

매칭 방법을 이용한 대구 아파트 실거래 가격지수 측정¹⁾

The Construction of Housing Price Indices Using Matching Approach : The Case of Apartments in Daegu

정태훈 Jung Taehun¹⁾, 김병조 Kim Byung Jo²⁾, 정창도 Jung Changdo³⁾

Abstract

Housing price index using matching approach is introduced to complement the problems of inaccuracy in hedonic price index and omission of data in Repeat Sales price index and this study marks the first attempt to adopt matching approach for regional analysis. Using actual transaction price data of apartments in Daegu, the results indicate that Matched Repeat Sales index follows the similar pattern as KAB (Korea Appraisal Board) index, KB index or Repeat Sales index. While the hedonic index tends to overestimate the housing price, McMillen index tends to underestimate it. The housing price index presented in this study is expected to contribute to providing additional information to predict housing market.

Keywords: Actual Transaction Price, Matching Approach, Hedonic, Repeat Sales

1. 서론

2013년 현재 가구의 평균 소득 대비 평균 주택가격 비율은 약 5배 정도이고 우리나라 가계의 자산보유 중 실물자산, 특히 부동산이 차지하는 비중이 아주 높은 것으로 알려져 있다.⁵⁾ 이러한 현실에서 부동산 가격을 대상으로 실시하는 정책이 가계의 행동 및 생

활에 지대한 영향을 주는 것은 자명하다. 따라서 부동산 가격과 관련한 경제정책을 실시함에 있어서 가능한 한 정확한 정보를 바탕으로 부동산의 전체적인 움직임을 파악해야 한다.

부동산의 움직임을 정확하게 파악하는 것이 상당히 어려움에도 불구하고 우리나라는 1986년부터 한국주택은행(현재 KB국민은행)에서 전국의 주택가

1) 본 논문은 한국은행의 '2013년도 지역전문가와의 공동연구: 대구 아파트 실거래가를 이용한 지수측정과 시사점' 보고서의 일부를 발췌 및 수정하여 작성한 것임. 본 논문은 2012학년도 경북대학교 학술연구비에 의하여 연구되었음.

2) 경북대학교 경제통상학부 부교수(제1차) | Associate Prof., School of Economics and Trade, Kyungpook National Univ. | Primary Author | taehunjung@gmail.com

3) 한국은행 지역통합실 지역경제팀 | Office of Branch Coordination, The Bank of Korea | bjkim@bok.or.kr

4) 경북대학교 대학원 경제학과 박사과정 수료 | Ph. D. Candidate, Dept. of Economics, Kyungpook National Univ. | metalnom@hanmail.net

5) KB은행은 가구를 소득별로 5분위로 나누고, 주택도 매매가격별로 5분위를 나눈 뒤, 주택가격 3분위에 속하는 주택의 평균매매가격을 소득 3분위에 속하는 가구의 평균소득으로 나눈 PIR(Price to Income Ratio) 값을 제공함. 이에 의하면 2008년 이후 PIR값은 계속해서 5.0 전후에서 움직이고 있다는 사실을 확인할 수 있음.

격 동향을 매월 발표하기 시작했다. 여기에서 제공되는 주택가격지수(KB지수)는 부동산 정책을 수립하고 실시하는 데에 많은 기여를 하였지만, 부동산업체로부터 제공 받는 호가정보에 의존하고 있어서 주택의 실거래가격과는 필연적으로 괴리가 발생할 수밖에 없다는 문제가 제기되어 왔다. 박연우, 방두완(2011)과 김명준 외(2008)에서도 밝히고 있듯이, 실거래가격을 이용하지 않고 부동산업체의 호가정보를 이용한 감정평가 가격을 이용할 경우 실제 거래가격보다 평활화(smoothing)하는 현상이 발생한다.⁶⁾

우리나라는 다행히도 2006년부터 건설교통부(현재 국토교통부)가 부동산의 실거래가격을 공시하고 있다.⁷⁾ 고질적으로 이어져오던 부동산 투기, 이중계약서 작성 등 잘못된 부동산 계약 관행을 없애고 투명한 거래 질서 확립을 목적으로 2006년 1월 1일부터 부동산거래신고제도를 도입하였다.⁸⁾ 이로 인하여 부동산 매매계약이 체결된 날로부터 60일 이내에 관할 관공서에 거래신고를 하게 됨으로써 부동산의 실거래가격에 대한 정보를 입수할 수 있게 되었다.

이렇게 공개되는 부동산 실거래가격 중 본 논문에서는 아파트의 실거래가격 정보를 이용한다. 아파트에 초점을 두는 이유는 크게 두 가지다. 첫째는 우리나라는 주택 가운데 아파트가 차지하는 비율이 상당히 높기 때문이다. 5년마다 실시하는 통계청의 인구주택총조사에 의하면, 주택 중 아파트가 차지하는 비율은 1980년에 7% 수준에 불과하던 것이 1990년

에는 22.7%, 2000년에는 47.7%, 2010년에는 59%로 계속적으로 증가하였다. 본 논문의 분석대상인 대구지역의 경우에도 주택 중 아파트가 차지하는 비율은 1985년에 25.3% 정도였으나, 1990년에는 32.5%, 2000년에는 56.7%, 2010년에는 67.2%로 지속적으로 증가하였다.

아파트 실거래가격에 관심을 가지는 두 번째 이유는 아파트가 다른 형태의 주택과는 달리, 어느 정도 표준화된 규격을 가지고 있기 때문이다. 단독주택의 경우 주택마다 천차만별의 특성을 가지고 있기 때문에 특정 주택의 실거래가 움직임을 분석에 이용한다는 것은 현실적으로 불가능한 측면이 있는 반면, 아파트는 완전히 동일하지는 않더라도 공통된 특성을 가진 경우가 많기 때문에 동일 부동산으로 간주하는 것이 가능하다는 편리한 점을 가지고 있다.

부동산의 실거래가를 이용하여 가격지수를 산정하는 방법으로는 헤도닉 가격지수(Hedonic Price Index)와 반복매매지수(Repeat Sales Price Index)가 많이 이용되어 왔다.⁹⁾ 이 두 가지 방식이 보편화된 이유는 헤도닉 가격지수는 부동산의 개별 특성을 직접적으로 통제한다는 점에서 이해하기가 쉬운 장점이 있고, 반복매매지수는 부동산의 개별 특성이 변하지 않는다는 가정하에 가격정보만을 이용하는 편리함이 있기 때문이다.¹⁰⁾

헤도닉 가격지수 산정방식에는 기본적으로 모형 설정의 어려움이 존재함과 동시에 특정 변수를 수집

6) 감정평가를 기반으로 할 경우 발생하는 평활화의 문제는 Barkham and Geltner(1996), Geltner(1991), Chau, MacGregor, and Schwann(2001), Geltner, MacGregor, and Schwann(2003) 등에서 다루고 있음.

7) 국토교통부는 1994년 12월까지 건설부와 교통부로 나뉘어 있었음. 그러다가 1994년 12월 23일 이 두 부처의 업무를 통합하여 건설교통부가 설립됨. 건설교통부는 2008년 2월 29일 해양수산부와 통합되어 국토해양부로 바뀌었다가, 2013년 3월 23일 해양 관련 업무를 해양수산부로 이관하고, 현재의 국토교통부로 설립되었음. 본 논문에서는 혼란을 피하기 위하여 일률적으로 '국토교통부'라는 명칭을 사용하기로 함.

8) 전세 및 월세의 실거래가격은 2012년 1월 1일부터 국토교통부를 통해서 공개되기 시작했음.

9) 감정평가가치에 의존하거나 부동산 중개업자에 의해서 제공되는 호가정보에 의존하는 것이 아니라, 실거래가를 이용하여 부동산 매매가격지수를 산정할 경우, 기존의 표본에 근거하는 라스파이레스(Laspeyres) 방식은 실질적으로 사용하기가 어려워짐. 그 이유는, 실거래가는 시세와는 달리 개별 아파트에 대한 가격정보가 매기 관측될 수 없기 때문임. 그래서 실거래가를 이용한 지수 산정에는 특정 표본에 대한 매기의 관측치가 없어도 되는 헤도닉 가격지수와 반복매매지수가 많이 이용되어 왔음.

하는 데에 많은 시간 및 비용이 소요되는 물리적인 어려움이 있다. 특히, 이러한 물리적인 어려움 때문에 특정 변수에 대한 정보가 부정확하게 수집될 가능성이 크다. 실제로 우리나라의 경우, 이러한 어려움 때문에 반복매매지수 방법을 채택하여 실거래 가격지수를 매월 공시하고 있다.

하지만 반복매매지수는 2회 이상 거래가 이루어진 부동산만을 대상으로 하기 때문에 표본의 수가 상당히 줄어든다는 단점을 가지고 있다. 물론, 동일 아파트로 간주하는 가정에 따라 관측 빈도가 줄어드는 문제를 완화할 수 있지만, 이에도 한계가 있어서 완전한 해결책이 되지 못한다. 실제로 한국감정원에서는 아파트 단지명이 같고 면적이 같은 아파트 중에서 1층, 최고층, 그 외의 층으로 세 분류를 한 뒤, 같은 분류에 들어가면 동일한 아파트로 간주하였다. 하지만, 이러한 방식에 대한 정당성을 부여할 만한 근거가 없을 뿐만 아니라, 이렇게 하더라도 표본의 수가 줄어드는 것을 완전히 해결하지는 못한다. 이처럼 표본의 수가 줄어드는 것을 해결하는 한 가지 방법으로 McMillen(2012)은 매칭(matching)을 이용한 방법을 제시하였다. 구체적인 방법은 다음 장에서 서술되겠지만, 간략히 이야기하면 2회 이상 거래되지 않아서 매매가격을 비교할 수 없는 아파트에 대해서 이와 가장 유사한 성질을 가진 아파트를 매칭한 후 지수를 산정하는 방법으로, 이 방법을 이용하면 일반적으로 표본수는 증가한다는 것이다.¹⁰⁾

본 논문은 기존에 사용되어 오고 있는 헤도닉 가격지수와 반복매매지수 이외에 매칭을 이용한 지수

등 다양한 지수 측정을 시도하고 이를 통해 주택시장에 대한 정확한 정보를 제공하려는 것이 목적이다. 헤도닉 가격지수가 갖는 자료의 부정확성, 반복매매지수가 갖는 자료의 누락을 보완하기 위하여 매칭을 이용한 두 가지 방법을 제시하였다. 특히, 이러한 매칭의 방법은 본 논문에서처럼 대구라는 지역을 대상으로 할 때에 더욱 더 중요성이 증가될 것이다. 서울 혹은 전국을 분석 대상으로 할 경우에는 일반적인 반복매매법을 이용하더라도 도시의 규모로 보아 표본수가 줄어들어도 통계적 분석에 치명적 결함이 되지 않을 정도의 표본은 구성될 것이다. 하지만 대구 정도의 규모, 더 나아가 대구시의 각 구 단위로 분석할 때에는 기본적인 표본수가 적은 상태에서 출발하기 때문에 표본수의 감소는 아주 중요한 문제가 된다. 이러한 점에서 본 논문은 그 의의를 찾을 수 있다.

