부동산학연구 제19집 제4호, 2013. 12, pp. 27～40

Journal of the Korea Real Estate Analysts Association
Vol.19, No.4, 2013. 12, pp. 27～40

Quantile Regression을 이용한 반복매매지수 산정에 관한 연구*

A Repeat Sales Price Index Using Quantile Regression

이 창 무 (Lee, Changmoo)**
류 강 민 (Ryu, Kangmin)***
김 지 연 (Kim, Jiyeon)****

< Abstract >

This study develops a repeat sales price index using quantile regrssion. Quantile regression(QR) allows us to mitigate the problem of outliers comparing with the OLS estimation commonly used. This advantage can be highligted when we need to generate a price index for a thin market where there are not enough samples.

In order to verify the effect, we constructed price indices for four different size groups of apartment condominiums in Seoul. A bigger group has a smaller sample size. The estimation results show that the quantile regression method is effective to smooth the peak points in indices generated by the OLS estimation. This stabilizing effect is obvious in the case of large size apartment condominiums with fewer samples.

We chose three evaluation indices to compare the performance of the OLS and QR indices including statistical reliability index(mean standard error), stability index newly developed, and sensitivity index(signal-to-noise ratio). Overall evaluation suggests that QR indices allow us much gain in stability without much loss in statistical reliability comparing with OLS indices.

주 제 어 : 반복매매지수, 분위회귀, 아파트, 주택가격지수
key word : Repeat-Sales-Index, Quantile Regression, Apartment, Housing Price Index

* 본 논문은 2012년 한국주택학회 정기학술대회에서 발표한 「Quantile regression을 이용한 반복매매지수 산정에 관한 연구」의 일부를 수정·보완한 것이며, 2012년 한양대학교 교내연구비 지원으로 연구되었음(HY-2012-G).

** 한양대학교 도시공학과 교수, changmoo@hanyang.ac.kr (주저자)
*** 한양대학교 도시공학과 박사과정, locsword@hanmail.net (교신저자)
**** 한양대학교 도시공학과 박사과정, kjy0510@hanmail.net

I. 서론

현재 국토해양부는 반복매매모형을 이용하여 아파트 실거래가 지수를 발표하고 있으며, 서울의 경우에는 5대 생활권과 생활권 내 면적별로 세분화된 지수가 발표되고 있다. 그러나 거래되는 빈도가 낮은 주택 하부시장의 경우 반복거래 표본을 확보하기 어려워 지수의 변동성이 크게 나타나고, 통계적인 신뢰도가 확보되지 못하는 문제가 발생한다. 국토해양부에서는 같은 아파트가 아니어도 몇 가지 조건을 충족하면 같은 아파트의 거래로 간주하는 동일 주택가정을 도입하여 이러한 문제를 해결하고자 하였다.

이러한 시도에도 불구하고 표본 수 부족으로 인한 문제는 여전히 존재한다. 표본 수가 적을 경우에는 표본 중 일부의 가격변동이 이상치로 작용하여 지수 값에 크게 영향을 미쳐 불안정한 추세의 지수가 산정된다. 이러한 문제를 최소화하기 위해 일정 기준을 정하여 이상치 값이라 판단되는 표본을 제거하는 방법이 사용되고 있으나 이 경우 표본수가 더욱 감소하여 지수의 안정성이 더욱 악화될 수도 있다. 또 어떤 기준으로 이상치를 제거할 것인가에 대한 논란에서 벗어나기도 힘들다.

최근 이상치의 영향을 줄일 수 있는 하나의 방안으로 분위회귀(Quantile Regression)가 논의되고 있다. 분위회귀를 이용해 지수를 산정하는 방법은 Coulson and McMillen(2007), McMillen and Thorsnes(2006) 등에 의해 제시된 것으로, 오차의 합을 최소화시키는 평균 개념의 계수를 추정하는 것이 아니라 절대편차의 합을 최소화하는 중위수 개념의 계수를 추정하는 것이다. 따라서 분위회귀의 경우 별다른 이상치 제거작업 없이도 이상치의 영향을 줄일 수 있는 추정방식이 된다.

