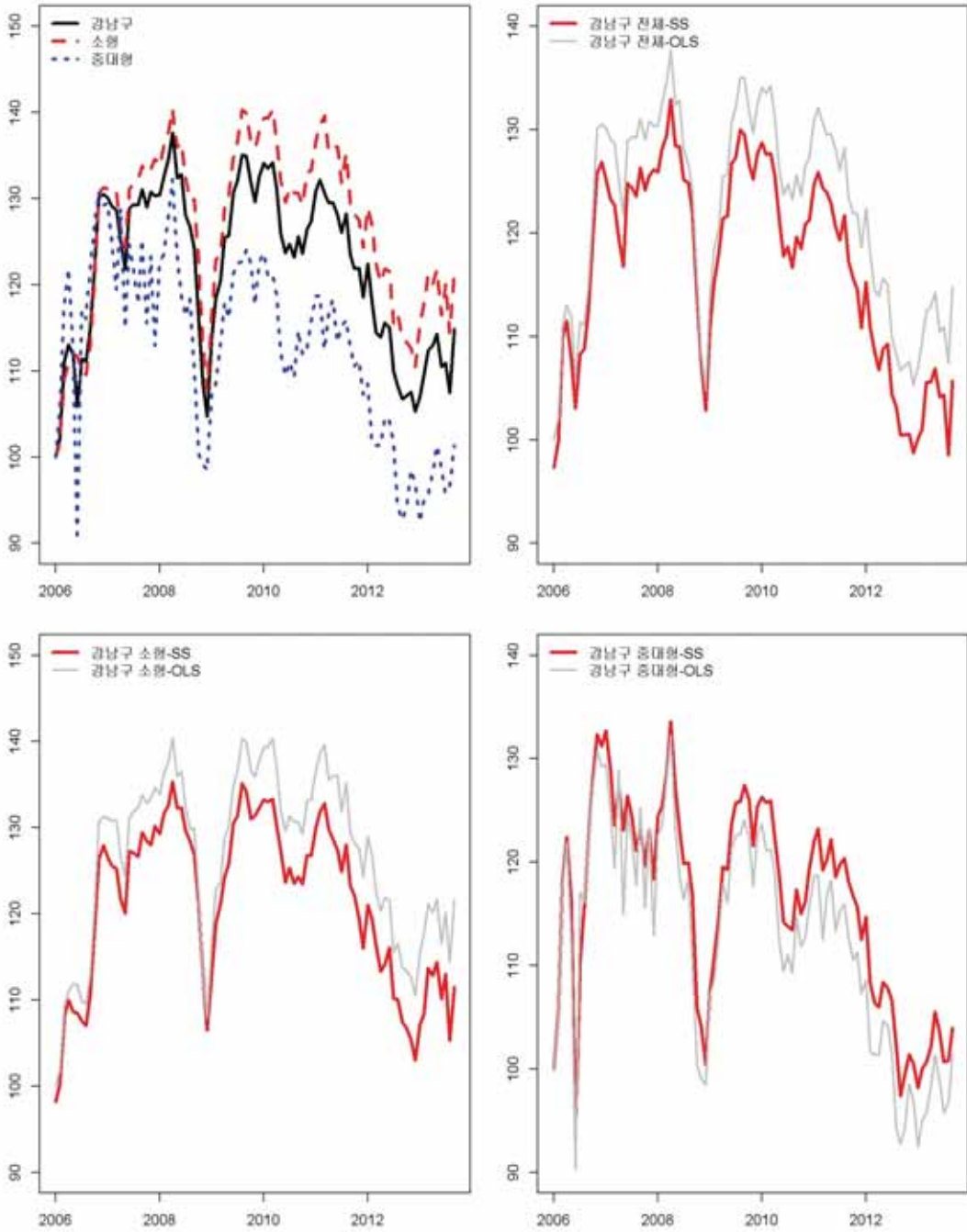
강남구의 전체와 주택유형별 실거래 주택가격지수 추정결과를 보면 OLS에 의한 지수와 상태공간모형에 의한 지수는 주택유형에 따라 차이가 나타나고 있다. 거래빈도가 상대적으로 많은 전체와 소형 아파트의 경우 대부분의 기간동안 OLS 지수에 비해 상태공간모형에 의한 지수가 낮게 나타나고 있는 반면 중대형은 차이가 없는 것으로 나타났다. 이는 두 가지로 해석이 가능하다.