부동산 가격지수와 관련된 선행연구에서 가장 많은 부분을 차지하는 것은 헤도닉 가격지수와 반복매매지수다. Rosen(1974)에 의하여 도입된 헤도닉 가격지수는 Mark and Goldberg(1984), Palmquist(1980), Thibodeau(1989), Shimizu et al.(2010) 등에 의해 이용되었다. 헤도닉 가격지수를 이용한 국내 문헌으로는 박현수(2009), 이용만(2007) 등이 있다. 반복매매지수의 경우 Bailey, Muth, and Nourse(1963)에 의해 개발되었고, Case and Shiller(1989)에 의해서 발전되었다. 현재에도 반복매매지수를 발전시킨 방법을 이용하여 S&P/Case-Shiller 주택가격지수를 발표하고 있다. 반복매매지수를 이용한 문헌으로는 Chau, Wong, and Yiu(2005), McMillen(2003), 이창무, 김

10) 헤도닉 가격지수의 경우 부동산의 개별 특성을 직접적으로 이용하는 방식인 반면, 반복매매지수의 경우 부동산의 개별 특성을 직접 이용하는 것이 아니라 이러한 개별 특성이 변하지 않는다는 조건을 이용하고 있기 때문에 헤도닉 가격지수에 비해 간접적인 통제방식이라 할 수 있음(McMillen 2012).

11) 매칭을 이용할 때에 언제를 기준으로 하느냐에 따라서 표본수가 증가할 수도 있고, 오히려 감소할 수도 있음. 논문에서도 2011년 3월을 기준으로 할 때에는 표본수가 증가하지만, 2006년 1월을 기준으로 하면 표본수가 감소한다는 사실을 확인할 수 있음. 시가고를 대상으로 하는 McMillen(2012)에서는 매칭으로 표본을 구성할 경우 그 수가 감소하지만, 싱가포르를 대상으로 하는 Deng, McMillen, and Sing(2012)에서는 표본수가 감소했음.

용경, 배익민(2007) 등이 있다.

본 논문에서처럼 매칭을 이용한 문헌은 아직 미미한 수준이다. McMillen(2012)은 미국 시카고의 주택 가격을 대상으로, Deng, McMillen, and Sing(2012)은 싱가포르의 주택 가격을 대상으로 매칭을 이용한 가격지수 산정과 이에 대한 논의를 전개하고 있다. 국내 문헌의 경우 매칭을 이용한 부동산 매매지수에 대한 논의를 전개하는 것은 저자가 알고 있는 한 본 논문이 최초다.

그리고 본 논문의 분석처럼 많은 지수를 생성한 후 비교한 선행연구로 Meese and Wallace(1997)과 Shimizu, Nishimura, and Watanabe(2010) 등이 있다. 특히, Shimizu, Nishimura, and Watanabe(2010)는 본 논문의 분석처럼 도쿄의 주택가격을 이용하여 여러 가지 부동산 지수를 비교하고 있지만, 매칭을 이용한 방법을 사용하지 않고, 기본적인 헤도닉 가격지수와 반복매매지수, 그리고 이 두 지수의 변형된 형태(Case-Shiller adjustment to the repeat sales index, age-adjustment to the repeat sales index, structural-change adjustment to the hedonic index)를 이용하고 있다는 점이 본 논문과는 다르다.¹²⁾

II. 주택가격지수 측정 방법

이 장에서는 실거래가를 이용한 지수 산정 방법으로 많이 이용되는 헤도닉 가격지수와 반복매매지수, 그리고 분석에 사용되는 여러 특성에 대한 정보의 부정확성 및 표본수 감소 등의 문제점을 극복하는 한 가지 방법으로 매칭을 이용한 반복매매지수에 대해서 논의하고자 한다.

1. 헤도닉 방법

Rosen(1974)에 의해 도입된 헤도닉 모형은 부동산의 가격에 영향을 미치는 여러 특성을 고려한 회귀모형이다. 일반적으로 부동산의 가격 수준을 결정짓는 요소로 구조적 특성을 나타내는 변수, 입지의 특성을 나타내는 변수, 그리고 그 외의 변수 등이 고려된다. 구조적 특성을 나타내는 변수로는 면적, 층수, 세대수, 방수, 욕실수, 준공연수 등이 있고, 입지의 특성을 나타내는 변수로는 부동산의 위치, 가장 가까운 지하철역까지의 거리, 도심 혹은 부심까지의 거리 등이 있다. 그 외의 변수로는 주변 환경, 건설사의 순위 등을 들 수 있다.

아파트의 매매가격과 이에 영향을 미치는 요소들과의 관계를 추정하기 위한 회귀식을 다음과 같이 설정한다.

$$\ln P_{it} = \sum_{i=1}^T \delta_i D_{it} + \beta X_{it} + u_{it} \quad \langle \text{식 1} \rangle$$

〈식 1〉에서 $\ln P_{it}$ 는 t 년에 거래된 아파트 i 의 매매 가격에 대한 로그값을 나타내고, X_{it} 는 아파트 i 의 가격에 영향을 미치리라고 생각되는 특성들을 나타낸다. 전술한 바와 같이 X_{it} 에는 구조적 특성을 나타내는 변수, 입지적 특성을 나타내는 변수, 그리고 그 외의 변수 등이 포함된다. D_{it} 는 시간더미를 나타내고, u_{it} 는 오차항을 나타낸다. 이 회귀식을 통하여 산출되는 시간더미의 추정계수로부터 헤도닉 가격지수를 다음과 같은 방법으로 계산해낸다. 먼저 기준연도(본 논문에서는 2006년 1월)의 값인 δ_0 을 0으로 표준화(normalize)하고, 그다음에 나머지 시간더미의 추정계수들인 $\hat{\delta}_2, \hat{\delta}_3, \dots, \hat{\delta}_T$ 들을 회귀식으로부터 구한다.

12) Shimizu, Nishimura, and Watanabe(2010)에 의하면, 헤도닉지수는 2002년 초반에 부동산 가격이 최저점에 달한 것으로 나타난 반면, 반복매매지수는 지속적으로 하락하다가 2004년 봄이 되어서야 상승하기 시작한 것으로 나타났음. 이처럼 부동산 지수에 따라서 시장의 상황을 나타내는 정도가 다른 만큼, 다양한 지수들을 제공하여 정책 실시의 적기를 놓치지 않도록 하는 것이 중요함.

마지막으로, 이렇게 하여 구해진 일련의 값으로부터 $\exp(0), \exp(\hat{\delta}_2), \exp(\hat{\delta}_3), \dots, \exp(\hat{\delta}_T)$ 와 같이 헤도닉 가격지수를 계산한다. 본 논문에서는 기준연도의 가격지수를 100으로 한 뒤, 매 기간의 가격지수를 조정하여 도출하였다.

2. 반복매매지수법

Bailey, Muth, and Nourse(1963)과 Case and Shiller(1989)에 의해서 개발되고 발전되어온 반복매매지수는 2회 이상 거래된 아파트를 대상으로 아파트 개별의 특성이 변하지 않는다는 가정하에서 만들어지는 지수다. 반복매매지수의 산정은 헤도닉 모형에서 사용하였던 <식 1>로부터 출발한다. <식 1>에서 특정 두 시점(t 와 s)에서 거래가 이루어졌다고 가정할 때, 이 두 시점 간의 아파트 가격의 변화는 다음과 같이 측정할 수 있다.

$$\ln P_{it} - \ln P_{is} = (\delta_t D_{it} - \delta_s D_{is}) + (\beta X_{it} - \beta X_{is}) + (u_{it} - u_{is}) \quad \langle \text{식 2} \rangle$$

<식 2>에서 아파트의 개별적인 특성의 변화가 없다면, 즉, X_{it} 와 X_{is} 가 같다면, 우변의 두 번째 항은 제거되고, 다음과 같은 간단한 형태의 추정식이 도출된다.

$$\Delta_{t,s} \ln P_i = D_i' \delta + \nu_{its} \quad \langle \text{식 3} \rangle$$

여기서 $\Delta_{t,s} \ln P_i = \ln P_{it} - \ln P_{is}$ 을 의미하고, $\nu_{its} = u_{it} - u_{is}$ 을 의미한다. D_i 는 시간더미 변수들의 벡터를 의미하는데, 첫 번째 매매가 일어나는 시점에는 -1을, 두 번째 매매가 일어나는 시점에는 1을, 매매가 일어나지 않은 시점에는 0의 값을 가진다.

<식 3>을 추정된 뒤 산출되는 $\hat{\delta}$ 값들을 가지고 지수를 계산해내는 작업은 헤도닉 가격지수 산정 때와 동일하다. 즉, 먼저 기준연도의 값인 $\hat{\delta}$ 을 0으로 표준화하고, 나머지 $\hat{\delta}$ 값들을 가지고, $\exp(0), \exp(\hat{\delta}_2), \exp(\hat{\delta}_3), \dots, \exp(\hat{\delta}_T)$ 을 계산한 뒤, 일률적으로 100을 곱하면 되는 것이다.