본 연구에서는 이러한 분위회귀를 이용하여 서울시 아파트 실거래가를 이용하여 반복매매지수를 산정하고, OLS를 이용하여 산정한 반복매매지수와의 비교를 통해 지수산정방식의 차이에 따른 특징에 대해 분석하고자 한다. 또한 동일한 자료를 이용하여 지수산정방식을 선택할 때 이용될 수 있는 비교지표를 개발하여 두 가지 지수산정방식에 대한 합리적인 평가를 시도하고자 한다.

II. 선행연구

지금까지 실거래가격에 기초하여 부동산 가격지수를 산정하기 위한 방법 중 하나로 반복매매지수의 이용이 확대되고 있다. 반복매매지수는 같은 부동산이 두 번 이상 거래되었을 경우 두 시점의 동일 부동산에 대한 거래가격의 비율을 이용하여 지수를 추정하는 방법이다. 이 방법은 Bailey et al.(1963)를 시작으로 Case(1986), Case and Shiller(1987), Shiller(1991), Goetzmann(1992) 등에 의해 발전되어 왔다. 국내의 경우 2006년 실거래가 자료가 축적되기 이전에는 시세자료를 가공하여 아파트 가격지수와 월세지수를 산정한 사례가 있으며, 실거래 자료가 공개되면서는 실거래가를 이용한 반복매매지수가 산정되고 있다(이창무 외, 2002; 이창무 외, 2003; 이창무 외, 2005; 이창무 외, 2007).

반복매매지수는 지수산정을 위해 많은 시간과 비용을 투자해야 하는 헤도닉 가격지수와 달리, 반복적으로 거래된 주택의 거래시점과 가격에 대한 정보만 있으면 지수를 산정할 수 있는 장점을

가지고 있다. 또한 지수를 산정하는데 있어서도 개별주택의 평균적인 가격 변동률을 추정하는 동일가중 방식과 총자산의 변동을 추정하는 가치가중 방식으로 지수를 산정할 수 있기 때문에, 목적에 따른 적절한 지수 선택이 가능하다는 장점도 있다.

그러나 반복매매지수는 기하평균이 가져오는 지수의 저평가 문제, 반복거래 된 자료만을 사용함으로써 발생하는 자료의 비효율성 문제 등 적지 않은 한계 또한 지니고 있다(Case and Shiller, 1987; Shiller, 1991; Goetzmann, 1992; Goetzmann and Peng, 2002; Peng, 2002; 이창무·배익민, 2008; 류강민 외, 2009, 류강민·이창무, 2011). 특히 Clapp and Giaccotto (1999)는 반복매매지수의 경우 반복 거래된 표본만이 지수 산정에 활용되므로 표본 편의가 발생한다는 분석결과를 제시하였다. Shiller(1991) 역시 리모델링 등을 통한 주택특성의 변화가 가격상승분으로 반영되어 표본의 편의가 존재할 수 있음을 언급하였고, 거래기간에 따라 가중치를 부여하는 가중반복매매모형을 제안하여 이러한 문제를 해결하고자 하였다.

