

군집 분석을 활용한 아파트 실거래가격지수 연구

- 서울시 사례를 중심으로 -

A Study on Apartment Transaction Price Index Using Cluster Analysis - Focused on the case of Seoul -

김경외* · 정성도**
Kim, Keung Oui · Jung, Sung Do

■ Abstract ■

This paper suggests a new method for calculating the transaction price index, which applies cluster analysis to ameliorate interregional heterogeneity and instability caused by missing data of transaction prices. In order to address the issue concerning missing values, absent data were estimated by using either the linear regression model, which utilized clustering information, or the linear interpolation model, depending on the continuity patterns of missing values. And in order to address the issue of interregional heterogeneity, the clustering analysis-applied apartment transaction price index for administrative district Gu was drawn from information on transaction prices of administrative district Dong. As a result of analyzing data of apartment transaction prices in Seoul from July 2011 to June 2016, it was found that the clustering analysis-based apartment transaction price index is more suitable than the existing index in terms of data applicability and the alleviation of regional inconsistencies.

Key Words : Transaction price index, Clustering analysis

* 서울대학교 기술경영경제정책대학원 경제학 박사과정 (주저자, awekim@snu.ac.kr)

** 한국엔지니어링협회 정책연구실 선임연구원 (교신저자, sungdo81@gmail.com)

1. 서 론

1. 연구의 배경

주택 가격은 주택시장 뿐만 아니라 개인, 정책당국, 금융기관 모두에 유의미한 영향을 미치는 요소로서 주택 가격의 안정화는 해당 시장을 포함한 국가 전체 경제의 안정화와 밀접하게 연관되어 있다(김경외 외, 2016). 이러한 측면에서 부동산가격지수는 정책당국자와 투자자가 시장의 안전성을 파악하고 개인이 자산 관리 의사결정을 내리는데 있어 중요한 판단 기준이 될 수 있다. 또한 부동산가격지수는 주택 가격에 영향을 미치는 요인들을 분석하고(이용민 외, 2010) 가격 변동을 예측하기 위한 연구 자료로서 그 활용 가치가 크다고 볼 수 있다.

부동산가격지수는 지수 작성에 활용되는 자료의 종류에 따라 평가가격에 기초한 지수와 실거래가격에 기초한 지수로 구분된다. 초기 부동산 연구에서는 시세 자료, 즉 부동산중개사와 같은 관련 전문가들이 표본주택을 대상으로 도출한 감정평가가격지수가 주로 사용되었다(이용만, 2007b; 이용만 외, 2008). 시세 정보를 활용한 감정평가가격지수는 거래가 없는 시점의 가격 정보를 제공할 수 있다는 장점이 있지만, 실제 가격보다 변동성이 작게 나타나는 평활화(smoothing) 현상(Chinloy et al., 1997; Fisher et al., 1994)과 실제 가격 변동과의 시차로 인해 결과적으로 부동산 가격의 변동성을 과소평가하는 문제점을 안고 있다(이창무 외, 2005; 이용

만 외, 2008; 남수현 외, 2009).

이에 반해 실거래가격지수는 실제 거래된 정보를 토대로 작성되는 실거래가격으로 작성되어 시장상황을 객관적으로 반영한다는 장점이 있다(이창무 외, 2008a). 그러나 실거래가격지수는 지수의 산정 방식에 있어서 몇 가지 문제점을 안고 있다. 먼저, 감정평가가격지수와 달리 실거래가격지수는 표본으로 삼는 주택들이 매기 거래된다는 보장이 없어 자료의 안전성 혹은 결측치 문제가 발생할 수 있다(이창무 외, 2008b). 뿐만 아니라, 감정평가가격지수와 같이 실거래가격지수는 국가에서 지정한 구나 시 단위로 산출이 되기 때문에 하위 지역 내에서 발생하는 가격 이질성을 고려하지 못한다. 그 예로, 시 단위로 산출된 실거래가격지수 값은 지역 내에서 발생하는 신규 아파트와 재건축 대상 아파트 간의 가격 차이를 고려하지 않기 때문에 편향된 가격 정보가 포함될 우려가 있다.

이처럼 기존의 감정평가가격지수와 실거래가격지수는 부동산가격지수의 본래 목적에 비추어 볼 때 자료의 특성과 지수 산출 방식에서 한계를 보인다. 따라서 이를 극복하기 위해서는 실거래가격 자료가 주는 신뢰성을 기반으로 자료의 안정성과 지역 내 주택 가격의 이질성을 극복할 수 있는 합리적인 지수 산출 방안이 모색되어야 한다.

2. 연구의 목적

본 연구에서는 국토교통부에서 제공하는 아

파트 실거래가격 정보를 활용해 기존의 실거래 가격지수의 문제점들을 보완한 새로운 아파트 실거래가격지수를 제안하고자 한다. 일반적으로 가격지수는 연구자가 추구하는 목표가 무엇 이냐에 따라 그 구성과 지향점이 달라지는데, 본 연구의 실거래가격지수는 실거래가격의 변동성을 보다 정확하고 구체적으로 보여줄 수 있는 신뢰도 높은 가격 정보 제공에 주안점을 두었다. 이를 위해 본 연구에서는 분석 대상들 간의 유사성을 띄는 집단을 정의하고(Řezanková, 2014), 집단 내 이상치를 제거하는데 효과적인 군집 분석(cluster analysis)을 적용하였다. 따라서 실거래가격을 대상으로 군집 분석을 실시하면 가격변동성이 비슷한 지역들을 보다 직관적이고 효율적으로 분류하는 것이 가능하며 이를 통해 추가적인 자료 해석과 시사점을 도출할 수 있다.

이에 본 연구의 목적은 크게 두 가지로 구성하였다. 첫 번째 단계에서는 선형보간법(linear interpolation)과 가격 변동성이 가장 유사한 군집의 회귀식 정보를 활용한 선형회귀대체법(linear regression imputation)을 사용하여 아파트 실거래가격 정보 상에서 나타나는 결측치를 추정하고자 한다. 두 번째 단계에서는 결측치를 추정한 행정동 실거래가격 정보로 군집 분석을 실시하여 행정구 단위의 군집별 실거래가격지수를 도출할 것이다. 그리고 이 군집별 아파트 실거래가격지수의 적절성은 기존의 실거래가격지수와 구(區)내의 군집 가격 지수와의 상관관계, 구별 분산, 구내의 세부 군집 가격지수의 분산 차이를 비교 한다.

II. 선행 연구

1. 주택 가격 지수 연구

주택가격지수는 주택 시장의 대표성을 나타내는 지표로서 시장의 상황을 진단하고 분석하는데 있어 기초가 된다(이용만, 2007). 국내에서 제공하고 있는 감정평가가격지수로는 국민은행과 부동산 114가 있으며 이들은 표본주택의 가격을 이용해 라스파이레스(Laspeyres) 방식으로 주택가격지수를 산정하였다. 라스파이레스는 주로 물가지수나 생산지수 작성에 주로 사용되는 방식으로 기준 시점의 수량과 가격을 고정시킨 상태에서 비교 시점의 가격의 변동분을 계산하는 방식이다(안성남, 2000). 따라서 이러한 산정 방식은 매월 선정된 표본주택에 한해 자료를 매월 수집할 수 있다는 시세 자료의 특성에 적합하다. 다만 라스파이레스 방식은 고정된 표본주택의 가격 변화를 제외한 다른 요소들을 고려하지 않기 때문에 왜곡된 결과를 낼 수 있으며, 기준점으로부터 시간이 경과할 경우 일정 정도의 편이가 발생할 수 있다. 뿐만 아니라 이러한 지수 산정 방식에서는 고가주택의 가격변화가 지수산정에 영향을 미칠 수 있고(이창무 외, 2002; 이창무 외, 2008b), 시세 자료의 특성으로 인한 시차 문제와 실거래가격 간의 차이가 있는 것으로 나타났다(이창무 외, 2005).

이에 반해 실거래가격지수는 실제 거래된 실거래가격으로 작성되기 때문에 주택 가격에 대한 보다 객관적인 정보를 제공할 수 있다. 우리나라에서는 2006년 1월부터 실거래가격 자료가

공개되기 시작하면서 매매와 전세 그리고 주택과 아파트에 대한 정보를 쉽게 구할 수 있게 되었다. 이를 토대로 한국감정원에서 매달 발표하고 있는 실거래가격지수는 동일가중 반복매매모형을 이용해 산정된 자료로서 개별 주택의 평균 변동률을 가격변동률로 가정하고 있다(Baily et al., 1963; 류강민, 2010). 이 중 아파트 실거래가격지수는 가계 자산 중 비중이 가장 큰 주택 종류 중 하나이면서 주택 시장의 큰 부분을 차지하는 아파트의 가격 정보를 제공하기 때문에 그 중요도가 높다.

일반적으로 실거래가격지수 산정에는 반복매매모형(Repeat sales model)이 사용된다. 반복매매모형은 최소 두 번 이상 거래가 이루어진 주택의 가격 증감율을 구하여 지수화하는 방식이다. 이러한 측면에서 반복매매모형은 별도의 통제가 필요 없고 동일 주택이라는 가정 하에서 별도의 표본 보정 없이 일관된 지수 산정이 가능하다는 장점이 있어 실거래가격지수 산정에 있어 적합하다고 볼 수 있다(이창무 외, 2005). 하지만 실거래가격이 제공하는 여러 장점에도 불구하고 기존의 실거래가격지수에는 여러 문제점들이 있다.