이와 같은 방법으로 구해진 반복매매지수는 두 번의 매매가 시간적으로 많이 떨어져 있을 경우 아파트 가격의 변동 폭이 더욱 더 증가할 수 있다는 의미에서의 이분산성(heteroskedasticity) 문제가 발생한다. Case and Shiller(1989)는 이러한 이분산성 문제를 극복하기 위한 한 가지 방법으로 다음과 같이 가정하였다.

$$E(u_{it} - u_{is})^2 = \xi_1(t-s) \quad \langle \text{식 4} \rangle$$

이를 바탕으로 Case and Shiller(1989)는 다음과 같은 절차로 반복매매지수를 수정하였다. 먼저 <식 3>을 OLS로 추정된 뒤 잔차를 구한다. 이렇게 구한 잔차의 제곱을 <식 4>처럼 상수항(ξ_0)과 두 매매 시점의 차이($t-s$)에 대해서 회귀하여 추정계수인 $\hat{\xi}_0$ 와 $\hat{\xi}_1$ 을 구한다. 마지막으로 $\sqrt{\hat{\xi}_0 + \hat{\xi}_1(t-s)}$ 을 가중(weight)으로 하는 GLS로 <식 3>을 다시 추정하여 값들을 구한다. 이 값들로부터 $\exp(0), \exp(\hat{\delta}_2), \exp(\hat{\delta}_3), \dots, \exp(\hat{\delta}_T)$ 을 구한 뒤, 100을 곱하여 반복매매지수를 계산한다.

3. 매칭을 이용한 방법

1) 매칭을 이용한 반복매매지수법

(Matched Repeat Sales Approach)

조사기간 중 2회 이상 거래되지 않아서 매매가격을 비교할 수 없는 아파트에 대해서 이와 가장 유사

한 성질을 가진 아파트를 매칭한다. 매칭은 Rubin (1974)과 Rosenbaum and Rubin (1983)이 제시한 성향 점수 매칭법(P propensity Score Matching: PSM)을 이용한다. 이렇게 매칭된 데이터를 이용하여 반복매매 지수를 구하는 방법이다. 성향점수 매칭법은 다음과 같은 알고리즘으로 실시하는데, 이 방법은 McMillen (2012)과 Deng, McMillen, and Sing (2012)도 이용하고 있다.

- ① 2006년 2월에서 2013년 4월까지의 기간 중 특정 시점인 p월에 거래가 이루어졌다면 종속변수의 값을 1로 부여한다.
- ② p월에 거래가 이루어지지 않았지만, 기준으로 삼은 월(예를 들어, 2006년 1월)에 팔렸던 아파트라면 종속변수의 값을 0으로 부여한다.
- ③ 설명변수는 헤도닉 분석에서 사용한 설명변수를 이용하여, 위의 종속변수에 대해 프로빗 추정을 한다.
- ④ 프로빗으로부터 추정된 성향점수(propensity score)를 기준으로, 각 p월에 종속변수가 1인 아파트(즉, 거래가 이루어져 매매가격을 가진)와 종속변수가 0인(거래가격이 존재하지 않는) 아파트의 성향점수를 비교하여 가장 가까운 것들을 비복원추출하여 매칭한다. 여기서 비복원추출 매칭이라 함은 성향점수가 가장 가까운 짝을 매칭시키고, 매칭이 이루어진 것을 제외한 나머지 중에서 다른 짝을 매칭시키는 것을 말한다.¹³⁾
- ⑤ 매칭된 짝에서 거래가 없는 아파트의 거래가격에 거래가 있는 아파트의 거래가격을 부여한다.

2) McMillen 방법

헤도닉 가격지수와 반복매매지수를 구할 때에 아파트의 특성과 관련된 많은 설명변수를 이용하는 추정식으로부터 출발하지만, 결국 관심이 있는 것은 시간터미의 추정계수들뿐이다. 그런데 McMillen (2012)과 Deng, McMillen, and Sing (2012)이 지적하고 있듯이, 이 시간터미의 추정계수들은 ‘처리효과(treatment effects)’의 형태를 띠는 것이다. 다음과 같은 두 기간(1기와 2기) 모형을 생각해 보자.

$$\ln P_{it} = \delta_1 D_{i1} + \delta_2 D_{i2} + \beta X_{it} + u_{it} \quad \langle \text{식 5} \rangle$$

〈식 5〉는 다음과 같이 변형될 수 있다.

$$\ln P_{it} = \delta_1 + (\delta_2 - \delta_1) D_{i2} + \beta X_{it} + u_{it} \quad \langle \text{식 6} \rangle$$

〈식 6〉은 1기에는 $\ln P_{i1} = \delta_1 + \beta X_{i1} + u_{i1}$ 이고, 2기에는 $\ln P_{i2} = \delta_2 + \beta X_{i2} + u_{i2}$ 이라는 사실에 포함하고 있다. 이러한 식의 형태는 처리효과를 다룰 때에 전형적으로 나타난다. 〈식 6〉의 경우 단순한 차분을 통해서 다음과 같은 식으로 변형될 수 있다.

$$\ln P_{it} - \ln P_{is} = (\delta_2 - \delta_1) D_{i2} + u_{it} - u_{is} \quad \langle \text{식 7} \rangle$$

〈식 7〉에서 알 수 있는 것은, 결국 1기와 2기의 매매가격의 차이는 다음과 같은 평균처리효과(Average Treatment Effect: ATE)의 형태로 나타난다는 것이다.

$$ATE = \frac{1}{n_2} \sum_{i=1}^n D_{i2} E[\ln P_{it_2} - \ln P_{it_1}] \quad \langle \text{식 8} \rangle$$

13) 매칭에는 통계패키지 STATA의 psmatch2를 이용하였다. 이 모듈(module)은 성향점수 매칭법을 이용한 연구에서 많이 사용됨. 매칭 절차에 대한 부연은 하면 다음과 같음. 가령, 2006년 1월을 매칭의 기준으로 삼을 경우, 2006년 2월의 매칭은 2006년 1월에는 거래되었으나, 2006년 2월에는 거래되지 않은 아파트에 0을, 2006년 2월에도 거래된 아파트에 1을 부여한 후 헤도닉 모형에서 사용된 설명변수를 이용하여 프로빗 추정을 한 뒤, 성향점수를 이용하여 매칭을 실시함. 이러한 프로빗 추정을 2013년 4월까지 계속 반복해야 하기 때문에 총 87번의 프로빗 추정이 필요함.

여기서 n_2 는 2기에 거래된 아파트의 수를 나타낸다. 1기를 지수를 산정하기 위한 기준연도로 설정한다면, 2기뿐만 아니라 그 이후의 시점에 대해서도 다음과 같이 동일한 방식을 적용할 수 있다.

$$ATE(t_j) = \frac{1}{n_j} \sum_{i=1}^{n_j} D_{i,j} E[\ln P_{it_j} - \ln P_{it_1}] \quad \langle \text{식 9} \rangle$$

여기서 매기 동일한 아파트가 거래되지 않았을 경우에는 해당 아파트는 <식 9>의 계산에서 제외시켜야 하는데 그러면 데이터 손실이 발생하게 된다. 이에 매칭을 이용한 방법이 이용되는데, 그 방법은 성향점수 매칭법을 이용한다. 성향점수 매칭법의 알고리즘은 앞에서 설명한 바와 같다.

III. 분석자료

본 연구에서 사용된 자료는 국토교통부에 의해 공시되는 아파트 실거래가격으로 2006년 1월부터 2013년 4월까지 대구에서 거래된 아파트의 매매가격을 이용하였다. 각 구별 거래량과 비중은 <표 1>과 같다. 아파트가 많은 달서구, 북구, 수성구의 거래량이 많았으며, 도심이라 주거지보다는 상업 및 관공서가 많은 중구는 1.7% 정도에 불과했다.

국토교통부의 자료들을 통해서 알 수 있는 변수들은 매매 실거래가격 이외에 거래된 아파트의 전용면적, 거래 층, 건축연도, 그리고 거래연월 등이다. 개별 아파트의 특성 중 국토교통부 자료에서 알 수 없는 것은 '부동산114'를 통하여 수집하였다. '부동산114'에서 수집한 개별 아파트의 특성 변수들은 아파트의

총 층수, 방수, 욕실수, 현관구조, 해당 동 세대수, 건설회사, 단지 동수, 단지 세대수, 총 주차대수, 난방방식, 난방연료, 교통 상황 등이었다. 하지만 이러한 많은 특성 변수값이 모든 아파트에서 나타나는 것은 아니었다. 각 특성별 구체적인 표본수에 대해서는 다음 <표 2>와 같다.

IV. 모형 및 실증분석

1. 헤도닉 방법

헤도닉 가격지수를 구하기 위한 회귀모형은 앞의 <식 1>과 같은 형태로 설정한다. 하지만, <표 2>에서 알 수 있듯이 특성 변수에 대한 정보를 모든 아파트가 다 가지고 있는 것이 아니기 때문에 분석하는 과정에서 상당한 수의 표본이 누락될 수밖에 없다. 그래서 어떤 변수들을 분석에 사용할 것인지에 대한 선택이 필요하다. 본 논문에서는 <표 2>의 변수 중에서 전용면적, 거래층, 방수, 욕실수, 단지 세대수, 경과연수, 난방방식, 난방연료, 현관구조의 정보를 사용하였다.¹⁴⁾ 총 주차대수의 경우 7만 1,800여 개의 데이터 손실이 발생하기 때문에 이에 대한 정보를 사용하지 않기로 결정하였다. 그리고 총 주차대수는 단지 세대수와 상관관계가 높은 변수이기 때문에 단지 세대수가 설명변수로 들어가면 다중공선성(multicollinearity) 문제에서 자유로워진다는 이점도 있어서 분석에서 제외해도 별 영향이 없다고 판단하였다. 단지 동수의 경우에도 단지 세대수와 높은 상관관계 때문에 분석에 사용하지 않았다. 교통환경의 경우 데이터 누

표 1 _ 행정구별 거래량과 비중

구분	달서구	북구	수성구	동구	달성군	서구	남구	중구	합계
거래량(건)	62,147	51,171	41,190	25,361	19,152	7,847	6,846	3,691	217,405
비중(%)	28.59	23.54	18.95	11.67	8.81	3.61	3.15	1.70	100.00