최근에는 Coulson and McMillen(2007)과 McMillen and Thorsnes(2006)가 이러한 표본의 편의 문제를 보완할 수 있는 분위회귀(Quantile regression) 반복매매지수를 제안하였다. 이들은 평균이 표본의 분포와 이상치에 의해 민감하게 반응하므로 OLS를 이용한 지수 추정보다는 이상치의 영향을 덜 받는 중간값을 추정하는 분위회귀가 보다 나은 대안이 될 수 있다고 제안하였다. 실제로 McMillen and Thorsens(2006)은 1993년부터 2002년까지 시카고 지역을 대상으로 OLS와 분위회귀를 이용하여 지수를 추정하였고, 그 결과 분위회귀를 이용한 지수가 OLS를 이용한 지수보다 낮게 추정되었으며, 보다 안정적이라는 결과를 도출하였다. 또한 이들은 리모델링으로 인한 가격상승의 효과와 표본의 편의를 분위회귀가 보정하는 효과를 가지고 있음을 주장하였다. 최근 Donald Epley(2012)는 자연적 재해 등 시장 외부의 충격이 가격변화를 가져오는 경우에는 반복매매모형을 통해 외부충격 전후의 시장을 정확히 평가하기 어렵다고 주장하였고, 이에 대한 대안으로 가격변동전후 시장가격의 중위값을 비교할 것을 제안하였다.

이에 본 연구는 반복매매지수의 한계로 언급되었던 표본의 편의, 또는 이상치 문제를 해결할 수 있는 방안으로 분위회귀를 이용하여 서울시를 대상으로 지수를 산정하고자 한다. 지수는 반복매매모형을 이용하였으며, OLS와 분위회귀 지수를 모두 추정하여 지수 간의 비교 및 함의를 도출하고자 하였다.

III. 분석모형

1. 반복매매지수

반복매매지수는 2회 이상 거래된 주택을 대상으로 첫 번째 거래시점과 두 번째 거래시점의 가격변동률을 이용하여 지수를 산정하는 방식이다. 로그형태의 헤도닉 가격함수를 가정하면 동일주택의 거래시점 간 가격 차이는 식(1)의 최우변과 같이 주택특성의 차이와 거래된 시점 간 거시요인의 차이로 구성된다.

$$\ln\left(\frac{V^s}{V^f}\right)_i = \ln\frac{V_i^s}{V_i^f} = \left(\sum_{l=1}^{k}\beta_{l,s}\ln X_{l,i}^s + M_s + \epsilon_i^s\right) - \left(\sum_{l=1}^{k}\beta_{l,f}\ln X_{l,i}^f + M_f + \epsilon_i^f\right) \quad (1)$$

여기서 $V_i^{s(f)}$ 는 두 번째(첫 번째) 거래된 주택가격, $X_{l,i}^{s(f)}$ 는 l 주택특성, $\beta_{l,s(f)}$는 l 주택특성의 한계가치, $M_{s(f)}$ 는 상수항으로 표현할 수 있는 주택특성을 제외한 거시요인이고, $\epsilon_i^{s(f)}$는 오차항이다.

두 시점의 거래된 주택이 동일함으로 주택특성의 변화가 없다고 가정할 경우 식 (1)은 식 (2)와 같이 단순화된다. 따라서 동일주택의 거래가격 변동률은 두 시점 간 각 주택특성의 한계가치의 변동($\Delta\beta_{l,s-f}$) 에 따른 요인와 거시요인의 변화에 따른 요인 두 가지에 의해 발생하게 된다.

$$\ln\frac{V_i^s}{V_i^f} = \sum_{l=1}^{k}\Delta\beta_{l,s-f}\ln X_{l,i} + (M_s - M_f) + \epsilon_i \quad (2)$$

이 두 가지 요인 즉 거시경제요인과 고정특성의 한계가치변동에 대한 효과를 포괄하는 효과를 r 로 표현하면 두 시점 s 와 f 간 시장 내 모든 부동산에 동일하게 발생한다고 가정하면 다음 식과 같이 표현할 수 있다.

$$\ln\frac{V_i^s}{V_i^f} = r_s - r_f + \epsilon_i \quad (3)$$

여기서 $r_{s(f)}$는 일정 시점 0 를 기준으로 산정되는 상대적인 가격변동률이 된다. 식 (3)을 기준시점 1 를 제외한 관측된 모든 시점을 대표하는 시점더미 D_t 도입하여 표현하면, 첫 번째 거래인 경우 −1, 두 번째 거래인 경우 1, 그 외 0인 값을 대입할 때 식 (4)와 같이 표현할 수 있다.