먼저 자료의 특성상 거래가 발생하지 않을 경우 나타나는 자료 활용의 비효율성 문제가 있다. 예를 들어 동일한 조건을 가진 아파트의 거래가격이 반복적으로 관측되지 않으면 결과적으로 지수산정에서 제외가 되기 때문에 지수의 안전성에 문제를 야기할 수 있다. 또한, 두 시점 사이에서 주택 특성변화가 나타나는 상황을 고려하지 못한다. 가령 지수가 산정되는 두 시점

사이에 재개발 등과 같은 효과로 인해 주택의 특성이 변하게 되면 이때의 관측되는 가격 정보는 이전 시점과 비교했을 때 왜곡된 결과로 이어질 수 있다. 이에 주택의 특성이 변한 사례를 제외하는 방안이 제시되기도 하였으나, 이는 오히려 반복매매모형의 자료효율성 문제를 악화시키는 것으로 나타났다(Case et al., 1991; 이용만, 2007a). 이에 더하여 매시점마다 관측되는 주택의 특성이 동일하지 않아 거래량의 빈도수에 따라 표본 편이가 발생할 수 있다. 만약 지수 산정에 있어서 자주 반복되어 거래되는 주택들이 어떤 특성을 공유하게 된다면 이때의 표본 편이는 클 수밖에 없어 지수전반의 신뢰도에 부정적인 영향을 미칠 수 있다. 따라서 부동산 시장 상황을 보다 객관적으로 분석하기 위해서는 기존의 지수산정방식의 문제점들이 개선되어야 한다.

2. 군집 분석 연구

군집 분석은 자료의 상호연관성에 근거하여 동질적인 집단을 분류하는 기법으로(Anderberg, 1973), 기존의 연구에서는 다변량자료의 차원 축소(dimension reduction)나 자료 내 집단 분류를 위해 사용되어 왔다. 특히 샘플에 대한 사전 정보가 없거나 샘플 내의 그룹을 구분하는 기준이 명확하지 않을 경우, 군집 분석은 요소 간의 거리를 측정하여 유사성을 파악할 수 있기 때문에 빅데이터 분석에서도 주로 사용된다. 군집 분석은 유사성 혹은 거리의 정의와 분류 절차에 의하여 그 종류가 구분되며, 부동산 연구에서는 변

수들을 분류하는 기법인 R-mode 분석이 주를 이루고 있다. 이에 본 절에서는 군집 분석을 활용한 부동산 연구 문헌들을 살펴보고자 한다.

부동산 연구에서 군집 분석은 동일 행정 구역 내에 존재하는 부동산이나 관련 요소들의 유사성과 이질성을 파악하고 이를 효과적으로 분류하기 위해 사용되어 왔다. 부동산을 활용한 군집 분석 연구들은 군집화하는 대상과 이를 활용하고자 하는 그 목적에 따라 각기 다르며, 크게는 부동산 외적인 요소를 분류하거나 부동산을 분류하는 두 가지 연구로 구분된다. 먼저 부동산 외적인 요소를 분류하는 연구에서는 군집 분석을 통해 부동산 관련 정보들을 연구하거나 이를 통해 새로운 분류 기준을 제시하고자 하였다(김성록, 2013; 김선진 외, 2015; 하정순 외, 2017). 김성록(2013)은 산업간 연계성을 활용한 군집 분석을 통해, 부동산산업과 타 산업 간의 연계 관계를 보였으며, 김선진 외(2015)는 부동산 경제 발전 수준으로 중국 지역을 분류하여, 부동산 가격 상승이 민간 소비에 미치는 부의 효과를 분석하였다. 하정순 외(2017)는 소비자의 부동산투자유형으로 군집을 분류하여 각각의 특성이 잘 반영된 시장 세분화의 기준이 될 수 있음을 보였다. 이처럼 부동산은 사회경제학적 요소들과 관련이 있기 때문에 다양한 부동산 자료들을 활용한 군집 분석이 가능하며, 그 결과 역시 자료 분석에 있어 타당한 것을 알 수 있다.

부동산 분류 연구에서는 부동산 관련 정보들이나 특성들을 토대로 부동산을 분류하고자 하였다(구성환 외, 2010; 정승영 외, 2010; 김재직, 2015). 정승영 외(2010)는 서울시 평균 상가권

리금과 상가임대료를 이용해 상가공간의 시장 군집을 구분한 결과, 두 기준 모두 높은 집단으로 분류되는 강남역 상권을 제외한 나머지 지역에서 상가권리금과 상가보증금에 의한 시장세분화의 차이가 존재하는 것을 확인하였다. 구성환 외(2010)는 기존의 행정구역에서 구별되지 못하는 인구구조, 경제구조를 반영하기 위해 군집분석을 실시하여 도시적 특성인 인구, 활동, 토지 및 시설을 기반으로 수도권 도시의 지역 유형을 여섯 가지로 구분하였다(구성환 외, 2010). 실거래가격을 활용해 부동산을 분류한 연구로 김재직(2015)은 서울시 64개 동 내의 아파트 전용면적별 실거래 매매가 자료로 군집을 산출하였으며, 그 결과, 군집 분석을 통해 분류된 행정동들을 실거래가격으로만 분류했음에도 불구하고 지리적인 특성들이 잘 반영됨을 보였다. 이는 실거래가격을 활용한 군집 분석 결과가 자료 중심적인 해석임에도 불구하고, 현실적인 특성을 잘 보여줄 수 있다는 사실을 시사한다.

일반적으로 부동산을 분류하는 가장 대표적인 기준인 행정구역으로 지역을 구분하게 되면, 자료의 지속성과 연속성 측면에서 장점을 취할 수 있지만, 하위 지역들의 이질적인 특징들을 구별하기에는 어려움이 있다. 그러나 기존의 연구에서는 일반적인 행정구역 중심의 분류에서 발생하는 이질성 문제를 극복하고자 군집 분석을 활용하였으며, 그 결과 기존의 행정구역이 구별해내지 못하는 지역 간의 이질성을 효과적으로 파악하였다. 이에 본 연구에서는 군집 분석을 이용해 실거래가격자료의 결측치를 추정하고, 지역 내 가격 상의 이질성이 최소화되어 본래의

지수가 목적으로 아파트 가격의 변동성을 정확하게 관찰할 수 있는 지수 산정 방식을 제안하고자 한다.

III. 데이터 및 분석

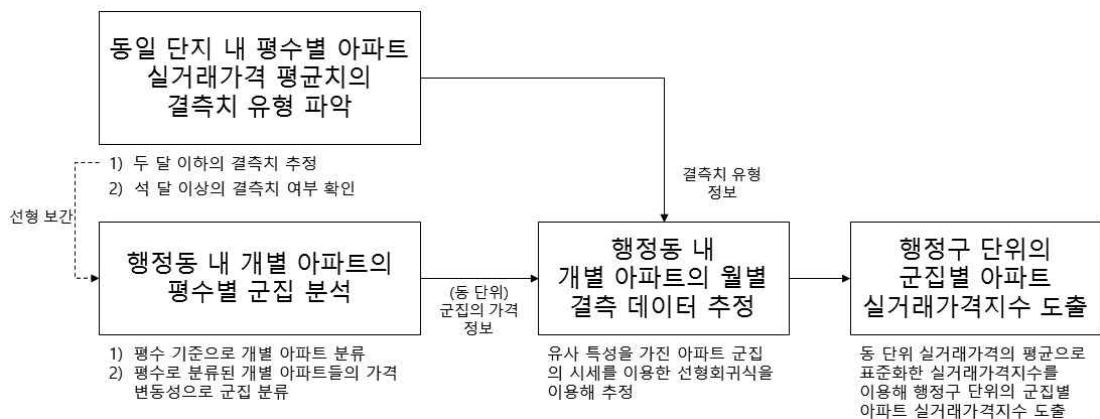
1. 데이터

본 연구에서는 2011년 7월부터 2016년 6월까지 총 60개월 동안 서울시 지역 내에서 기록된 아파트 매매 실거래가격자료를 이용해 분석하였다. 실거래가격 자료는 실제로 거래된 가격 정보를 제공하기 때문에 기존의 시세 중심의 정보보다 객관적이고 신뢰성 높은 판단이 가능하며, 매월 갱신된 자료를 쉽게 구할 수 있다는 장점이 있다(고필송 외, 2010). 우리나라에서는 2006년 1월 1일부터 체결되는 모든 부동산과 부동산 권리에 대한 계약을 60일 이내에 관할 시군구청에 신고하도록 되어 있으며, 관련 정보는

국토교통부를 통해 제공되고 있다. 국토교통부에서 제공하는 실거래가격 자료는 전국 단위로 크게 매매와 전월세로 구분되며 유형별로는 아파트, 연립/다세대, 단독/다가구로 분류된다. 본 연구에서는 아파트가 일반 주택보다 상대적으로 거래량이 많고 국내 주택 시장에서 차지하는 비중이 크기 때문에(이영수, 2014), 아파트 매매 자료를 선정하였으며, 지면의 한계로 국내 부동산 시장에서 부동산 거래가 가장 활발한 서울시에 한해 실증 연구를 실시하였다.

2. 분석

본 연구에서는 실거래가격 자료를 관찰한 후, 결측치 유형에 따라 이를 추정하고 이를 토대로 행정구 단위의 군집별 아파트 실거래가격지수를 산출하는 총 네 가지의 연구 단계로 구성하였다(〈그림 1〉).