표 2 _ 아파트 특성 변수와 표본수

아파트 특성 변수	표본수	평균	표준편차	최솟값	최댓값
전용면적(m ²)	217,405	80,3369	28,8132	18,6384	296,6
거래층	217,405	8,9209	6,0782	1	56
경과연수	217,405	10,8921	7,5226	-3	40
총 층수	201,996	-	-	-	-
방수	194,592	3,0394	0,6139	1	7
욕실수	194,547	1,5942	0,5112	1	5
현관구조	192,018	-	-	-	-
건설회사	192,599	-	-	-	-
단지 동수	201,940	7,3429	4,2081	1	21
단지 세대수	201,934	735,902	431,307	14	2160
총 주차대수	145,674	848,368	603,116	10	4202
난방방식	193,892	-	-	-	-
난방연료	192,068	-	-	-	-
교통환경	159,142	-	-	-	-

주: 1) 경과연수는 해당 연도에서 건축연도를 뺀 값임. 경과연수에 음의 값이 나오는 것은 거래가 먼저 일어나고 난 뒤, 건축이 완료되는 것을 말하는데, 이에 해당하는 아파트는 11개임. 남구에 1개, 달서구에 7개, 북구에 1개, 서구에 1개, 수성구에 1개 아파트가 여기에 해당함.
 2) 현관구조는 계단식, 복도식, 복합식, 중앙코어형, 타워식 등으로 분류되어 있음.
 3) 난방방식은 개별난방, 중앙난방, 지역난방 등으로 분류되어 있음.
 4) 난방연료는 도시가스, 기름, LPG, 열병합, 폐열 등으로 분류되어 있음.

락 여부를 떠나서 정보의 불규칙성¹⁴⁾ 때문에 ‘부동산 114’가 제공하는 데이터는 사용하지 않았다. 하지만 교통 환경은 아파트 매매가격에 영향을 미치는 아주 중요한 요소라고 판단되기 때문에 다음 두 가지 방식으로 해결하기로 하였다. 첫째, 모든 아파트에 대한 주소 정보를 이용하였다.

둘째, 지리정보시스템(Geographic Information System: GIS)을 이용하여 각 아파트에서 대구시의

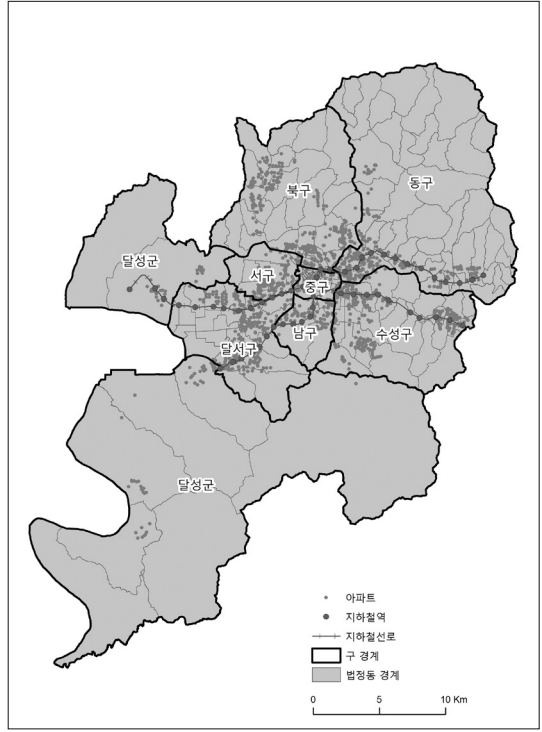
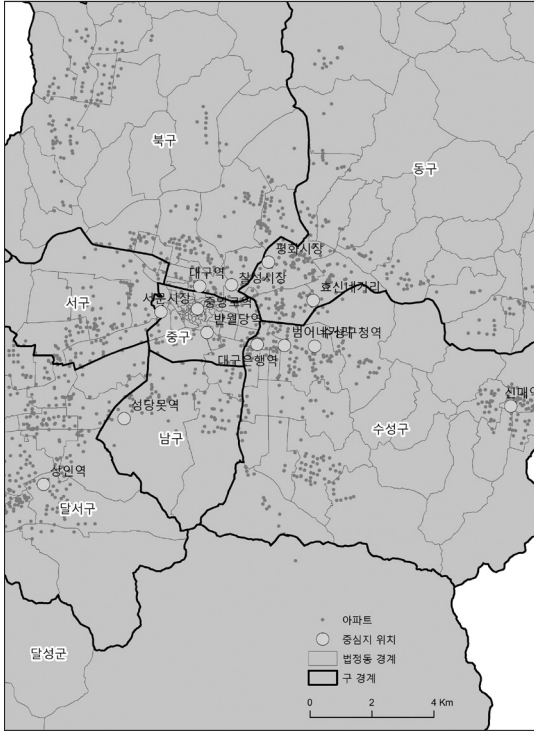
중심지까지 거리 정보와 가장 가까운 지하철역까지의 거리 정보를 이용하였다. 먼저 지리학에서 지정하고 있는 대구시의 중심지 및 지하철에 대한 정보는 <그림 1>과 같다.

대구시의 중심지는 2011년 국토교통부 부동산 공시지가의 표준지 지가를 바탕으로 최고 지가 20% 임계치를 이용하여 설정하였다. 그 결과 13개의 중심지가 선정되었다. 구체적으로는 기존 도심인 동성로

14) 헤도닉 모형에서 설명변수의 선정은 선행연구에서 사용하고 있는 일반적인 특성변수들을 사용하였음. 헤도닉 모형의 이론적 근거를 제시한 Rosen(1974)까지 거슬러 올라가면, 대부분의 연구는 Rosen(1974)의 주장에 따라 부동산의 구조적인 특성, 지리적인 접근성, 주거지 근린의 특성 등을 부동산의 가격에 영향을 미치는 주요한 변수로 채택하고 있음. 가령, 국내 문헌에서 대기질과 아파트 가격의 관계를 분석한 최종일, 심성훈(2002)의 경우 아파트의 구조변수로 평수, 준공연도, 난방방식, 세대수 등을, 접근도 변수로 지하철 및 쇼핑센터까지의 거리를, 근린변수로 공원 유무, 초등학교 유무 등을 독립변수로 사용하고 있음. 부산의 아파트를 대상으로 연구한 김기호, 이성우(1998)와 김성우, 정건섭(2010)도 아파트의 특성과 지역 특성을 설명변수로 사용하고 있음. 김명준 외(2008)도 아파트의 구조적, 지리적 특성 외에 본 논문에서처럼 건설회사에 대한 특성도 함께 고려하고 있음. 그 외 박현수(2009), 이용만(2007) 등도 본 논문과 비슷한 아파트의 특성 변수들을 설명변수로 사용하고 있음.

15) ‘부동산114’에서 제공하는 개별 아파트 교통환경의 경우 가장 가까운 역까지 도보로 걸리는 시간 혹은 거리 등으로 다양하게 표현되는 경우가 많아 계량분석에는 적합하지 않음.

그림 1 _ 대구시의 도심과 부심, 지하철에 대한 정보



일대(대구역, 중앙로역, 반월당역)와 중구 서문시장, 북구 칠성시장, 수성구 범어네거리, 효신네거리, 지하철 2호선인 신매역, 대구은행역, 수성구청역, 동구 평화시장, 지하철 1호선 성당못역, 상인역 등이었고, 각 아파트에서 이들 13개 중심지와와의 거리를 계산하여 분석해 이용하였다. 다음으로 각 아파트에서 가장 가까운 지하철역까지의 거리를 계산하였는데, 아파트의 주소 데이터를 좌표화(geocoding)해서 가장 가까운 지하철역을 할당한 후, 직선거리로 계산하였다.

구체적인 회귀모형은 <식 1>의 형태로 설정하고, 통제변수에 해당하는 X_{it} 에는 전용면적의 로그값, 경과연수, 방수 및 욕실수의 로그값, 단지 세대수의 로그값, 현관구조 더미변수, 도급순위 더미변수, 난방방식 더미변수, 난방연료 더미변수, 1층 여부 더미변수, 그리고 중심지 및 가장 가까운 지하철역까지의 거리의 로그값, 마지막으로 각 아파트가 소재한 행정

구 더미가 포함되었다. 경과연수는 해당 연도에서 건축연도를 뺀 값으로 계산하였고, 현관구조 더미변수는 계단식일 때 1, 그 외에는 0을 부여하였다. 난방방식 더미변수는 개별난방일 때 1, 그 외에는 0을 부여하였고, 난방연료 더미는 도시가스일 때 1, 그 외에는 0을 부여하였다. 도급순위 더미변수는 국토교통부에서 공개하는 2012년 건설업체 시공능력 평가공시 정보를 이용하였다. 이에 의하면 10위 이내의 건설업체는 현대건설, 삼성물산, 대우건설, GS건설, 포스코건설, 대림산업, 롯데건설, 현대산업개발, SK건설, 두산중공업 순이었다. 이들 건설회사의 아파트면 도급순위 더미변수값을 1, 그렇지 않으면 0을 부여하였다.