$$\ln\left(\frac{V^s}{V^f}\right)_i = \sum_{t=2}^{T}\beta_t D_t + \epsilon_i \quad (4)$$

여기서 T 는 관측자료의 마지막 시점이고, 시점더미 D_t 의 추정계수인 β_t는 기준 시점 대비 t 시점의 가격비 로그값의 추정치가 된다. 추정된 $\hat{\beta}_t$ 을 이용해 다음과 같은 과정을 통해 기준시점이 0 인 경우의 가격지수로 산정된다.

$$I_t = \exp(\hat{\beta}_t) \times 100 \quad (5)$$

2. 분위회귀(Quantile Regression)

분위와 분위수의 정의는 다음 식 (6) 이하와 같이 표현할 수 있다. 식 (6)에서 임의의 변수 Y가 y 보다 작거나 같을 때의 확률을 $F(y)$라 하면 $F(y)$는 변수 Y의 확률분포함수에서의 한 지점에서의 분위가 된다.

$$F(y) = \mathrm{P}rob(Y \le y) \quad (6)$$

이 때 확률밀도함수 $F(y)$는 0과 1 사이의 값을 가지며, y값이 클수록 1에 가까워지며, y값이 작을수록 0에 가까워진다. 또한 $F(y)$ 값 중 하나인 γ는 변수 Y의 누적확률인 분위라 표현할 수 있다. 예를 들어 γ이 0.2이면, 변수 Y를 작은 것부터 큰 순으로 차례로 늘어놓았을 때 하위 20% 되는 분위를 의미한다. 특정 γ에 대해 γ보다 크거나 같은 $F(y)$가 여러 개 존재할 수 있는데, 이 중 최대하계(infimum)[1]를 만족하는 $F(y)$를 도출

1) 하계란 집합의 모든 원소들보다 작거나 같은 수를 의미하며, 최대하계란 하계 중 가장 큰 수를 의미한다. 예를 들어 (-1,1)인 집합에서 하계는 -3, -20, -1000 등이 될 수 있으며, 최대하계는 -1이 된다. 최대하계는 집합에 해당하지 않아도 되기 때문에 최소값과는 다르다.

하는 y를 γ분위의 분위수 $Q(\gamma)$라 정의한다.

$$Q(\gamma) = inf\{y : F(y) \geq \gamma\}, here\ 0 \leq \gamma \leq 1 \quad (7)$$

변수 Y 에 대한 임의의 표본을 $\{y_1, \ldots, y_n\}$ 라 하고 임의 표본에 대한 분위수를 ξ라 할 때, 절대 표준편차인 $|y_i - \xi|$의 합을 최소화하는 분위수 ξ는 분위가 0.5일 때의 중위수가 된다.

$$\min_{\xi \in R} \sum_{i=1}^{n} |y_i - \xi| \quad (8)$$

또한 위 식은 가중치 도입을 통해 분위가 0.5인 중위수뿐만 아니라 임의의 분위 γ에 대한 분위수 $\xi(\gamma)$를 추정할 수 있다. 이때의 가중치 ρ_r은 표본 y_i가 해당분위보다 큰 경우에는 해당분위 γ이 가중치가 되며, 해당분위보다 작은 경우에는 1-γ이 가중값을 가진다. 예를 들어 분위 γ가 0.2일 경우 표본의 분위가 0.2보다 크면, 0.2의 가중치가 적용되며 작은 경우 0.8(=1-0.2)의 가중치를 가진다. 식(9)와 같이 가중치가 고려된 오차의 절대편차를 최소화하는 분위수 ξ은 해당 분위 γ의 분위수 $\xi(\gamma)$가 된다.