〈그림 1〉 연구 과정

1) 동일 단지 내 평수별 아파트 실거래가격 평균치의 결측치 유형 파악

아파트 실거래가격 자료의 최소 단위는 아파트의 거래 건수이며, 각 개별 아파트들의 거래 시기나 주기는 그 특성과 유형에 따라 다르게 관찰된다. 보통 아파트 매매 혹은 전세의 법적 거래 기간이 24개월이라는 사실을 미루어볼 때, 60개월의 기간 내에 개별 아파트 건수로 결측치를 추정하는 것은 사실상 불가능하다. 이에 본 연구에서는 아파트의 여러 특성 중 동일 단지 내의 아파트 평수를 기준으로 가격 정보를 묶어 도출한 평균값을 사용하였다. 한 단지 내에 속한 아파트들은 단지의 브랜드나 지리적인 특성을 공유하지만, 동시에 면적, 남향 여부, 층수 등에서 차이점을 안고 있다. 이러한 특성들은 아파트 가격에 영향을 미치는 요인들로서, 김진희(2014)에 따르면, 아파트 가격은 면적을 비롯한 여러 헤도닉 요인들에 영향을 받는 것으로 나타났다. 본 연구에서는 실거래가격정보가 제공하는 여러 정보들 중에 동일 단지 내에 위치한 아파트들의 가격변동성을 가장 효과적으로 구분할 수 있다고 보고, 개별 단지의 아파트 평수별로 평균한 실거래가격 자료의 결측치 유형을 파악하였다.

결측치 추정에 있어 결측치의 빈도나 연속성은 매우 중요한 판단 요소이며, 보다 정확한 결측치 추정을 위해서는 연속된 결측치의 수로 결측치의 유형을 파악해야 한다. 가령 연속된 결측치가 하나 혹은 두 개일 경우에는 선형보간법(linear interpolation)을 이용한 단순 추정이 가능하다. 선형보간법은 이산적인 자료에서 관측

되지 않은 사이 값을 추정하거나 결측치를 추정할 때 사용되는 방식으로서 이를 포함한 여러 보간법들은 주로 영상처리 관련 연구들에서 많이 응용되고 있으며, 부동산 시세관련 연구에서는 분석에 앞선 데이터 전처리 과정에 사용된다(정원구 외, 2007).

그러나 연속적으로 발생하는 결측치의 수가 늘어나 유효 데이터 사이의 거리가 커지는 경우에 선형보간법을 사용하면 실제 거래가격의 상승 혹은 하락의 전이가 발생하는 상황에서 오류를 범할 가능성이 있다. 이에 본 연구에서는 앞서 도출한 개별 단지의 아파트 평수별 실거래가격 평균 자료 상에서 나타나는 결측치 유형을 크게 두 달 이하와 석 달 이상으로 분류하고, 본 단계에서는 전자의 경우에 한해서만 선형보간법을 이용해 결측치를 추정하였다.

2) 행정동 내 개별 아파트의 평수별 군집 분석

두 번째 단계에서는 석 달 이상의 연속된 결측치 추정에 앞서 아파트 가격변동성으로 군집 분석을 실시해 아파트들의 유사성을 파악하였다. 군집 분석은 특정 그룹 내의 요소들을 다른 그룹의 요소들과 비교해 보다 유사한 것들로 구성되도록 그룹화하는 작업으로서, 분석의 기준이 되는 유사도를 정의하기 위해서는 우선적으로 유사의 정도를 계산하기 위한 특성을 정의해야 한다. 이에 본 연구에서는 아파트 군집의 유사 특성을 파악하고 이를 활용하기 위해 같은 행정동 내 개별 단지의 평수별 아파트의 실거래가격 평균치로 군집 분석을 실시하여 유사한 가격 흐름을 나타내는 군집을 형성하였다.

군집 분석을 위한 유사성의 측정기준은 요소들 간의 가격 상관관계를 이용한 거리의 개념을 활용하였다. 군집 분석에서 활용되는 거리 산출법으로는 유클리드(euclidean), 맨하탄(manhattan), 민코우스키(minkowski), 통계적(statistical) 거리 등이 있으며 본 연구에서는 유클리드 방식으로 요소들 간의 거리를 측정하였다. 예를 들어, 아파트 a1, a2, a3의 가격변동이 pa1, pa2, pa3라고 했을 때, 각각의 상관관계를 구해준 다음(식 1), 각 상관관계 간의 거리를 유클리드 방식으로 계산하였다(식 2). 이 경우, 타 요소들과의 가격 변동이 유사할수록 요소들 간의 거리는 줄어들게 되며, 결과적으로 거리가 가까운 요소들은 동일한 군집으로 분류되게 된다.

$$\rho_{p_{a1}p_{a2}} = \frac{cov(p_{a1}, p_{a2})}{\sigma_{p_{a1}} \sigma_{p_{a2}}} \quad (\text{식 1})$$

$$d_{Euclidean}(p_{a1}, p_{a2}) = \sqrt{(\rho_{p_{a1}p_{a2}} - \rho_{p_{a1}p_{a3}})^2 + (\rho_{p_{a1}p_{a3}} - \rho_{p_{a2}p_{a3}})^2} \quad (\text{식 2})$$

군집 분석에서 군집 생성을 위해 사용되는 방법으로는 계층적(hierarchical) 군집 기법, 중심기반(centroid-based) 군집 기법, 분포기반(distribution-based) 군집 기법, 밀도기반(density-based) 군집 기법 등이 있으며 본 연구에서는 분석 후 결과 설명 및 해석에 있어 시각적으로 용이한 계층적 군집 기법을 선정하였다. 계층적 군집 기법의 방식은 bottom-up 방식의 병합기법과 top-down 방식의 분할기법으로 나뉘는데, 본

연구에서는 n개의 군집들로 시작해 순차적으로 군집을 병합하는 병합기법을 사용하였다.

군집 간의 거리는 계층적 군집연결기법에서 정의하는 방식에 따라 와드법, 최장연결법, 최단연결법, 평균연결법 등으로 나뉜다. 대표적으로 사용되는 연결기법인 최단연결법과 최장연결법(식 3)은 두 군집(C₁, C₂) 간의 최단 혹은 최장거리로 군집간 거리로 정의하는 방식으로, 이들은 군집을 형성해나가는 과정에서 차이를 보인다. 가령 전자가 비슷한 크기의 여러 개 군집을 만든 다음 연결시킨다면, 후자는 큰 군집을 순차적으로 만들어간다는 특징이 있다. 이에 본 연구에서는 결과적으로 군집 간의 차이를 더 잘 나타낸다고 판단되는 최장연결법을 사용하였다.

$$d\{C_1, C_2\} = \max \left\{ d(\rho_{p_{a1}p_{a2}}, \rho_{p_{a1}p_{a3}}) \mid \rho_{p_{a1}p_{a2}} \in C_1, \rho_{p_{a1}p_{a3}} \in C_2 \right\} \quad (\text{식 3})$$

군집 분석에 있어 군집의 개수를 정하는 작업은 실거래 가격 정보를 활용한 다양한 분석에서 결과의 정확성과 목적 부합성의 문제와 직결되는 문제이다. 하지만 적절한 군집의 개수를 정하는 명확한 기준이 없어 이 부분에 있어서는 연구자의 해석에 따라 차이가 나타날 수 있다(송민경 외, 2010). 이에 본 연구에서는 적절한 원소의 수를 가질 수 있는 군집을 형성하기 위해 분석 기간에 포함되는 모든 월별 자료에 적어도 하나 이상의 실거래가격이 존재하도록 군집수를 설정하였다. 예를 들어, 동 내에 총 20개의 아파트 평수별 실거래가격이 있다고 가정했

을 때의 거래 가격과 거래량이 주어졌다고 가정해보자(〈표 1〉). 이 때 60개월 동안 적어도 한번 이상의 거래가 발생한 개월의 수가 45개라고 한다면, 이 때 군집의 수는 한 군집이 최소 다섯개 이상의 군집을 가질 수 있는 수로 설정된다.

〈표 1〉 군집 개수 설정 사례

i	apt	1	2	3	...	60	avg
1	a1	$p_{a1,1}$	$p_{a1,2}$	$p_{a1,3}$...	$p_{a1,60}$	30
2	a2	x	x	$p_{a2,3}$...	x	33
3	a3	$p_{a3,1}$	x	$p_{a3,3}$...	$p_{a3,60}$	40
...
20	e3	$p_{e3,1}$	x	x	...	$p_{e3,60}$	59
avg							45

3) 행정동 내 개별 아파트의 월별 결측 데이터 추정

앞의 두 단계에서 얻은 정보를 토대로 석 달 이상의 연속된 결측치를 선형회귀대체법을 이용해 추정하였다. 결측치 처리는 통계조사의 결과와 자료의 신뢰도 향상에 있어 매우 중요하다. 결측치처리에 있어 가장 보편적인 방법은 결측치를 잘 대체할 수 있는 보조 자료를 활용하여 샘플의 분산과 편향을 줄이는 것이다(최필근 외, 2008). 이를 위해서는 모형의 복잡성을 높이지 않고 정보의 손실을 최소화 하는 선에서 보조 자료를 선정해야 한다. 이에 본 연구에서는 앞서 파악한 개별 아파트의 결측치 유형과 동 단위 군집의 가격지수 변동을 토대로 선형회귀식을 이용하여 결측치를 추정하였으며, 그 절차는 다음과 같다.