<표 3>은 헤도닉 모형의 실증분석 결과를 나타낸다. 대부분의 추정계수는 납득할 수 있는 부호 및 통계적인 유의성을 보여주고 있다. 먼저 전용면적의 경우, 추정계수가 약 0.87이며 이는 전용면적이 1% 증

표 3_ 헤도닉 모형의 실증분석 결과

설명변수	추정계수	설명변수	추정계수
Log_전용면적	0.8741*** (0.0027)	Log_중심지와의 거리 (범어네거리)	-0.0983*** (0.0041)
경과연수	-0.0239*** (0.0000)	Log_중심지와의 거리 (효신네거리)	0.0894*** (0.0022)
Log_방수	0.0907*** (0.0033)	Log_중심지와의 거리 (신매역)	0.0269*** (0.0011)
Log_욕실수	0.0536*** (0.0020)	Log_중심지와의 거리 (대구은행역)	0.0716*** (0.0032)
현관구조 터미 (계단식=1)	0.0947*** (0.0013)	Log_중심지와의 거리 (수성구청역)	-0.1249*** (0.0033)
도급순위 터미 (10위 이내=1)	0.0512*** (0.0014)	Log_중심지와의 거리 (평화시장)	0.0160*** (0.0026)
Log_단지 세대수	0.0273*** (0.0007)	Log_중심지와의 거리 (성당못역)	-0.0586*** (0.0023)
난방방식 터미 (개별난방=1)	-0.0007 (0.0014)	Log_중심지와의 거리 (상인역)	-0.0188*** (0.0011)
난방연료 터미 (도시가스=1)	-0.0340*** (0.0016)	달서구	0.0709*** (0.0039)
1층 여부 터미 (1층=1)	-0.0747*** (0.0016)	달성군	-0.0496*** (0.0044)
Log_가까운 지하철역까지의 거리	-0.0638*** (0.0006)	동구	0.0828*** (0.0041)
Log_중심지와의 거리 (대구역)	0.1787*** (0.0076)	북구	0.2342*** (0.0043)
Log_중심지와의 거리 (중앙로역)	-0.4345*** (0.0162)	서구	0.1274*** (0.0044)
Log_중심지와의 거리 (반월당역)	0.0666*** (0.0056)	수성구	0.2923*** (0.0040)
Log_중심지와의 거리 (서문시장)	0.1128*** (0.0055)	중구	0.0158*** (0.0061)
Log_중심지와의 거리 (칠성시장)	0.0465*** (0.0039)		
관측치 수 = 180,563		R-squared = 0.9059	

주: ***는 1% 유의수준에서 유의함을 나타내고, 괄호 안은 표준오차를 의미함.

가할 때 아파트 가격이 0.87% 정도 증가함을 보여준다. 경과연수의 계수는 음의 값을 가지며, 이는 오래된 아파트일수록 매매가격은 하락함을 보여준다. 김명준 외(2008)의 해석처럼 경과연수의 추정계수가 양의 값을 가진다면, 이는 재건축의 이익에 대한 기대감이 반영된다고 해석할 수 있다. 하지만 대구의 경우 서울처럼 재건축의 기대감은 그다지 크지 않은

것으로 판단된다. 당연한 결과이지만, 방수 및 욕실수의 경우 그 수가 많을수록 아파트 가격은 올라갔다. 현관구조가 계단식일 경우에 역시 아파트 가격은 올라가는데, 이 역시 아파트 구조 면에서 유용성이 상대적으로 낮고 엘리베이터에서 내려서 많이 걸어야 하는 복도식보다는 계단식을 선호하는 현상을 반영하고 있다. 도급순위 10위 이내의 건설업체가 만드는

아파트의 가격이 비싼 경향을 나타냈으며, 단지 세대수가 많을수록 아파트 가격은 상승하는 것으로 나타났다. 이는 단지 세대수가 많을수록 근린 편의시설이 잘 구비되고, 관리비 절감 및 장기적으로 집을 내놓는다거나 재건축에 들어갈 때의 장점 등이 반영된 것으로 판단된다. 난방방식의 경우, 개별난방일수록 아파트 매매가격에 양의 영향을 끼칠 것으로 예상되었으나, 통계적 유의성을 가지지 못한 결과가 나왔다. 그리고 도시가스의 경우에는 예상과는 달리 도시가스가 아파트 매매가격에 음의 영향을 끼치는 것으로 나타났다.¹⁶⁾ 매매된 아파트가 1층인지의 여부를 묻는 더미변수의 추정계수는 음의 값을 가졌다. 아파트가 1층일 경우, 아파트 가격이 평균적으로 약 7.5% 하락하는 것으로 보인다.

아파트 가격에 큰 영향을 미칠 것으로 예상되는 아파트의 입지 조건에 대한 분석도 대체적으로 사회적 통념에 부합하는 결과를 보였다. 먼저 교통환경 중 가장 가까운 지하철역까지의 거리에 대한 추정계수는 음의 값을 나타냈다. 이는 지하철역에서 멀리 떨어져 있는 아파트일수록 가격이 하락함을 의미한다. 중심지와와의 거리가 아파트 가격에 미치는 영향은 13개의 중심지 중 어디를 기준으로 하느냐에 따라 아파트 가격에 미치는 영향이 달라졌는데, 중앙로역, 범어네거리, 수성구청역, 성당못역, 상인역으로부터는 거리가 가까울수록, 대구역, 반월당역, 서문시장, 칠성시장, 효신네거리, 신매역, 대구은행역¹⁷⁾, 평화시장으로부터는 거리가 멀수록 아파트 가격이 올라가는 것으로 나타났다. 특히, 범어네거리와 수성구청역으로부터 가까울수록 아파트 가격이 올라가는 것은 학군 때

문인 것으로 보이고, 서문시장, 칠성시장, 평화시장 등 전통시장으로부터 멀수록 아파트 가격이 올라가는 것은 시장 주변의 교통혼잡 등으로 주거에 불편함이 있기 때문인 것으로 보인다. 다음으로 행정구 더미의 추정계수도 설명이 가능한 결과를 보였다. 수성구라는 위치가 아파트의 가격을 올리는 역할을 한 반면, 달성군이라는 위치가 아파트의 가격을 내리는 역할을 하였다. 수성구에 위치한 아파트의 경우, 수성구에 있다는 자체만으로 가격을 약 30% 가까이 올리는 것으로 나타난 반면, 농촌 지역을 많이 포함하고 있는 달성군은 상대적으로 아파트 가격이 약 5% 정도 낮게 형성되는 것으로 나타났다.

2. 반복매매지수법

반복매매지수 산정을 위해서 필요한 데이터는 2회 이상 거래된 아파트이다. 국토교통부 자료에서는 특정 아파트의 특정 호수가 계속 거래된 것인지의 여부를 알 수 없기 때문에 일본처럼 기업형 주택임대관리 시장이 형성되지 않는 현 상황에서는 완전히 동일한 아파트의 거래 횟수를 구한다는 것은 불가능하다. 그래서 행정구와 아파트단지 및 전용면적이 동일한 상태에서 첫째, 같은 층일 경우에 동일한 아파트로 간주하는 방법, 둘째, 최저층을 제외한 나머지 층을 같은 층으로 간주하는 방법, 셋째, 최저층과 최고층을 제외한 나머지 층을 같은 층으로 간주하는 방법, 이렇게 크게 세 가지 방법을 사용할 수 있다. 이 중 한국감정원이 사용하는 방법은 세 번째인데, 본 논문의 경우에는 최고층의 정보를 구할 수 없기 때문에 첫 번째 방

16) 도시가스는 대구지역을 포함한 광역시의 경우 대부분의 아파트가 선호할 뿐만 아니라 가장 많이 이용하고 있는 난방연료이기 때문에 양의 영향을 끼치거나 혹은 아파트 간 차별성이 거의 없다는 점을 감안하여 아파트 가격에 아무런 영향을 끼치지 않으리라고 예상되었음. 하지만 오히려 음의 영향이 나타났고, 그 원인에 대해서는 앞으로 좀 더 면밀히 살펴보아야 할 것임.

17) 대구은행역에서 멀수록 아파트 가격이 올라가는 이유는 지하철 인근이라는 교통 측면에서의 편리성보다는 인근의 범어네거리 또는 수성구청역 등에 비해 학군 면에서 상대적으로 불리한 점이 더 크게 작용하고 있는 것으로 추정됨.

법과 두 번째 방법이 사용될 수밖에 없다.

그런데 행정구, 단지, 전용면적이 같고, 최저층을 제외한 나머지 층은 동일한 아파트로 간주하는 두 번째 방법의 경우, 동일한 아파트로 간주될 가능성이 첫 번째 방법보다 커지기 때문에 표본수는 많이 줄어든다. 층수까지 같아야만 동일한 아파트로 간주할 때의 약 16% 수준밖에 되지 않았다. 이처럼 최저층을 제외한 나머지 층은 동일한 아파트로 간주하게 되면 그만큼 동일한 아파트가 많아지게 되고 자연스럽게 아파트 가격은 평균값을 사용하게 되므로 평활화(smoothing)된 값을 가지게 되겠지만, 표본수가 급격히 감소하는 문제점이 발생하게 된다. 이러한 점을 감안하여 본 논문에서는 행정구, 단지, 전용면적, 층수 이렇게 4가지 특성이 같을 때에만 동일한 아파트로 간주하는 첫 번째 방법을 이용하기로 한다.

2회 이상 매매가 이루어진 아파트를 대상으로 반복매매법을 위한 회귀식의 설명변수와 종속변수를 생성해야 한다. 가령, 동일한 아파트가 1, 2, 3기에 P_1, P_2, P_3 의 가격으로 3회 거래되었다면, 종속변수는 $\ln(P_2/P_1), \ln(P_3/P_2)$ 로 2개의 관측치가 생기고, $\ln(P_2/P_1)$ 에 대응하는 설명변수로 1기에 -1, 2기에 1을 부여하는 시간 더미변수를, $\ln(P_3/P_2)$ 에 대응하는 설명변수로 2기에 -1, 3기에 1을 부여하는 시간 더미변수를 생성하면 된다. 이런 방식으로 모든 거래에 대해서 적용할 경우 관측치수는 <표 4>와 같다. 이렇게 하여 생성된 관측치수는 13만 6,841개이고 이 중 달서구의 경우 2회 이상 거래된 아파트의 수가 많았기 때문에 관측치수는 4만 2,151개로 다른 구보다 월등히 많았다. 그다음으로 북구, 수성구 순이었고,

관측치 수가 가장 작은 구는 중구로 2,000개에도 미치지 못하였다.