$$\min_{\xi \in R} \sum_{i=1}^{n} \rho_r |y_i - \xi| \quad (9)$$

한편 식(9)에서 임의 표본 y_i가 해당하는 분위값 ξ보다 크면 $y_i - \xi$는 음의 값을 가지며, 이때 가중치는 1-γ이 된다. 가중치를 1-γ 대신에 γ-1을 도입하면 다음 식과 같이 절대값 부호를 쓰지 않고 표현할 수 있다.

$$\min_{\xi \in R} \sum_{i=1}^{n} \rho_r (y_i - \xi) \quad (10)$$

위 식을 선형 조건부 분위함수인 $Q(\gamma | X = x) = x'\beta(\gamma)$를 도입하면, 추정된 계수 벡터 $\beta(\gamma)$는 변수 벡터 x에 따라 달라지는 분위수를 추정하기 위한 기울기가 된다.[2)]

$$\hat{\beta}(\gamma) = \mathrm{arg}min_{\beta \in R^p} \sum_{i=1}^{n} \rho_r (y_i - x_i{}'\beta) \quad (11)$$

본 연구에서 분위회귀모형을 이용한 반복매매지수 추정은 분위 γ=0.5일 때이며, OLS 모형과 비교를 통해 두 모형의 특성을 비교하고자 하였다.

Ⅳ. 실증분석

1. 자료

본 연구는 지수산정을 위해 국토해양부에서 발표한 2006년 1월부터 2011년 12월까지의 실거래가 정보를 이용하였다. 지수는 서울시를 대상으로 산정하였으며, 국토해양부에서 발표하고 있는 면적별 기준에 따라 소형, 중형, 중대형, 대형[3)]으로 아파트 지수를 세분화하여 표본 수 변화에 따른 지수변화를 살펴보고자 하였다.

[표 1]은 본 연구에서 활용한 서울시 실거래건

2) 식(11)과 같은 방법으로 오차제곱합인 OLS를 표현할 경우에는 다음 식과 같이 나타낼 수 있다.

$$\hat{\beta} = \mathrm{arg}min_{\beta \in R^p} \sum_{i=1}^{n} (y_i - x_i{}'\beta)^2$$

3) 전용면적으로 60㎡ 이하는 소형, 60㎡ ~ 85㎡ 이하는 중소형, 85㎡ ~ 135㎡ 이하는 중대형, 135㎡ 초과는 대형으로 구분하고 있다.

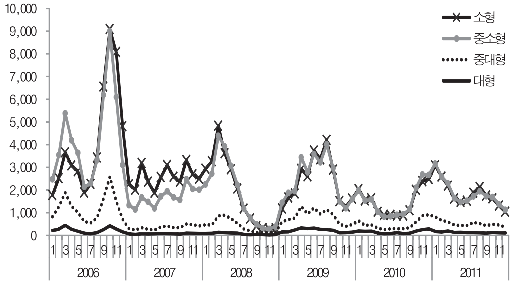
수로, 소형과 중소형 규모의 아파트 거래는 많지만 대형 아파트는 평균 거래건수가 소형의 6% 수준이며, 최소 월별 실거래건수는 17건에 불과한 달도 있다는 것을 알 수 있다. 따라서 대형아파트의 경우에는 표본수의 한계에 의해 지수 변동이 나타날 가능성이 높을 것으로 예상된다.

[표 1] 서울시 규모별 월별 실거래건수

구분	소형	중소형	중대형	대형
평균	2,392.7	2,269.8	630.9	145.4
최소값	273.0	284.0	97.0	17.0

[그림1]은 2006년부터 2011까지의 서울시 주택의 실거래 건수의 추이를 나타낸 것으로, 2008년 11월에는 모든 규모의 아파트가 가장 낮은 실거래건수를 보이고 있다는 것을 알 수 있다. 이는 글로벌 금융위기의 영향이 아파트 거래에 전반적으로 영향을 미친 것으로 보인다. 특히 대형의 경우는 단 단위의 거래량이 관측되는 시점도 다수 존재한다.