우선 결측치를 추정하려고 하는 아파트의 가격 변동성과 가장 상관관계가 높은 개별 아파트의 평수별 군집을 파악하였다. 그런 다음, 개별 아파트의 실거래가격을 종속변수, 군집의 가격지수를 독립변수로 하는 선형회귀식을 수립하여 둘 간의 차이를 나타내는 계수를 구하고, 이를 이용해 아파트의 결측치를 추정하였다. 예를 들어, 아파트 a_1 과 가격 변동성이 가장 유사한 군집이 $C_1 = \{a_1, a_2, \dots, a_n\}$ 이고, 이 군집의 t 기 가격지수를 $p_{C_1,t}$ 라 하면, $p_{C_1,t}$ 는 군집에 포함된 아파트가격(p_{a_n})중 t 기에 거래정보가 있는 개별 아파트의 가격지수들의 평균값을 이용한다(식 4). 이때 t 기에 개별 아파트의 가격지수는 해당 아파트의 t 기 실거래가격을 평균값으로 정규화시킨 값(a_{kt}/\bar{a}_k)을 이용하며, 추정을 위한 선형회귀식은 (식 5)와 같다. 여기서 가격 변동성의 상관관계가 높다는 사실은 이 둘 간의 전반적인 가격의 흐름이 매우 유사하다는 사실을 시사한다. 이러한 사실에 기반하여 (식 5)를 통해 추정된 $\hat{\beta}_0$ 와 $\hat{\beta}_1$ 은 $p_{C_1,t}$ 의 정보를 통해 $p_{a1,t}$ 를 추정할 수 있는 계수로서 사용된다(식 6).

$$p_{C_1,t} = \frac{\sum_{a_{kt} \in C_{1,t}} a_{kt}/\bar{a}_k}{n(C_{1,t})} \quad (\text{식 4})$$

$$p_{a1} = \beta_0 + \beta_1 C_1 + u \quad (\text{식 5})$$

$$p_{a1,t} = \hat{\beta}_0 + \hat{\beta}_1 p_{C_1,t} \quad (\text{식 6})$$

4) 행정구 단위의 군집별 아파트 실거래가격지수 도출

기존의 시 혹은 구 단위 실거래가격지수는 해당 지역 내 이질적인 가격 정보를 포함하고 있어 가격 변동의 편향성 문제를 내포하고 있다. 따라서 행정구역 하위 지역들의 정확한 가격 변동성을 파악하기 위해서는 특정 지역 내에 여러 하위 지역들 중 가격 변동성이 유사한 지역들과 이질적인 지역들을 구별할 수 있어야 한다. 본 단계에서는 결측치를 복원한 실거래가격 자료를 이용해 앞선 단계와 동일한 접근 방식으로 군집 분석을 실시하여 이른바 행정구 단위의 군집별 아파트 실거래가격지수를 산출하고자 하였다.

행정구 단위의 가격 지수는 실제 행정구 단위가 아니라 행정구 내의 여러 행정동으로 나뉜 군집의 가격정보로 산출된다. 본 연구에서는 시간에 따른 군집간 변동폭을 비교하기 위해 초기 지수값으로 표준화하여 사용하였다. 또한, 행정구 단위의 실거래가격지수 산출에 있어 개별 행정동이 포함하는 아파트들의 월별 가격을 그 아파트의 60개월 평균가격으로 나눈 아파트들의 월평균 값을 구한 뒤, 이를 이용해 행정동 단위의 값을 도출하는 방식을 채택하였다(식 4)의 a_{nt}/\bar{a}_n).

이는 행정동 단위로 군집 분석을 하여 산출한 행정동 대표 실거래가격의 하위 구성요소인 개별 아파트의 절대 가격 차이의 정도가 결과물에 영향을 미칠 수 있기 때문이다. 가령 개별 아파트 간의 절대 가격 차이가 크다면, 특정 시기의 행정동 대표 실거래가격지수는 실제 아파트들의

등락률 보다는 해당시기에 발생한 절대가격이 큰 아파트의 매매 여부에 의해 결정될 수 있다.

지수 산출과 관련하여 예상할 수 있는 다른 문제는 개별 아파트를 단순 평균할 경우 개별 아파트들의 거래 가중치를 고려하지 못한다는 것이다. 그러나 실거래가격 자료상에서 개별 아파트사이의 거래량의 편차가 크기 때문에 단순 평균을 하더라도 거래량에 의한 가중치가 부과되는 효과를 얻을 수 있다. 예를 들어, 거래량이 많은 아파트와 거래량이 적은 아파트의 가격을 평균할 경우, 거래가 많은 아파트의 정보가 보다 많은 월에 나타나므로 지수에 많이 반영되고, 나아가 해당 월에 거래가 없는 아파트의 경우 가격 지수 산출에 반영되지 않으므로 가중치가 부과되는 결과를 얻을 수 있다.

IV. 분석 결과

1. 실거래가격 결측치 추정 결과

결측치 추정을 위한 본격적인 분석에 앞서 분석에 포함이 될 수 있는 실거래가격 자료를 선별하였다. 자료 검토 결과, 60개월 동안 한 건 이상의 매매가 발생한 25,094개의 단지 내 평수별 아파트의 유형 중에서 거래가 발생한 월의 평균값은 전체 기간 60에 훨씬 못 미치는 8.61로 나타났다(〈표 2〉). 이는 추정 단위를 단지 내 평수별 아파트 실거래가격 평균값으로 설정하고, 분석 대상을 우리나라에서 아파트 거래량이 가장 많은 서울시로 선정하였음에도 불구하고,

자료상에서 결측치가 차지하는 비중이 여전히 매우 크다는 사실을 시사한다. 뿐만 아니라, 개월 수와 누적 거래량의 비율은 거의 선형으로 증가하지만, 아파트의 유형은 이와 다르게 상당히 편향되었음을 확인할 수 있었다(〈표 3〉). 실제로 30개월 이상 거래가 있었던 7.8%의 아파트 유형이 전체 기간의 거래량에서 차지하는 비중은 49.5%로 특정 아파트 유형이 차지하는 비중이 매우 큰 것으로 나타났다. 이처럼 거래가 거의 없는 아파트 유형의 결측치를 선형회귀대체법을 통해 추정하게 되면 정보가 충분하지 않아 오히려 자료의 신뢰성이 떨어질 수 있다. 또한, 거래가 집중된 특정 아파트 유형만을 이용해 결측치를 추정 할 경우, 이를 제외한 다양한 아파트 유형의 자료 손실이 발생한다. 데이터 분석에서 소수의 요소를 가지는 자료를 분석에서 제외하는 것은 이상치를 제거하는 것과 같은 효과를 낼 수 있다(Sarle, 1983). 그렇기 때문에 적절한 실거래가격 추정을 위해서는 거래량과 아파트 유형이 동시에 고려되어야 한다.

〈표 2〉 매매가 발생한 달 통계

단지 내 평수별 아파트 유형	평균	중간 값	표준편차
25,094	8,61	3	11,56

〈표 3〉 개월 수에 따른 전체 대비 누적 거래량 비율(%)
및 누적 아파트 유형(%)

매매 개월 수	< 11	< 21	< 31	< 41	< 51
전체대비 거래량	17,6	33,7	50,5	67,2	82,4
아파트 유형	74,4	86,5	92,8	96,6	98,8

이러한 측면에서 본 연구는 추정하고자 하는 값의 수가 적어도 보유한 자료의 수보다는 작게 유지될 수 있게끔 하였다. 또한, 최소 2개 이상의 군집을 만들기 위한 조건으로 행정동 내에 30개월 이상 거래가 이루어진 아파트 유형의 수가 4개 이상이 되어야 된다고 판단하였다. 이는 유사성을 갖는 요소들의 군집을 나누는 군집 분석의 목적을 만족시키는 최소한의 기준이라 볼 수 있다. 이에 본 연구에서는 서울시 60개월 동안 한번이라도 거래가 있었던 달이 30개 이상인 아파트 단지를 최소 네 개 이상을 가진 133개의 행정동 자료만을 이용하였다. 그 결과, 거래량과 아파트 유형의 수가 적은 행정동들을 제외하면서, 동시에 선정된 행정동 내의 다양한 아파트 유형들을 고려하는 효과를 얻을 수 있었다.

이를 토대로 자료 내에서 추정 가능한 결측치의 유형을 살펴본 결과, 60개월 동안 발생한 총 1,505,640개의 자료 중 4.26%가 선형보간법, 81.38%가 선형회귀대체법 대상으로 분류됨을 확인하였다(〈표 4〉). 이 정보를 토대로 선형보간법과 선형회귀대체법을 통해 각각 확보한 자료는 실측치의 30%와 21.6%로, 전체 자료의 활용률을 14.35%에서 21.71%로 증가시킨 것으로 나타났다(〈표 5〉). 또한, 선형회귀대체법을 활용해 추정된 결측치가 가격의 선형적인 흐름뿐만 아니라 비선형적인 가격의 변동까지도 고려할 수 있다는 것을 확인할 수 있었다.