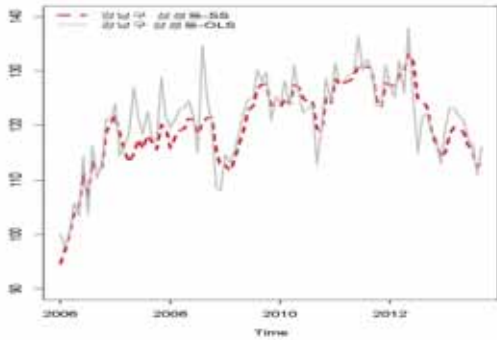
〈그림 1〉 종로구 아파트 매매가격지수 추정



〈그림 2〉 강남구 아파트 매매가격지수 추정



(그림 3) 강남구 삼성동 아파트매매가격지수 추정



첫째, 거래빈도가 많을 경우 동일주택 각각이 거래된 시점 간 차이가 작아 OLS에 의한 추정에서 수익률이 상대적으로 과다추정된 것으로 볼 수 있다<sup>5)</sup>. 또 다른 해석은 개별주택들의 거래쌍 간의 계열상관이 존재하지만 이를 모형에서 고려하지 못할 경우 생기는 편의(biased)로 볼 수 있다.

## V. 결론

본 논문은 거래빈도가 낮은 시장에서의 안정적인 실거래 주택가격지수 작성방안에 대해 연구하였다. 2006년 1월부터 2013년 9월까지 아파트 실거래가 자료를 사용하여 서울시 종로구와 강남구 지역에 대한 아파트 실거래가격지수를 추정한 결과 시사점은 다음과 같다.

첫째, 하위시장으로 주택시장을 세분할 경우 거래건수가 적어 기존의 지수작성방법으로는 지수를 작성할 수 없거나 작성하더라도 지수의 변

동성이 높아지는 것을 확인할 수 있었다. 지수의 변동성은 투자의 위험을 나타내는 주요 지표이다. 주택 하위시장에서 지수의 변동성이 높아지는 것은 투자의 위험성이 높아진 이유라기보다는 지수작성 방법의 문제이다. 정부는 서울시를 4개의 권역으로 묶어서 지수를 발표하고 있지만 이는 가격지수를 평활화(smoothing)하여 발표함으로써 시장참여자에게 부동산투자에 대한 위험(risk)을 상대적으로 낮게 평가하게 하는 또 다른 문제를 제기하고 있다.

둘째, 지수를 작성하기 위해 시간에 대한 효과를 단순히 더미변수로 설정하는 것은 지수작성상의 문제뿐만 아니라 지수를 활용하는 측면에서 주택가격수준을 사전에 예측하지 못하는 한계가 있다. 본 연구에서 제안한 상태공간모형은 실거래된 자료가 공통적으로 가지는 미관측 가격지수에 대한 상태(state)와 상태의 변화를 확률과정으로 모형화함으로써 예측 가능한 주택가격지수를 개발하는데 이론적 틀을 제공할 것이다<sup>6)</sup>. 이러한 지수 개발은 사후처방적 대책이 아닌 사전예방적 대책으로 정책의 실효성을 높이는데 기여할 것이다.

셋째, 세분화된 하위주택시장에 대해 상태공간모형을 통한 지수를 작성한 결과 더미변수방법에 의한 지수보다 변동성이 낮게 추정되어 안정적인 지수임을 확인할 수 있었다. 특히, 거래가 발생하지 않은 시기에 대해서도 안정적인 지수를 작성할 수 있음을 보여주었다.