3. 매칭을 이용한 방법

매칭법에서는 동일한 아파트가 존재하지 않는 경우, 성향점수(propensity score)를 이용하여 가장 유사한 아파트를 동일한 아파트로 간주하게 된다. 매칭을 이용할 때, 비교시점에 매매가격을 가지고 있는 표본수에 영향을 받기 때문에 어떤 시점을 기준으로 삼느냐에 따라 표본수가 달라진다. 가령 200개(A_1, \dots, A_{100})의 아파트 매매가격을 가지고 있는 2006년 1월을 기준시점으로 삼았을 때를 생각해 보자. 2006년 2월의 경우 매매가격을 가지고 있는 아파트는 300개($A_1, \dots, A_{150}, C_1, \dots, C_{150}$)이고, 이 중 2006년 1월에도 매매가격을 가지고 있는 아파트는 150개(A_1, \dots, A_{150})였다고 하자. 그러면 2006년 1월에 매매가격을 가지고 있으면서 2006년 2월에 매매가격을 가지지 않는 아파트 50개(A_{151}, \dots, A_{200})에 대해서는 성향점수를 이용하여 이 50개의 아파트와 가장 유사한 아파트를 일대일 매칭한다. 가령 아파트 A_{151} 과 가장 유사한 아파트 C_1 이 결정되었다면, 2006년 2월에 아파트 A_{151} 의 가격 대신에 아파트 C_1 의 가격을 이용하는 것이다. 단, 매칭의 특성상 모든 아파트에 대해서 일대일 매칭이 이루어지는 것은 아니기 때문에 2006년 1월에 매매가격을 가지고 있으면서 2006년 2월에 매매가격을 가지지 않는 아파트 중 매칭이 이루어지는 것은 50개 이하가 된다. 이렇게 하여 2006년 1월의 표본수의 변화는 없지만, 2006년 2월의 표본수는

표 4 _ 반복매매지수 산정을 위한 관측치수

남구	달서구	달성군	동구	북구	서구	수성구	중구	합계
4,021 (2.9)	42,151 (30.8)	11,950 (8.7)	15,028 (11.0)	30,516 (22.3)	4,480 (3.3)	26,764 (19.6)	1,931 (1.4)	136,841 (100.0)

주: 괄호 안의 값은 비중(%).

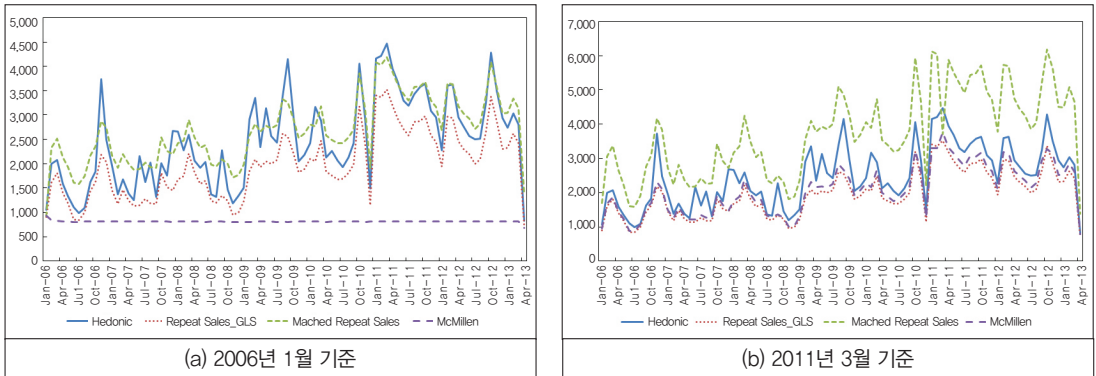
매칭으로 인하여 바뀌는 것이다.¹⁸⁾ 같은 방식을 다른 비교시점에도 적용하여 매칭된 아파트의 가격을 계산해낸다.

매칭된 표본이 완성되면 지수를 작성하는 데에 두 가지 방법을 이용한다. 한 가지는 Matched Repeat Sales 방법이고, 또 다른 하나는 McMillen 방법이다. 이 두 가지 방법은 매칭된 표본을 구성하는 단계까지는 동일하고, 그 이후 단계에서 차이가 난다. Matched Repeat Sales 방법은 매칭된 표본을 가지고 반복매매법을 그대로 이용하는 반면, McMillen 방법은 항상 기준 연월의 표본이 기준이 된다. 앞에서 설명했던 것처럼 가령 200개의 아파트 매매가격을 가지고 있는 2006년 1월을 기준 시점으로 삼고 McMillen 방법을 적용한다고 생각해 보자. McMillen 방법은 항상 기준 시점의 아파트만을 분석 대상으로 하므로, 그 외의 아파트는 표본에서 빠지게 된다. 따라서 2006년 2월 이후의 비교 시점의 아파트를 매칭하여 구한 표본은 매기 최대 200개의 아파트만 표본으로서 존재하게 되

는 것이다. 이러한 방식으로 표본수의 추이를 계산한 것이 <그림 2>다. <그림 2>의 (a)와 (b)는 각각 2006년 1월과 2011년 3월을 기준으로 하였을 때의 표본수의 추이를 나타낸다.

<그림 2>를 보면, 지수작성 방법에 따라 그리고 기준 시점을 언제 잡느냐에 따라 표본수는 상당히 달라진다는 사실을 알 수 있다. 먼저 표본이 시작되는 2006년 1월을 기준으로 한 경우를 보면, 기준 연월의 표본수는 1,060개였다. 헤도닉 방법의 경우, 표본을 전부 사용하므로 1,060개에서 시작한다. 반복매매법, Matched Repeat Sales 방법, McMillen 방법의 경우에는 일단 그룹핑(grouping)을 하기 때문에 2006년 1월의 표본수는 기본적으로 줄어든다. 이 중 반복매매법은 2회 이상 거래된 아파트만 남겨 둔 후, 2006년 1월의 표본수를 헤아려본 결과 901개만 남았다. 매칭을 한 뒤에 반복매매법을 적용하는 Matched Repeat Sales 방법의 경우 2006년 1월의 표본은 925개였다. McMillen 방법의 경우에는 2006년 1월이 항

그림 2 _ 매칭에 의한 표본수의 추이



18) 2006년 2월의 경우, McMillen 방법을 이용할 때와 Matched Repeat Sales 방법을 이용할 때에 표본수가 달라질 수 있음. McMillen 방법을 이용할 때 2006년 2월의 표본수는 최대 200개로 2006년 1월의 표본수가 거의 비슷하지만, Matched Repeat Sales 방법을 이용할 때에는 다른 시기(예를 들어, 2006년 3월)에 거래된 아파트 중 2006년 2월의 아파트(C_1, \dots, C_{150})와 겹치는 부분이 있다면, 이 아파트들의 실거래 가격정보는 이용하기 때문에 표본수는 200개를 초과할 수 있는 것(즉, 2006년 1월과 2월의 반복쌍의 수와 2006년 2월과 3월의 반복쌍의 수가 달라질 수 있는 것)임. McMillen 방법에서는 매칭을 한 후, 매칭이 안 된 데이터는 버려서 데이터가 줄어들 수 있는 반면, 본 논문의 Matched Repeat Sales 방법은 실거래 가격이 존재함에도 불구하고 데이터를 버려야 하는 점을 보완하는 방법을 모색하고 있음.

상 기준인데, 2006년 1월의 1,060개 중 그루핑으로 그 수가 줄어들어 938개가 되었음을 확인할 수 있다.

이렇게 2006년 1월을 시작으로 매기의 표본수를 살펴보면, 헤도닉 방법과 Matched Repeat Sales 방법에서 표본수가 상당히 많고, McMillen 방법에서 표본수가 가장 적음을 알 수 있다. 아파트 개별 특성에 대한 정보가 아주 정확하다면 헤도닉 방법을 사용하는 것이 최선이지만, 물리적으로 정확한 정보를 얻는다는 것이 현실에서는 힘들기 때문에 반복매매법을 이용하게 된다. 하지만 반복매매법의 경우, 데이터의 손실이 크다는 단점이 있다. 그리하여 새로이 제시되는 방법이 매칭을 이용하는 방법인데, 2006년 1월을 기준으로 할 경우 McMillen 방법에서는 표본수가 상당히 누락된다는 사실을 알 수 있다. 따라서 표본을 늘릴 수 있는 것은 Matched Repeat Sales 방법이 될 것이다.

하지만 기준 시점을 선택할 때, 아파트의 거래가 격이 가장 많은 연월을 기준 시점으로 잡는다면, 표본수의 누락이라는 문제점을 극복할 수 있을 것이다. 본 논문의 분석 기간 중 가장 표본수가 많았던 기간은 2011년 3월로 표본수가 4,466개였다. 2011년 3월을 기준 시점으로 각 방법의 표본수를 나타낸 것이 <그림 2>의 (b)이다. 이 그림에서 알 수 있듯이 매칭을 이용한 두 가지 방법이 반복매매법보다는 표본수가 많아진다는 것을 알 수 있다.

그렇다면 표본수가 적은 2006년 1월을 기준으로 삼는 것이 최적인가, 아니면 표본수가 많은 2011년 3월을 기준으로 삼는 것이 최적인가 하는 문제가 남게 된다. McMillen(2012)에 의하면, 어떤 시점이든 매칭의 기준이 될 수 있지만, 거래건수가 많은 시점을 기준으로 삼으면 매칭이 안 되는 표본들이 나타나기 때문에 이로 인한 노이즈(noise)를 감안하면, 거래건수가 적은 시점을 기준으로 삼는 것이 적절하다고 하였다. 본 논문에서는 이러한 점을 고려하여, 매칭을 이

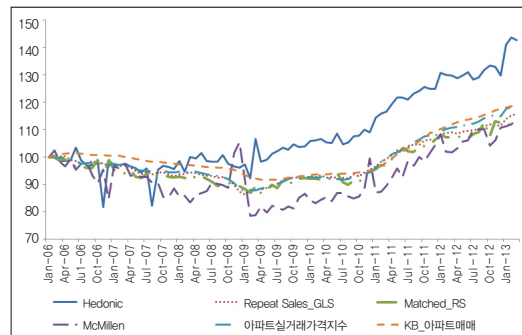
용하여 데이터를 구성할 때, 두 가지 기준을 모두 사용하였다. 즉, McMillen 방법에서는 거래건수가 적은 2006년 1월을 기준으로 하였고, Matched Repeat Sales 방법에서는 거래건수가 많은 2011년 3월을 기준으로 하였다.

V. 지수 산정 및 비교

2006년 1월부터 2013년 4월까지 헤도닉 방법, 반복매매법, Matched Repeat Sales 방법, McMillen 방법을 이용하여 앞에서 시행한 실증 결과를 바탕으로 대구시의 아파트 실거래가 가격지수를 구하였다. <그림 3>은 대구시의 아파트 실거래가 가격지수를 나타낸다.