[그림 1] 서울시 규모별 실거래 건수 추이



2. OLS 및 분위회귀 지수 추정

현재 국토교통부 아파트 실거래가 지수의 경우 동일주택 가정을 통해 거래쌍을 구성하고 있다. 동일주택 가정은 유사한 가격수준을 유지하는 단지와 평형, 동, 층(3가지 층구분, 1·2층/최상층/중간층)이 같을 경우 동일한 주택의 거래로 간주하는 것이다. 이 4가지 기준에 의해 동일한 아파트로 분류된 경우 가상적인 동일주택의 거래량이 2건 이상일 경우에는 평균값을 이용하며 거래량에 가중치를 부여하여 지수를 산정한다. 본 연구에서 산정한 반복매매지수는 국토교통부에서 이용하는 자료와 달리 동 정보가 없는 관계로 단지와 평형, 층이 같을 경우 동일한 주택이라 가정하여 지수를 산정하였다.

규모별 지수추정 결과는 [그림 2] 에 제시되어 있다. 일반적으로 표본수가 적은 상태에서는 개별 관측치가 지수산정에 과도하게 영향을 미치게 된다. 이는 관측치가 이상치인 경우 산정된 지수의 안정성을 저해한다. 시각적인 관측에 기초할 때 표본수가 많은 소형, 중소형의 경우에는 두 지수의 안정성에 큰 차이를 보이지 않는다. 그러나 분위회귀 지수가 OLS 지수에 비해 상대적으로 다소 낮게 산정되는 경향성을 보인다. 이는 Coulson and McMillen(2007)이 제시한 것처럼 국내 아파트의 경우도 간헐적으로 발생하는 각 세대별 내부수선의 효과가 OLS 지수에는 통제되지 못한 결과일 수도 있다.