〈표 4〉 추정 가능한 결측치 유형 분석 결과

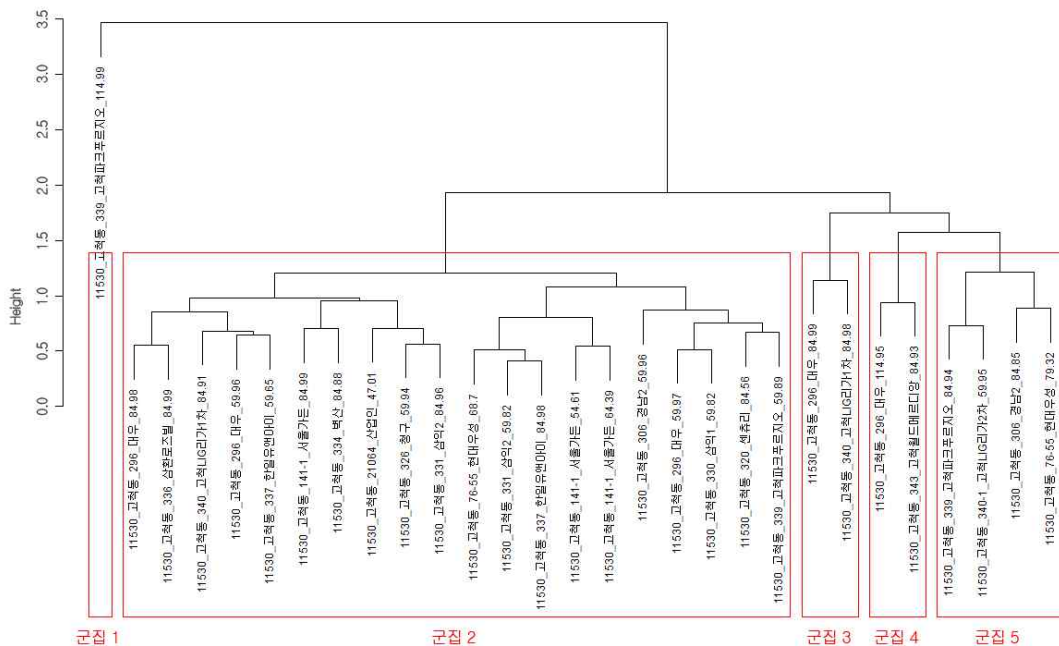
실측치	선형보간	선형회귀대체	합계
216,119	64,173	1,225,348	1,505,640
14,35%	4,26%	81,38%	100,00%

〈표 5〉 결측치 추정 결과

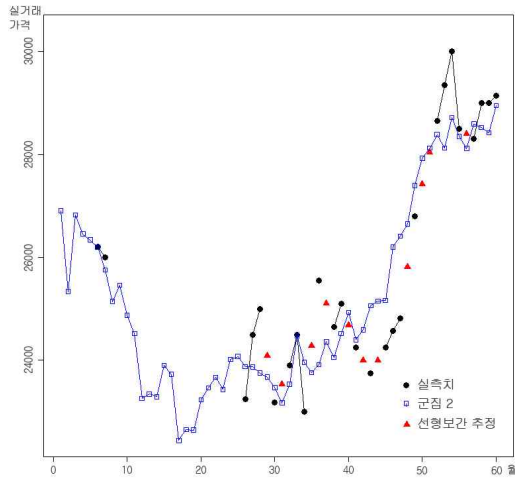
실측치	선형보간	선형회귀 대체	결측치	합계
216,119	64,173	46,674	1,178,674	1,505,640
14.35%	4.26%	3.10%	78.28%	100.00%

보다 자세한 설명을 위해 서울시 구로구 고척동 현대우성 아파트 68.7m²의 실거래가격 결측치 추정 결과를 살펴보자. 고척동 내 아파트 유형의 가격변동성으로 군집 분석을 실시한 결과, 총 다섯 개의 군집들 중에서 해당 아파트 유형은 군집 2에 포함됨을 확인할 수 있었다(〈그림 2〉). 이는 해당 아파트 유형의 가격변동성은 군집 2에 포함된 아파트 유형들의 가격변동성과

매우 유사하다는 사실을 시사한다. 따라서 군집 2의 가격 정보는 해당 아파트 유형의 자료 내에서 발생하는 세 개 이상의 결측치 추정을 위한 선형회귀대체법의 참고 자료로써 사용되어지게 된다. 이 사례에 대한 결측치 추정 결과는 〈그림 3〉과 같다. 여기서 동그라미는 실측치, 세모는 선형보간법을 이용해 추정한 값, 네모는 선형회귀대체법 추정에 사용되는 군집 2의 가격정보를 나타낸다. 〈그림 3〉을 통해 확인해볼 수 있듯이, 세 개 이상의 연속된 결측치의 경우, 가격 변동이 유사한 군집의 가격 정보를 이용한 선형회귀대체법으로 추정되기 때문에 이 때 발생하는 비선형적인 가격 변동은 군집 2의 정보를 통해 추정이 가능하다.



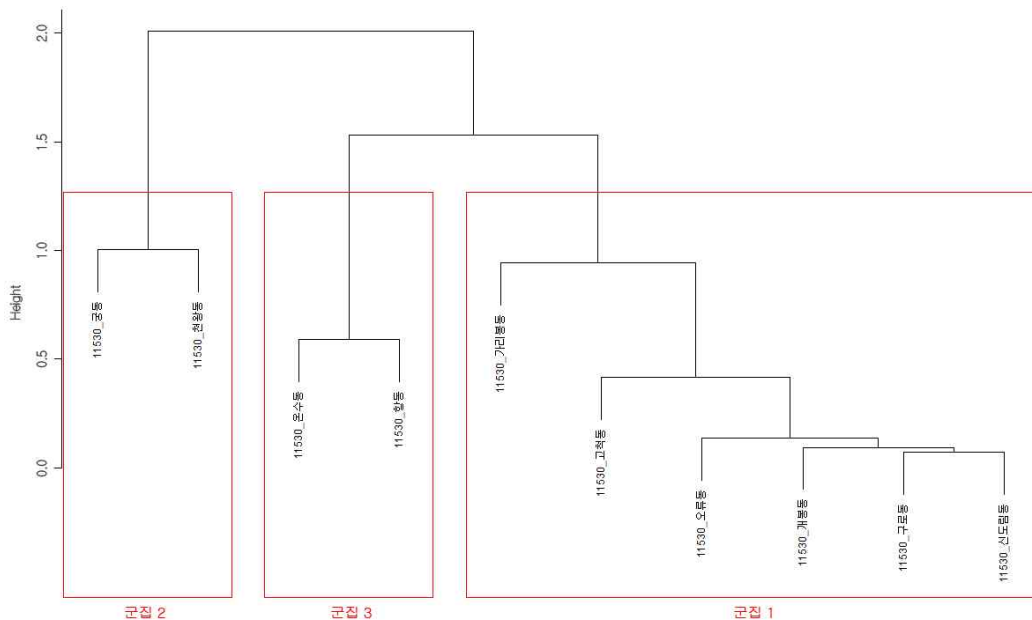
〈그림 2〉 행정동 내 아파트 유형의 가격변동성에 따른 군집 분석 결과 예시(고척동)



〈그림 3〉 실거래가격 추정 결과 예시
(고척동 현대우성 아파트 68.7㎡)

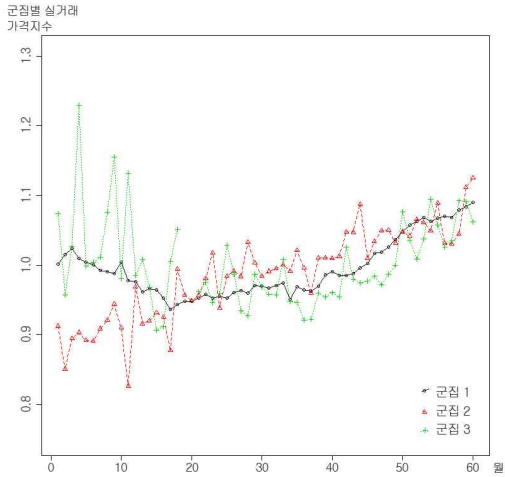
2. 행정구 단위의 군집별 아파트 실거래가격지수 도출 결과

결측치를 복원한 서울시 행정동의 실거래가격으로 군집 분석을 실시하여 행정구 단위의 군집별 아파트 실거래가격지수를 도출하였으며, 각 군집에 대한 상세 정보는 부록 1에 첨부하였다. 그 결과, 기존의 가격지수에서 고려하지 않았던 동일 지역 내의 가격 이질성을 파악할 수 있었다. 앞서 예시로 소개한 구로구의 경우를 살펴보면, 총 3개의 군집이 형성되었으며, 각 군집은 각각 2개, 2개, 6개의 동으로 이루어져있음을 확인할 수 있다(〈그림 4〉). 구로구의 군집별 아파트 실거래가격지수를 확인해본 결과, 각 군집의 가격변동성에 따른 이질성의 정도가 상당히 크다는 것을 확인할 수 있었다(〈그림 5〉). 각 군집 간의 가격 변동을 직접적으로 비교하기 위해 초기값으로 표준화하여 차이를 명확히 확인할 수 있었다(〈그림 6〉). 이는 단순 행정구 수준으로 실거래가격지수를 도출할 경우 놓칠 수 있

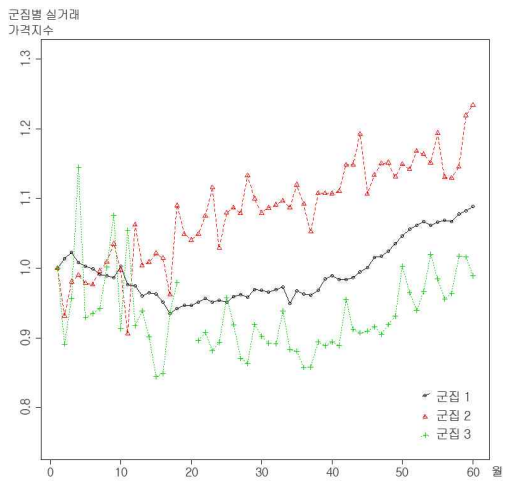


〈그림 4〉 행정구 내 아파트 유형의 가격변동성에 따른 군집 분석 결과 예시(구로구 고척동)

는 하위지역들의 특성들을, 행정동 단위의 군집 분석을 통해 파악할 수 있음을 보여준다.