<그림 3>에서 알 수 있듯이 한국감정원지수(아파트 실거래가지수)와 KB지수, 그리고 본 논문에서 구한 반복매매지수는 거의 비슷한 값을 가진다. 한국감정원지수를 기준으로 보았을 때, KB지수는 약간 과대평가를, 반복매매지수는 약간 과소평가를 하고 있다는 사실을 알 수 있다. KB지수는 부동산중개업소에서 수집되는 호가정보를 기준으로 하고 있기 때문에 실거래가를 이용하는 한국감정원지수보다 약간 높게 나타난 것으로 보이고, 한국감정원지수와 반복매매지수는 동일한 방법을 이용하고 있지만, 동일한 아파트로 간주하는 방식에서 약간의 차이가 존재하

그림 3 _ 대구시 아파트 실거래가 가격지수

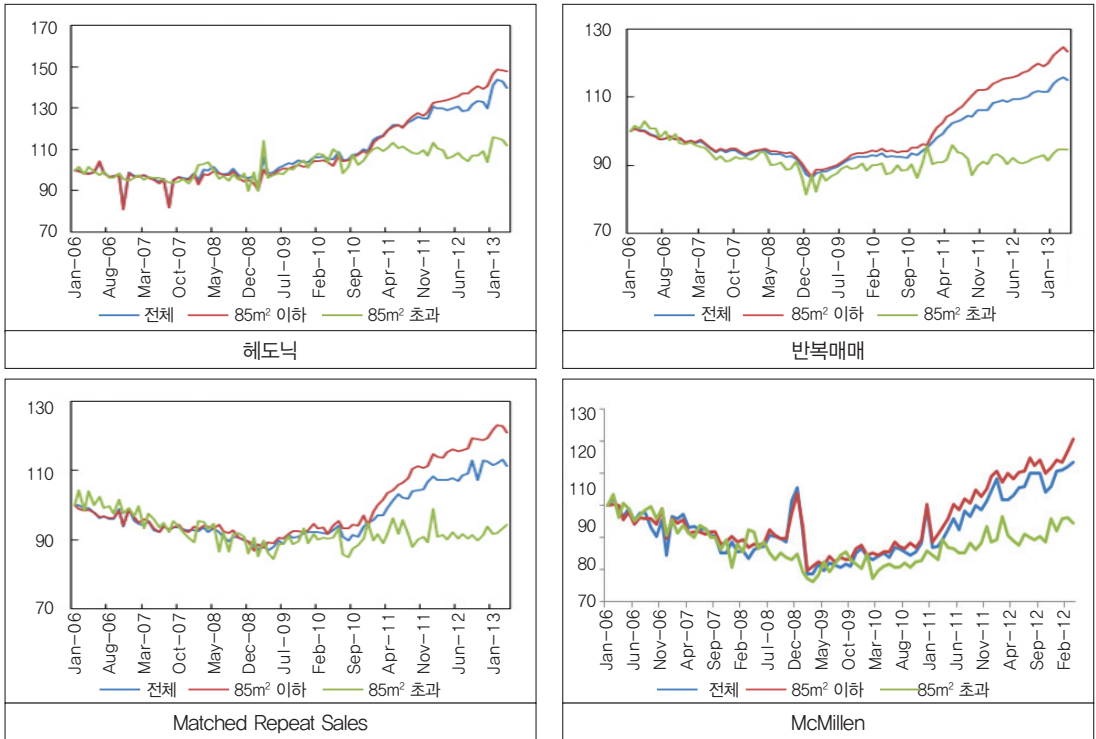


고 그 차이가 지수의 차이를 유발하는 것으로 보인다. 이 세 지수와 크게 다른 값을 보이고 있는 것은 헤도닉 가격지수다. 헤도닉 가격지수는 다른 지수들보다 상대적으로 과대평가되고 있으며, 그 차이도 꽤 있는 것으로 파악된다.

McMillen 지수는 다른 모든 지수보다 전반적으로 과소평가를 하고 있는 것으로 나타났으며, Matched Repeat Sales 지수는 McMillen 지수를 제외한 다른 지수들(한국감정원지수, KB 지수, 반복매매지수)과 비슷한 움직임을 보이지만, 지수값이 조금씩 더 작은 것으로 나타났다. 특히, McMillen 지수의 경우 2009년 초반부터, Matched Repeat Sales 지수의 경우 2011년 중반부터 그 값이 다른 지수들이 제시하는 것보다 작게 나타나기 시작했음을 알 수 있다.¹⁹⁾

〈그림 4〉는 전용면적별 아파트 매매가격지수를 비교하고 있다. 〈그림 4〉에 의하면, 전반적으로 부동산 매매지수는 상승해왔지만, 이러한 상승세를 이끄는 것은 85m² 이하 아파트라는 사실을 알 수 있다. 헤도닉 가격지수의 경우에는 85m² 이하 아파트와 85m² 초과 아파트의 매매가격 움직임이 반대로 움직이기 시작한 것은 2011년 초순부터다. 2011년에 들어서 85m² 이하 아파트 매매가격은 상승하기 시작했고, 85m² 초과 아파트 매매가격은 하락하기 시작했다. 전체적인 아파트 매매가격의 움직임은 85m² 이하 아파트가 주도하였다. 이같이 85m² 이하 아파트와 85m² 초과 아파트의 상반되는 움직임은 반복매매지수, Matched Repeat Sales 지수, McMillen 지수에서도 나타나는데, 그 시작점은 헤도닉 가격지수

그림 4 _ 전용면적별 아파트 매매가격지수 비교



19) 본 논문에서 보고하고 있지는 않지만, 구 단위까지 지수를 산정하였음. 구별로는 대체적으로 모든 지수의 움직임이 들쭉날쭉하였고, 특히, 표본수가 적은 남구, 서구, 중구에서는 그 움직임이 상당히 크게 변동하였음.

와 다르다. 반복매매지수, Matched Repeat Sales 지수, McMillen 지수는 85m² 이하 아파트와 85m² 초과 아파트의 상반된 움직임은 헤도닉 가격지수보다 약 1년 빠른 2010년부터 나타나기 시작했다.

〈표 5〉는 아파트 매매가격지수별 성장률의 표준편차를 보여준다. 이에 의하면, 감정가를 이용하여 산출하는 KB지수는 실거래가를 이용하는 다른 지수들보다 표준편차가 상당히 작다는 것을 알 수 있다. 즉, 실거래가를 사용하지 않으면 나타날 수 있는 평활화 문제가 KB지수에는 내재되어 있음을 보여주는 것이다.

실거래가를 이용하고 있는 한국감정원지수의 경우, 그 변동성이 다른 지수보다는 상당히 작다는 것을 알 수 있다. 반면에 헤도닉 가격지수와 McMillen 지수의 경우, 그 변동성이 다른 지수보다 상당히 크다는 사실도 알 수 있다. 반복매매지수의 경우에는 그 변동성이 많이 줄어드는데, 그 이유는 한국감정원지수처럼 그루핑을 통하여 매매가격을 어느 정도 평활화하기 때문일 것으로 판단된다. 한국감정원지수와 반복매매지수는 그 추정방법이 동일하지만, 차이점은 동일한 아파트로 간주할 때에 나타난다. 한국감정원지수의 경우에는 최저층, 최고층, 그 외 나머지는 모두 같은 층으로 분류하여 그룹마다 평균 가격을 이용하기 때문에 이 단계에서 평활화가 많이 이루어진다. 이러한 문제를 극복하기 위하여 본 논문에서 사용한 반복매매지수의 경우에는 한국감정원지수보다는 엄격한 조건을 적용하여 같은 층일 때만 동일한 아파트

로 간주하였다.

마지막으로 Matched Repeat Sales 지수는 매칭의 방법을 이용한 뒤 반복매매법과 동일한 방식으로 지수를 산정하였지만, 반복매매지수와 비교해 보면 그 변동성이 상당히 크다는 것을 알 수 있다.

김명준 외(2008)는 한국감정원지수 및 반복매매지수법의 경우 실거래가를 사용함에도 불구하고 KB지수와 비슷한 표준편차 값을 가지고 있기 때문에 이로부터 발생하는 변동성 축소의 문제점을 해결할 수 있는 방법이 모색되어야 한다고 지적하고 있다. 즉, 김명준 외(2008)는 변동성의 축소가 시장의 위험을 실제보다 과소평가할 가능성이 있다고 보고 있는 것이다.

하지만, 실거래가를 사용하여 지수를 만들 경우, 아파트의 특성에 따른 가격 차이를 충분히 통제하지 못한다면, 아파트 매매지수에 노이즈가 많이 들어가게 되어 변동성이 커질 가능성이 존재한다. 따라서 변동성만으로 시장의 움직임을 잘 반영한다고 보기 어려운 점이 있다는 것을 인지해야 한다.

〈표 6〉은 일반적으로 부동산 지수의 신뢰도를 평가하는 지표로 사용되고 있는 상대표준오차(Coefficient of Variation: CV)를 나타낸다. 상대표준오차는 추정값의 표집오차를 나타내기 위한 것으로 다음과 같이 계산된다.²⁰⁾

$$CV(\%) = [\text{표준오차} / \text{추정값}] \times 100$$

표 5 _ 아파트 매매가격지수 변동률의 표준편차(%)

구분	헤도닉	반복매매	Matched Repeat Sales	McMillen	한국감정원지수	KB지수
대구시 전체	4.55	0.93	1.83	5.69	0.79	0.56

20) 통계학에서는 일반적으로 변이계수라는 용어를 많이 사용함. 이 계산식은 통계청과 한국감정원의 '전국주택가격동향조사'의 통계설명자료에 따른 것임.