본 논문은 개별주택의 반복매매쌍들 간 계열

- 5) 미국의 Case-Shiller 지수는 이러한 문제를 감안하기 위해 거래기간의 역수를 설명변수로 추가하여 보정하였다.
- 6) Robert Shiller 교수는 중장기적으로 주택가격은 예측할 수 있다는 이론적 틀을 제시하였으며, S&P의 미국 대도시 지역에 대한 주택가격지수인 Case-Shiller Index를 개발하는데 공로를 인정받아 2013년 노벨경제학상을 수상하였다.

상관을 고려하지 못하였고, 반복매매쌍에서 거래 시점의 차이가 짧을수록 수익률이 높아질 수 있는 가능성을 모형에서 고려하지 못하였다. 또한 서로 다른 주택들간에는 상호 독립적인 가정을 전제로 하였다. 그러나 실제 지리적으로 인접한 주택들 간에는 상호 의존적인 공간자기상관이 존재하기 때문에 이에 대한 보완이 필요하다. 이에 대해서는 추후 연구로 남긴다.

## 참고문헌

1. 남수현 · 이장우, “한국 부동산 선물가격지수 개발에 관한 연구”, 금융공학연구, 한국금융공학회, 제8권 3호, 2009, pp.75-103
2. 유정석, “부동산가격지수를 활용한 파생금융상품 개발방안 연구”, 부동산학보, 한국부동산학회, 제45집, 2011, pp.127-143
3. 한국감정원, 「실거래가격지수 개발에 관한 연구」, 서울: 한국감정원, 2007
4. 통계청, 「2013년 가계금융·복지조사」 통계청, 2013
5. Bailey, M. R. Muth, and H. Nourse, “A Regression Method for Real Estate Price Index Construction,” Journal of the American Statistical Association, Vol.58, 1963, pp. 933-942
6. Case, B., and J. M. Quigley, “The Dynamics of Real Estate Prices,” Review of Economics and Statistics Vol.73, 1991, pp.50-58
7. Case, K., and R. Shiller, “The Efficiency of the Market for Single-Family Homes,” American Economic Review, Vol.79 No.1, 1989, pp.125-137
8. 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, Vol.9 No.2, 1992, pp.137-164
9. Geltner, D., “Smoothing in Appraisal-Based Returns,” Journal of Real Estate Finance and Economics, Vol.4 No.3, 1991, pp.327-345
10. Geltner, D., “Temporal Aggregation in Real Estate Return Indices,” Journal of the American Real Estate and Urban Economics Association, Vol.21 No.2, 1993, pp.141-166
11. Goetzmann, W., “The Accuracy of Real Estate Indices: Repeat Sale Estimators,” Journal of Estate Finance and Economics, Vol.5 No.1, 1992, pp.5-14
12. Hwang, Min and John M. Quigley, “Housing Price Dynamics in Time and Space: Predictability, Liquidity and Investor Returns,” Journal of Real Estate Finance and Economics, Vol.41, 2010, pp.3-23
13. Kuo, C., “A Bayesian Approach to the Construction and Comparison of Alternative House Price Indices,” Journal of Real Estate Finance and Economics, Vol.14, 1997, pp.113-132
14. Meese, R., and N. E. Wallace, “Nonparametric Estimation of Dynamic Hedonic Price Models and the Construction of Residential Housing Price Indexes.” Journal of the American Real Estate and Urban Economics Association,

Vol.19, 1991, pp.308-332

15. Quan, D., and J. Quigley,. "Price Formation and the Appraisal Function in Real Estate Markets," Journal of Real Estate Finance and Economics, Vol.4, 1991, pp.175-190
16. Schwann, G.M, "A Real Estate Price Index for Thin Markets." Journal of Real Estate Finance and Economics, Vol.16 No.3, 1998, pp.269-287
17. Shiller, R.J., "Arithmetic Repeat Sales Price Estimators," Journal of Housing Economics, Vol.1, 1991, pp.110-126
18. Shumway, R. H. and Stoer, D. S., "An approach to time series smoothing and forecasting using the EM algorithm." Journal of Time Series Analysis, Vol.3, 1982, pp.253-264
19. Watson, M. W. and Engle, R. F., "Alternative algorithms for the estimation of dynamic factor. Mimic and varying coefficient regression models." Journal of Econometrics, Vol.23, 1983, pp.385-400