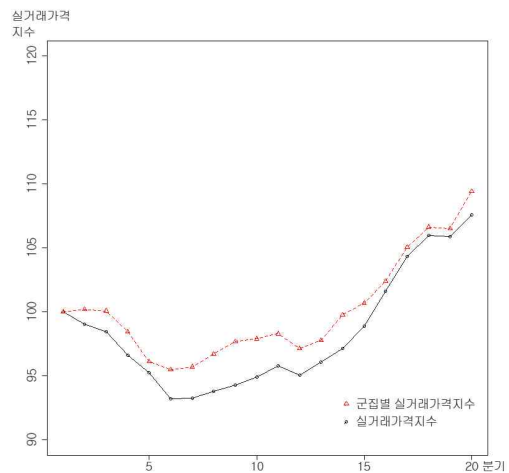
〈그림 5〉 행정구 단위 군집별 아파트 매매 실거래가격지수(구로구)



〈그림 6〉 초기값으로 표준화한 행정구 단위 군집별 아파트 매매 실거래가격지수(구로구)

마지막으로, 군집별 실거래가격지수의 실효성을 보이기 위해 기존의 실거래가격지수와와의 비교 분석을 실시하였다. 한국감정원에서 제공

하는 실거래가격지수는 시, 도 단위와 구 단위로 구분되는데, 전자의 경우에는 월별로 발표가 되는 반면 후자는 분기별로 작성이 되고 있다. 이에 본 연구에서는 행정구 내에 속한 전체 행정동들의 월별 가격을 평균하고, 이를 다시 분기별 평균으로 계산하였다. 직접적인 비교를 위해 본 자료상에서 1기인 2011년 7월의 초기 값으로 표준화를 시켜 100부터 가격변동이 시작하게끔 하였다. 그 결과, 두 지수의 전반적인 가격 변동성의 움직임이 매우 유사함을 확인할 수 있었다(〈그림 7〉). 실제로 구로구를 포함한 전체 행정구에서 산출된 두 지수의 통계적인 상관관계는 0.984로 매우 높게 나타났으며, 전반적인 실거래가격지수의 표준편차는 4.251, 군집별 실거래가격지수의 표준편차는 3.970으로 후자가 더 적은 값을 보였다(〈표 6〉). 군집을 이루고 있는 행정동의 수가 많은 군집일수록 행정구 단위의 실거래가격지수와 유사하여 대표성을 띤다고 할 수 있었고, 이질적인 양상을 보이는 군집 역시 효과적으로 확인할 수 있었다.



〈그림 7〉 실거래가격지수와와의 비교 결과(구로구)

〈표 6〉 실거래가격지수와 비교 분석 결과(구로구)

상관관계	실거래가격지수 표준편차	군집별 실거래가격지수 표준편차
0.984	4.521	3.970

〈부록 2〉에서 서울시의 모든 구에 대한 군집별 실거래가격지수를 비교할 수 있다. 〈부록 2〉의 용산구를 보면 다수를 차지하는 군집1(산천동, 이촌동, 원효로4가 외 7개동)이 용산구 전체의 가격지수와 0.97의 상관관계를 보이거나 군집3(신창동, 용문동, 문배동 외 6개 동)의 경우 0.69의 상관관계를 보여 다소 다른 가격추이를 보이는 큰 그룹이라고 할 수 있다. 특히 그 군집3에 포함된 동들의 가격변화가 군집2에 비해 보다 가격 편차가 적음을 알 수 있다.

V. 결 론

1. 결론 및 시사점

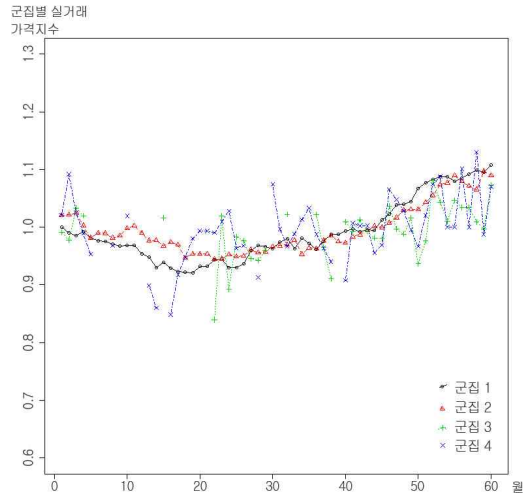
본 연구에서는 실거래가격 자료가 주는 자료의 신뢰성을 기반으로 군집 분석을 통해 기존의 지수에서 지적되어온 자료의 안정성과 지역 내 주택 가격의 이질성 문제를 극복할 수 있는 결측치 추정 방식과 새로운 지수 산출 방안을 제시하였다. 이를 위한 실증 사례로서 2011년 7월부터 2016년 6월까지 총 60개월 동안 서울특별시 지역 내에서 일어난 아파트 매매 실거래가격

자료를 이용해 분석을 실시하였다. 그 결과, 선형보간법과 선형회귀대체법을 이용해 각각 실측치의 30%와 21.6%의 자료활용률을 높일 수 있었다. 이를 토대로 도출한 행정구 단위 군집별 아파트 실거래가격지수의 실효성을 보이기 위해 기존의 실거래가격지수와 비교 분석을 실시한 결과, 두 지수의 상관관계는 매우 유사하다는 것을 확인할 수 있었다. 다만 전반적인 지수의 크기는 실거래가격지수가 군집별 실거래가격지수보다 낮게 나타났는데 이는 기하평균을 사용하는 반복매매모형의 특성이나 지수 산정에 사용된 샘플 차이에 기인하는 것으로 보인다. 군집별 실거래가격지수의 표준편차는 서울시 내 모든 구에서 실거래가격지수보다 다소 작게 나타났다. 이는 본 지수의 가격변동성이 상대적으로 더 작다는 특징을 보이고 있음을 알 수 있다.

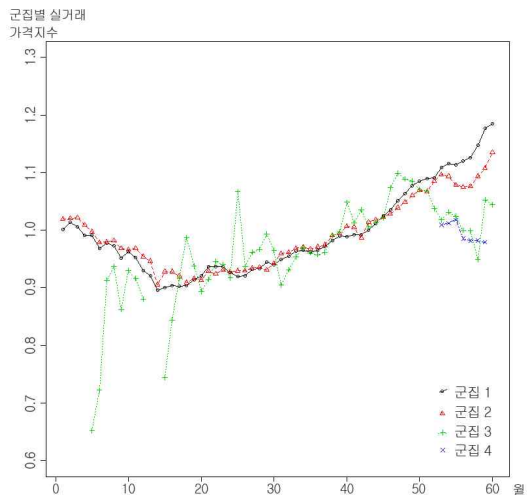
본 연구의 시사점은 다음과 같다. 먼저, 본 연구에서 제안한 선형보간법과 선형회귀대체법을 활용한 추정 방식은 실거래가격의 자료활용률을 효과적으로 높일 수 있다. 기존의 연구에서 지역이나 시기에 따른 거래량의 편차는 실거래가격자료 활용의 문제점으로 지적되어 왔다. 이러한 측면에서 단순 선형보간법이나 유사 가격변동성을 띄는 군집의 가격 정보를 이용해 추정하는 선형회귀대체법은 실거래가격지수의 결측치로 인해 발생하는 자료의 안전성 문제에 대한 해결 방안이 될 수 있다. 보통 특정 아파트(혹은 지역)의 가격을 추정할 때, 유사한 조건을 가진

인근 아파트의 정보를 이용하는 것처럼, 가격변동성이 유사한 아파트 군집의 가격 정보를 활용하는 이 방식은 보다 직관적이고 객관적인 추정 방식이라 할 수 있다.

뿐만 아니라 군집 분석을 이용해 도출한 군집별 아파트 실거래가격지수는 가격지수의 사용 가치를 높일 수 있다. 군집별 아파트 실거래가격지수의 가장 큰 장점은 동일 지역 내의 가격 이질성을 효과적으로 파악할 수 있다는 것이다. 따라서 가격중심으로 작성된 본 가격지수의 가격 이질성에 대한 결과는 아파트 가격에 대한 보다 정확하고 세부적인 정보를 제공할 수 있다. 성동구(〈그림 8〉)와 강남구(〈그림 9〉) 사례에서 확인할 수 있듯이, 군집별 실거래가격지수 정보는 전반적인 가격 흐름을 주도하는 지역은 어디인지, 해당 행정구 내에서 이질적인 변동성을 보이는 지역이 무엇인지 혹은 어느 군집의 거래량이 자주 일어나는지를 쉽게 파악할 수 있게 해준다. 또한 이 군집별 실거래가격지수 정보는 관련된 다양한 연구 분야에서 활용될 수 있다. 만약 분석의 대상이 되는 지역의 가격지수가 이질적이라면, 즉, 지역 내의 가격 변동성의 일관성이 높지 않다면, 이에 상응하는 변수를 찾기가 쉽지 않고 또 그 변수의 설명력 또한 크지 않을 것이다. 문제는 기존의 가격지수가 동일 지역 내에서 발생하는 하위 지역 간의 가격 이질성을 보여주지 못한다는 것이다. 그렇기 때문에 이러한 접근 방식은 실거래가격지수의 활용 가치를 더 높일 수 있는 방안으로 사용될 수 있다.