표 6 _ 아파트 매매가격지수의 상대표준오차(%)

구분	헤도닉	반복매매	Matched Repeat Sales	McMillen	한국감정원지수	KB지수
대구시 전체	1.43	0.82	0.80	1.27	0.92	0.84

이 계산식에 따라 각 지수의 상대표준오차를 구한 결과가 <표 5>이고, 이에 의하면 Matched Repeat Sales 지수가 가장 작은 값을 가짐으로써 다른 지수들보다 상대적으로 안정적인 것으로 나타났다. 반면, 헤도닉 가격지수와 McMillen 지수는 1%를 넘는 값을 가졌다.²¹⁾

VI. 결론

본 논문은 기존에 사용되고 있는 헤도닉 가격지수와 반복매매지수 이외에 매칭을 이용한 지수 등 다양한 지수 측정을 시도하고 이를 통해 주택시장에 대한 정확한 정보를 제공하려는 것이 목적이다. 구체적으로는 헤도닉 가격지수가 갖는 자료의 부정확성, 반복매매지수가 갖는 자료의 누락을 보완하기 위하여 매칭을 이용한 방법을 제시하였다. 대구시의 아파트 가격 정보를 이용하여 분석한 결과, 한국감정원지수, KB지수, 반복매매지수는 거의 비슷한 값을 가지는 것으로 나타난 반면, 헤도닉 가격지수는 다른 지수들에 비해 상대적으로 과대평가를, McMillen 지수는 상대적으로 과소평가를 하는 경향이 있었다. Matched Repeat Sales 지수는 한국감정원지수, KB지수, 반복매매지수 등과 비슷한 움직임을 보이면서도 조금씩 다른 특징을 나타냈다.

이처럼 본 논문은 실거래가를 이용하여 부동산과 관련한 다양한 지수를 지역시장에 적용하였고, 특히, 매칭을 이용한 방법을 지역 차원에서 분석한 최초의

연구라는 데에 그 의의가 있다. 대구의 아파트 실거래가를 이용한 다양한 지수 측정은 대구지역의 시장 상황에 대한 정확한 정보 제공을 가능하게 하여 부동산 시장의 안정화 및 지역 특성을 이해하는 데 도움이 될 것으로 기대하고, 이러한 방법론으로 앞으로 체계적인 검증을 받는다면 다른 지역의 부동산 시장을 분석하는 데에도 많이 이용될 것으로 보인다. 특히 지역 단위의 부동산 시장 또는 가격에 대한 연구들이 많지 않은 상황에서 앞으로 진행될 여타 논문들에도 중요한 참고자료가 될 것으로 판단된다. 게다가 다양한 지수들 간에 발생하는 차이를 면밀히 분석함으로써 부동산 시장을 예측하는 데에 기여할 것이다. 가령 매칭을 이용한 지수가 다른 지수보다 특정 시차를 두고 선행 혹은 후행하는지의 상관관계가 파악된다면, 지역 주택시장이 하강 국면에 접어들었는지 상승 국면에 접어들었는지를 파악하는 데에 도움이 될 것이다. 이러한 주택시장에 대한 정확한 파악은 정책 실시 적기를 놓치지 않기 위한 필수불가결한 요소임이 틀림없다.

본 논문은 이러한 의의를 가지고 있음에도 불구하고 다음과 같은 한계를 가진다. 첫째, 본 논문에서 시도한 매칭이 가지고 있는 본연의 문제점으로 규모 혹은 위치가 다른 아파트가 매칭될 수 있다는 것이다. 앞으로 이에 대한 방법론적인 해결책이 연구되어야 한다. 둘째, 다양한 지수를 제시한 뒤, 다른 지수들과 비교하고 있음에도 어떤 지수가 우월한지에 대한 명확한 근거가 부족하다는 것이다. 본 논문에서는 표본

21) 2014년 2월 20일 한국감정원에서 개최된 부동산통계 지역전문가 간담회에서 저자가 CV값의 크기와 지수의 신뢰도에 대해서 문의한 결과, 한국감정원에서는 CV값이 1.5% 이내이면 지수가 안정적이라고 판단하고 있다는 답변을 얻었음.

의 크기와 지수의 변동성, 그리고 상대표준오차를 비교하고 있는데, 이러한 값들이 아파트의 가격 상황을 제대로 적시에 반영하는지에 대해서는 여전히 의문의 여지는 있다. 따라서 지수의 우월성을 보증할 수 있는 근거를 찾는 것이 앞으로의 과제로 남아 있다.

참고문헌 •••••

김기호, 이성우. 1998. 해안변 아파트 단지의 주택가격 분포특성에 관한 연구: 부산시 사례를 중심으로. 국토계획 33권, 2호: 119-134.

김명준, 박광우, 신용현, 조훈, 현정순. 2008. 주택 가격지수 산정: 서울 아파트 실거래가격을 이용한 실증연구. 한국은행 금융경제연구원 Working Paper 348호.

김성우, 정진섭. 2010. 부산 아파트 실거래가를 이용한 전통적 헤도닉모형과 공간계량모형간의 적합도에 관한 비교 연구. 부동산학연구 16집, 3호: 41-55.

박연우, 방두완. 2011. 국내 부동산 가격지수 개발방향. SIRFE Occasional Paper 11-P01. 서울: 서울대학교 금융경제연구원.

박현수. 2009. 특성가격모형을 활용한 아파트 실거래가격지수 산정방법에 관한 연구. 부동산학연구 15집, 3호: 111-125.

이용만. 2007. 특성가격함수를 이용한 주택가격지수 개발에 관한 연구: 시간변동계수모형에 의한 연쇄지수. 부동산학연구 13집, 1호: 103-125.

이창무, 김용경, 배익민. 2007. 반복매매모형을 이용한 아파트 실거래지수 운영특성 분석. 부동산학연구 13집, 2호: 21-40.

최종일, 심성훈. 2002. 서울시 아파트 가격에 대한 대기질의 영향: 헤도닉 기법을 이용하여. 자원·환경경제연구 11권, 2호: 261-278.

Bailey, Martin J., Richard F. Muth, and Hugh O. Nourse. 1963. A regression method for real estate price index construction. *Journal of the American Statistical Association* 58: 933-942.

Barkham, Richard, J., and David Geltner. 1996. Price discovery and efficiency in the UK housing market. *Journal of Housing Economics* 5, no.1: 41-63.

Case, Karl E., and Robert J. Shiller. 1989. The efficiency of the market for single-family homes. *American Economic Review*

79: 125-137.

Chau, Kwong Wing, Bryan D. MacGregor, and Gregory M. Schwann. 2001. Price discovery in the hong kong real estate market. *Journal of Property Research* 18, no.3: 187-216.

Chau, K. W., S. K. Wong, and C. Y. Yiu. 2005. Adjusting for non-linear age effects in the repeat sales index. *Journal of Real Estate Finance and Economics* 31: 137-153.

Deng, Yongheng, Daniel P. McMillen, and Tien Foo Sing. 2012. Private residential prices in singapore: A matching approach. *Regional Science and Urban Economics* 42: 485-494.

Geltner, David. 1991. Smoothing in appraisal-based returns. *Journal of Real Estate Finance and Economics* 4, no.3: 327-345.

Geltner, David, Bryan D. MacGregor, and Gregory M. Schwann. 2003. Appraisal smoothing and price discovery in real estate market. *Urban Studies* 40: 1047-1064.

Mark, J. and M. Goldberg. 1984. Alternative house price indices: an evaluation. *American Real Estate and Urban Economics Association Journal* 12: 30-49.

McMillen, Daniel P. 2003. The return of centralization to chicago: Using repeat sales to identify changes in house price distance gradients. *Regional Science and Urban Economics* 33: 287-304.

McMillen, Daniel P. 2012. Repeat sales as a matching estimator. *Real Estate Economics* 40: 745-773.

Meese, Richard A. and Nancy E. Wallace. 1997. The construction of residential housing price indices: A comparison of repeat-sales, hedonic-regression, and hybrid approaches. *Journal of Real Estate Finance and Economics* 14: 51-73.

Palmquist, R. 1980. Alternative techniques for developing real estate price indexes. *Review of Economics and Statistics* 66: 394-404.

Rosen S. 1974. Hedonic prices and implicit markets: Product differentiation in pure competition. *Journal of Political Economy* 82: 34-55.

Rosenbaum, Paul R., and Donald R. Rubin. 1983. The central role of the propensity score in observational studies for causal effects. *Biometrika* 70: 41-55.

Rubin, Donald R. 1974. Estimating causal effects of treatments in randomized and nonrandomized studies. *Journal of Educational Psychology* 66: 688-701.

Shimizu, C., K. G. Nishimura and T. Watanabe. 2010. Housing Prices in Tokyo: A Comparison of Hedonic and Repeat Sales Measures. *JSPS Grants-in-Aid for Creative Scientific Research Understanding Inflation Dynamics of the Japanese Economy, Working Paper Series* no. 62.

Shimizu, C., H. Takatsuji, H. Ono, and K. G. Nishimura. 2010. Structural and Temporal Changes in the Housing Market and Hedonic Housing Price Indices. *International Journal of Housing Markets and Analysis* 3, no.4: 351-368.

Thibodeau, T. 1989. Housing price indices from the 1974-83 SMSA annual housing surveys. *American Real Estate and Urban Economics Association Journal* 17: 110-117.

- 논문 접수일: 2014. 4. 8
- 심사 시작일: 2014. 4. 21
- 심사 완료일: 2014. 9. 10

요약

주제어: 실거래가, 매칭법, 헤도닉 가격지수, 반복매매지수

본 논문은 기존에 많이 사용된 헤도닉 가격지수와 반복매매지수 이외에 매칭을 이용한 지수 등 다양한 지수 측정을 시도하고 이를 통해 주택시장에 대한 정확한 정보를 제공하려는 것이 목적이다. 특히, 헤도닉 지수가 갖는 자료의 부정확성, 반복매매지수가 갖는 자료의 누락을 보완하기 위하여 이용한 매칭법은 지역 차원에서 분석한 최초의 연구라는 데에 그 의의가 있다. 대구시의 아파트 가격정보를 이용하여 분석한

결과, Matched Repeat Sales 지수는 한국감정원지수, KB지수, 반복매매지수 등과 비슷한 움직임을 보였다. 헤도닉지수는 다른 지수들에 비해 상대적으로 과대평가를, McMillen 지수는 상대적으로 과소평가를 하는 경향이 있었다. 본 논문에서 제시한 매칭 방법을 이용한 지수는 다른 다양한 지수들과의 차이에 대한 면밀한 분석을 통하여 부동산 시장의 움직임을 예측하는 데에 많은 도움이 될 것으로 기대된다.