〈그림 8〉 행정구 단위 군집별 아파트 매매 실거래가격지수 활용 사례(성동구)



〈그림 9〉 행정구 단위 군집별 아파트 매매 실거래가격지수 활용 사례(강남구)

마지막으로 본 연구는 우리나라 부동산 가격 정보 체계 구축에 기여할 수 있다. 이상적인 부동산 가격지수는 정기적으로 측정가능하고, 표본

의 모집단에 대한 대표성을 띄어야 하며, 타지수와 비교 가능하고, 지수작성의 연속성과 시계열의 기간, 관측빈도를 만족해야 한다(Arthur, 2003). 가격지수의 다양성이 국가의 부동산 가격 정보 체계 구축에 있어 중요한 원칙 중 하나라는(Borio and Lowe, 2002) 관점에서 이 가격지수는 이를 모두 만족함과 동시에 기존의 가격지수와는 차별화된 정보를 제공할 수 있다는 장점이 있다. 따라서 앞서 언급한 지수 설계의 목적대로 이 가격지수는 소비자의 경제활동과 정부기관의 정책수립 활동을 위한 보다 정확하고 구체적인 가격변동성 정보를 제공할 수 있다.

2. 한계점

본 연구에서는 분석의 특성상 몇 가지 한계점을 안고 있다. 먼저, 자료 추가 시 발생하게 되는 문제가 고려되지 않았다. 군집 분석의 경우, 자료가 추가되면 기존에 도출한 군집이 바뀔 수 있기 때문에 자료 업데이트에 대한 기준이 마련되어야 한다. 가령 자료가 추가될 때 군집을 수정할 것인지, 만약 그렇다면 군집 업데이트의 주기를 어떻게 설정한 것인지에 대한 고민이 필요하다. 또한, 실증 분석이 서울에만 한정되어 있다. 본 연구에서 실증 사례로서 선정한 서울의 경우, 국내에서 아파트 거래량이 가장 많은 지역이기 때문에 상대적으로 실증 분석에 용이하다고 볼 수 있다. 물론 본 연구에서 제시하는 지수 산정 방식은 일반화할 수 있는 접근법이지만 본 연구에서 제시하고 있는 결측치 추정이나 군집 분석의 특성상 거래량이 적은 지역에서의 활용 가능성 여부가 추가적으로 고려되어야 한다.

참고문헌

1. 구성환 · 장훈 · 송민경, 2010, 「인자분석과 군집 분석을 이용한 수도권 지역의 지역유형 구분에 관한 연구」, 한국지형공간정보학회 추계학술대회, 2010. 9., 43-48.
2. 고필송 · 박창수, 2010, 「아파트 실거래가 지수를 이용한 부동산 시장 평가」, 『한국전자통신학회지』, 4(1): 63-68.
3. 김경희 · 김영효, 2015, 「모델의 불확실성을 반영한 아파트가격지수 예측 모형 연구: BMS, BMA를 중심으로」, 『부동산분석』, 1(1): 27-49.
4. 김선진, 2015, 「중국 부동산가격의 민간소비에 대한 부의 효과」, 『중소연구』, 39(3): 83-110.
5. 김성록, 2013, 「부동산산업의 연계구조 및 생산유발 효과 추이 분석」, 『부동산학연구』, 19(3): 117-128.
6. 김진희, 2014, 「우리나라 아파트의 특성과 아파트 가격: 분위회귀분석」, 『산업경제연구』, 27(1): 173-195.
7. 남수현 · 이장우, 2009, 「한국 부동산 선물가격지수 개발에 관한 연구」, 『금융공학연구』, 8(3): 75-103.
8. 안성남, 2010, 「연쇄지수와 지수의 이론」, 『통계연구』, 5(2): 1-22.
9. 이용만, 2007a, 「특성가격함수를 이용한 주택가격지수 개발에 관한 연구: 시간변동계수모형에 의한 연쇄지수」, 『부동산학연구』, 13(1): 103-125.
10. 이용만, 2007b, 「주택가격지수의 목적과 방법을 둘러싼 쟁점: 실거래가격에 기초한 지수를 중심으로」, 『부동산학연구』, 13(3): 147-167.
11. 이용만 · 이상한, 2008, 「국민은행 주택가격지수의 평활화 현상에 관한 연구」, 『주택연구』, 16(4): 27-47.
12. 이용민 · 노창래 · 이창무, 2010, 「실거래가를 이용한 오피스텔 호별 가격결정 모형: 강남구를 중심으로」, 『부동산 · 도시연구』, 3(1): 41-59.
13. 이창무 · 김병욱 · 이현, 2002, 「반복매매모형을 활용한 아파트 매매가격 지수」, 『부동산학연구』, 8(2): 1-19.

14. 이창무 · 김진유 · 이상영, 2005, 「공동주택 실거래가 지수 산정에 관한 연구: 서울시 아파트 시장을 중심으로」, 『국토계획』, 40(4): 121-134.
15. 이창무 · 김종현 · 박한 · 김형태, 2008a, 「실거래가를 활용한 비아파트 중위수 매매가격 지수」, 『부동산학연구』, 14(3): 5-16.
16. 이창무 · 배익민, 2008b, 「시세가격을 활용한 아파트 실거래가 반복매매지수 산정」, 『부동산학연구』, 14(2).
17. 정승영 · 도희섭, 2010, 「상가권리금의 군집분석을 이용한 시장세분화에 관한 연구: 서울시를 중심으로」, 『대한부동산학회지』, 28(2): 291-309.
18. 최필근 · 이내성 · 박소현 · 정미옥 · 정선경, 2008, 「국가통계의 품질제고를 위한 방법론 연구: Imputation 기법과 행정자료 활용을 중심으로」, 통계개발원 연구보고서.
19. 하정순 · 정삼석 · 정상철, 2017, 「부동산투자스타일 세분화에 따른 선호속성과 구매의도에 관한 연구」, 『부동산학보』, 68: 82-96.
20. Arthur, S. V., 2005, "Obtaining real estate data: Criteria, difficulties and limitations," *Real estate indicators and financial stability*, 21: 63-69.
21. Anderberg, M. R., 1973, "Cluster analysis for applications," Academic Press, New York.
22. Borio, C., and Philip, L., 2002, "Asset prices, financial and monetary stability: Exploring the nexus," *BIS working paper*, No.114, July.
23. Case, B., Pollaskowski, H. O., Watcher, S. M., 1991, "On choosing among house price index methodologies," *Journal of AREUEA*, 19: 286-307.
24. Case, K., and Shiller, R., 1989, "The efficiency of the market for single-family homes," *American Economic Review*, 79: 125-137.
25. Chinloy, P., Cho, M., Megbolugbe, I. F., 1997, "Appraisals, transaction incentives, and smoothings," *Journal of Real Estate Finance and Economics*, 14(1): 89-111.
26. Fisher, J., Geltner, D., Webb, R., 1994, "Value indices of commercial real estate: A comparison of index construction methods," *Journal of Real Estate Finance and Economics*, 9(2): 137-166.
27. Řezanková, H., 2014, "Cluster analysis of economic data," *Journal of Statistika*, 94(1): 73-86.
28. Sarle, W. S., 1983, "Cubic Clustering Criterion", SAS Technical Report A-108, Cary, NC: SAS Institution.

논문접수일 : 2016년 10월 4일

심사(수정)일 : 1차 2016년 11월 3일

2차 2017년 5월 12일

게재확정일 : 2017년 5월 26일

국문초록

본 연구에서는 군집 분석을 도입하여 실거래가격 자료의 결측 문제로 인해 나타나는 실거래가격지수 산정 과정에서의 불안전성 문제와 지역 간의 이질성 문제를 보완할 수 있는 새로운 아파트 실거래가격지수 산정 방식을 제시하였다. 실거래가격 자료상에서 발생하는 결측치의 문제를 해결하기 위해 연속된 결측치 유형에 따라 선형보간법과 군집 정보를 활용한 선형회귀대체법으로 결측치를 추정하였으며, 지역 간 이질성 문제를 해결하기 위해 행정동 실거래가격 정보로 행정구 단위의 군집별 아파트 실거래가격지수를 도출하였다. 2011년 7월부터 2016년 6월까지 총 60월 동안에 서울에서 이루어진 아파트 매매 실거래가격 자료를 이용해 분석하였으며, 그 결과, 군집별 아파트 매매 실거래가격지수가 자료 활용성과 지역 간 이질성 측면에서 더 적합함을 확인할 수 있었다.

주제어 : 실거래가격지수, 군집 분석

부록 1. 서울시 내 행정구별 군집 정보

지역명 (동수)	군집 번호	법정동 코드	지역명 (동수)	군집 번호	법정동 코드
종로구	1	수송동, 당주동, 청운동, 견지동	종구	1	남대문로5가, 오장동
	2	동송동, 교북동		2	신당동, 묵정동, 흥인동, 인현동2가, 남산동2가
	3	구기동, 명륜2가, 필운동, 연건동		3	남창동, 장충동1가
	4	무악동, 창신동, 사직동, 익선동, 송인동, 평창동, 신영동, 내수동, 인의동, 충신동, 명륜1가		4	황학동, 순화동, 종림동, 만리동1가, 충무로4가
	5	부암동, 신교동		5	을지로5가, 예장동, 만리동2가
	6	이화동		6	회현동1가, 회현동2가
	7	행촌동			
용산구	1	산천동, 이촌동, 원효로4가, 효창동, 도원동, 이태원동, 한남동, 서빙고동, 보광동, 한강로2가	성동구	1	금호동1가, 금호동3가, 금호동4가, 마장동, 옥수동, 응봉동, 하왕십리동, 행당동, 금호동2가, 상왕십리동
	2	동빙고동, 동자동		2	성수동1가, 성수동2가, 도선동, 송정동, 사근동
	3	신창동, 용문동, 문배동, 신계동, 용산동5가, 한강로3가, 후암동, 한강로1가, 원효로1가		3	용답동
	4	용산동2가		4	홍익동
	5	원효로3가			
	6	청암동			
	7	청파동3가			
광진구	1	광장동, 구의동, 자양동, 군자동	동대문구	1	답십리동, 이문동, 장안동, 전농동, 휘경동, 회기동
	2	화양동, 능동		2	신설동
	3	중곡동		3	용두동, 제기동, 청량리동
중랑구	1	망우동	성북구	1	길음동, 돈암동, 석관동, 정릉동, 종암동, 하월곡동, 동소문동4가, 동소문동7가, 삼선동2가, 안암동1가, 보문동3가, 동소문동5가, 삼선동3가, 삼선동4가
	2	면목동, 묵동, 상봉동, 신내동, 중화동		2	보문동4가, 동선동1가
				3	동선동3가, 동선동4가
				4	동선동5가
				5	보문동2가, 성북동
				6	상월곡동, 장위동, 보문동6가
				7	안암동3가, 안암동4가

부록 1. 서울시 내 행정구별 군집 정보(계속)

지역명 (동수)	군집 번호	법정동 코드	지역명 (동수)	군집 번호	법정동 코드
강북구	1	미아동, 번동, 수유동	도봉구	1	도봉동, 쌍문동, 창동
	2	우이동		2	방학동
노원구	1	공릉동	은평구	1	구산동, 불광동, 수색동, 신사동, 응암동, 진관동, 갈현동, 역촌동, 증산동
	2	상계동, 월계동, 중계동, 하계동		2	녹번동
				3	대조동
서대문구	1	남가좌동, 북가좌동, 연희동, 홍은동, 홍제동, 총정로3가, 천연동, 현저동, 대현동, 냉천동, 영천동, 북아현동	마포구	1	공덕동, 도화동, 삼암동, 성산동, 신공덕동, 중동, 창전동, 용강동, 마포동, 대흥동, 염리동, 노고산동, 신수동, 망원동, 연남동, 토정동
	2	창천동		2	아현동, 신정동, 상수동
	3	총정로2가		3	서교동
	4	합동		4	하중동
				5	합정동
				6	현석동
양천구	1	목동, 신정동	강서구	1	가양동, 내발산동, 등촌동, 방화동, 염창동, 화곡동, 마곡동
	2	신월동		2	공항동
구로구	1	개봉동, 고척동, 구로동, 신도림동, 오류동, 가리봉동	금천구	1	가산동
	2	천왕동, 궁동		2	독산동, 시흥동
	3	온수동, 향동			
영등포구	1	당산동4가, 당산동5가, 대림동, 문래동3가, 신길동, 양평동2가, 양평동3가, 여의도동, 영등포동, 영등포동8가, 당산동2가, 당산동3가, 당산동, 문래동5가, 문래동6가, 양평동1가, 양평동4가, 양평동5가, 당산동1가, 문래동2가, 문래동4가	동작구	1	노량진동, 대방동, 사당동, 상도동, 신대방동, 흑석동, 본동
	2	당산동6가		2	동작동
	3	영등포동7가, 도림동, 양평동6가		3	상도1동
	4	영등포동1가			
	5	영등포동2가			
	6	영등포동3가			
	7	영등포동4가			
	8	영등포동5가			
동작구	1	노량진동, 대방동, 사당동, 상도동, 신대방동, 흑석동, 본동	관악구	1	남현동
	2	동작동		2	봉천동, 신림동
	3	상도1동		3	

부록 1. 서울시 내 행정구별 군집 정보(계속)

지역명 (동수)	군집 번호	법정동 코드	지역명 (동수)	군집 번호	법정동 코드
서초구	1	내곡동	강남구	1	개포동, 삼성동, 수서동, 역삼동
	2	반포동, 방배동, 서초동, 양재동, 잠원동, 신원동		2	논현동, 대치동, 도곡동, 압구정동, 일원동, 청담동, 신사동, 자곡동
	3	우면동		3	세곡동
송파구				4	율현동
	1	가락동, 거여동, 문정동, 방이동, 송파동, 신천동, 오금동, 잠실동, 장지동, 풍납동, 마천동	강동구	1	강일동, 길동, 둔촌동, 성내동, 암사동, 천호동
	2	삼전동		2	고덕동, 명일동, 상일동
	3	석촌동			

부록 2. 군집별 실거래가격지수와 실거래가격지수와의 비교 분석 결과

	구 단위 실거래가격지수 (a,b)	군집별 실거래가격지수 (b)	군집 1 (c,d,e)	군집 2 (c,d,e)	군집 3 (c,d,e)	군집 4 (c,d,e)	군집 5 (c,d,e)	군집 6 (c,d,e)	군집 7 (c,d,e)	군집 8 (c,d,e)
종로구	0.92, 3.82	2.9	4, 6.42, -0.09	2, NA, NA	4, 4.36, 0.7	11, 3.02, 0.95	2, NA, NA	1, NA, NA	1, NA, NA	-
중구	0.9, 4.73	2.41	2, NA, NA	5, 3.77, 0.9	2, 5.51, -0.47	5, 2.65, 0.91	3, NA, NA	2, NA, NA	-	-
용산구	0.98, 5.26	3.28	10, 4.69, 0.97	2, NA, NA	9, 3.27, 0.69	1, 5.47, -0.25	1, NA, NA	1, NA, NA	1, NA, NA	-
성동구	1, 5.77	4.69	10, 5.46, 0.99	5, 4.14, 0.97	1, 3.7, 0.56	1, 4.51, 0.66	-	-	-	-
광진구	0.99, 5.03	3.7	4, 4.23, 0.99	2, 4.5, 0.67	1, 2.65, 0.69	-	-	-	-	-
동대문구	1, 4.96	4.22	6, 4.6, 1	1, 4.3, 0.82	3, 3.84, 0.97	-	-	-	-	-
종랑구	1, 4.31	3.79	1, 2.75, 0.87	5, 4.08, 0.99	-	-	-	-	-	-
상복구	1, 5.47	4.6	14, 5.09, 0.99	2, NA, NA	2, NA, NA	1, NA, NA	2, NA, NA	3, 4.8, 0.97	2, 4.12, 0.35	-
강북구	0.99, 4.92	3.01	3, 3.53, 0.99	1, 2.15, 0.55	-	-	-	-	-	-
도봉구	0.99, 4.83	3.34	3, 3.81, 0.99	1, 2.3, 0.77	-	-	-	-	-	-
노원구	1, 5.52	4.41	1, 2.25, 0.64	4, 5.09, 1	-	-	-	-	-	-
은평구	0.99, 4.1	3.5	9, 3.86, 0.99	1, 2.99, 0.77	1, 1.75, 0.75	-	-	-	-	-
서대문구	0.99, 4.77	4.31	12, 4.71, 0.99	1, 2.97, 0.52	1, 3.03, 0.76	1, 3.72, 0.6	-	-	-	-
미포구	0.99, 5.54	4.23	16, 4.76, 0.99	3, 4.45, 0.95	1, 3.34, 0.71	1, 6.7, 0.1	1, 3.5, 0.13	1, 4.26, 0.65	-	-
양천구	1, 5.52	4.4	2, 4.99, 1	1, 3.3, 0.97	-	-	-	-	-	-
강서구	1, 6.74	6.6	7, 7.21, 1	1, 4.25, 0.17	-	-	-	-	-	-
구로구	0.98, 4.52	3.97	6, 4.16, 1	2, 6.89, 0.56	2, 4.55, 0.69	-	-	-	-	-
금천구	0.99, 5.24	4.35	1, 3.03, 0.92	2, 5.08, 0.99	-	-	-	-	-	-
영등포구	0.99, 5.03	4.67	21, 5.05, 1	1, NA, NA	3, 3.61, 0.92	1, 5.96, 0.57	1, NA, NA	1, NA, NA	1, NA, NA	1, NA, NA
동작구	0.99, 4.63	4.06	7, 4.26, 1	1, 4.69, 0.87	1, 4.03, 0.2	-	-	-	-	-
관악구	0.99, 4.41	3.51	1, 2.92, 0.87	2, 3.96, 0.99	-	-	-	-	-	-
서초구	0.99, 7.33	6.23	1, NA, NA	6, 6.09, 0.99	1, 8.14, 0.84	-	-	-	-	-
강남구	1, 7.98	6.12	4, 7.5, 0.99	8, 5.9, 1	1, NA, NA	1, NA, NA	-	-	-	-
송파구	0.99, 5.07	4.01	11, 4.41, 0.99	1, NA, NA	1, 2.13, 0.44	-	-	-	-	-
강동구	0.98, 4.47	4.43	6, 4.12, 0.93	3, 5.43, 0.97	-	-	-	-	-	-

a : 구 단위 실거래가격지수와 행정구 단위의 군집별 실거래가격지수와의 상관관계, b : 표준편차, c : 개별 군집의 행정동 수,

d : 개별 군집의 표준편차, e : 개별 군집의 실거래가격지수와 구 단위 실거래가격지수와의 상관관계.

* NA : 결측치가 많아 추정이 불가능한 경우.