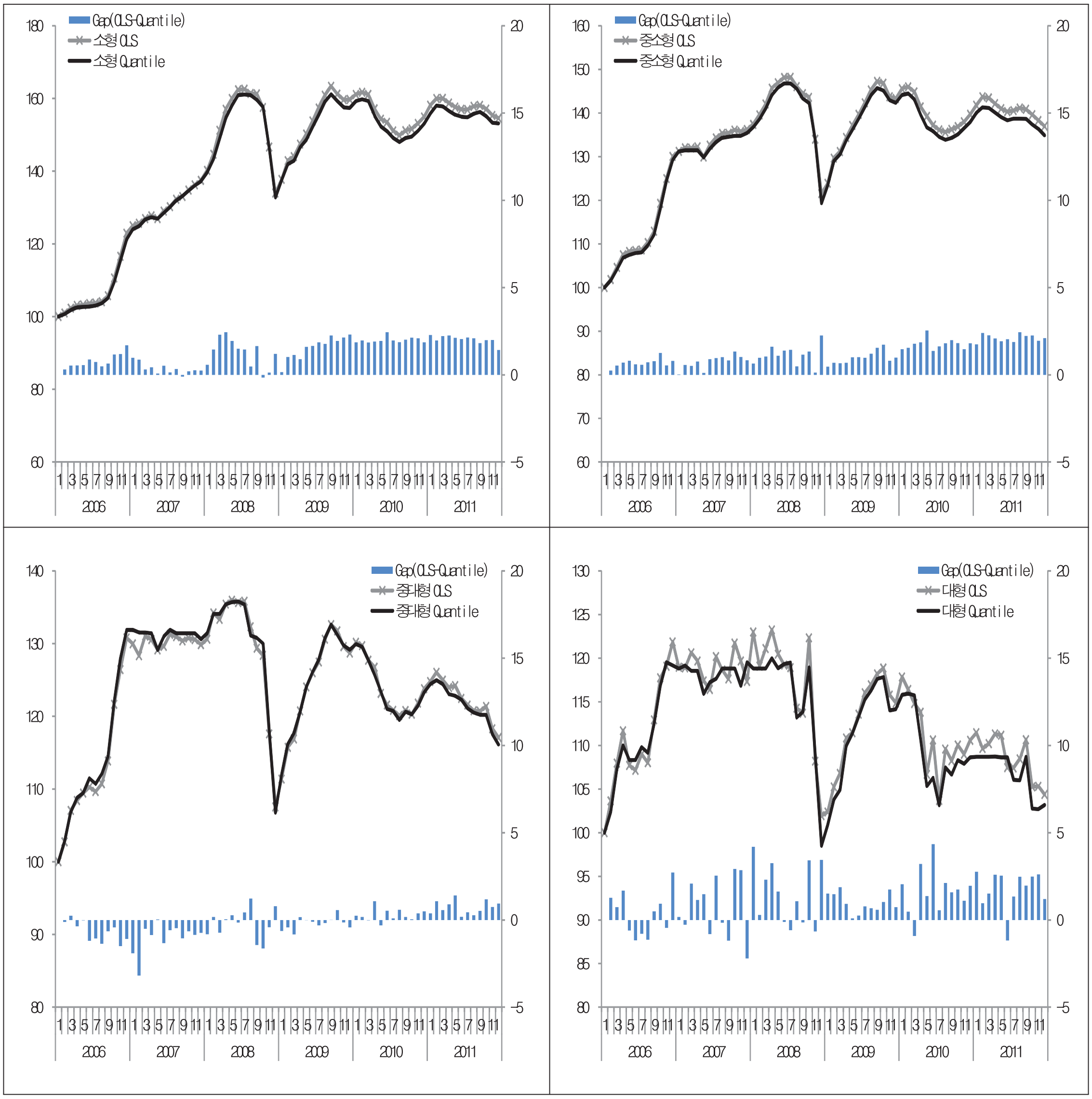
표본수가 적은 중대형, 대형의 경우에는 지수의 수준차를 시각적으로 인지할 만큼 안정적인 추세를 보이지 못하고 있다. 다만 중위수 추정지수가 OLS 지수보다 다소 변동폭이 적고 안정된 추이를 보이고 있음을 알 수 있다. 특히 대형의 경우 OLS 지수에서 관측되는 각 시점별 피크가 분위회귀 지수에서는 평활화되는 경향성을 관측할 수 있다.

이는 중위수가 갖는 특성상 이상치의 값이 지수를 산정하는데 있어 크게 영향을 주지 않으므

로 발생하는 결과로 판단된다. 따라서 표본수가 적은 경우에는 분위회귀 지수의 이용이 상대적으로 안정적인 지수를 산정하는데 도움이 될 것으로 판단된다. 다만 분위회귀를 이용하여 지수를 산정할 경우 국부시장의 가격변동률이 나머지 시장의 중위수준 가격변동률에 비해 지속적으로 높다라면 그러한 국부시장의 가격변동을 담지 못하는 둔감한 지수가 산정될 수 있는 한계가 있다.

이상의 논의는 산정된 지수 그래프를 시각적으로 분석한 결과에 불과하다. 다음 장에서는 좀 더 종합적으로 지수산정방식의 선택을 위해 이용될 수 있는 판단지표를 논의하고 그 지표들을 이용한 지수간의 비교를 진행하도록 하자.

[그림 2] 서울시 규모별 아파트 실거래가 지수(2006.1=100.0)



V. 지수간 비교 평가

1. 평가지표 개발

다양한 지수산정방식에 따라 산출되는 지수를 비교하기 위해서 통계적인 신뢰도, 지수의 안정성, 추가적으로 지수의 민감도를 판단할 수 있는 평가지표를 살펴보도록 하자.

1) 통계적인 신뢰도

추정치의 통계적 신뢰도를 나타내는 가장 기초적인 평가지표는 추정치의 표준오차이다. 다만 추정치의 단위나 수준이 다른 경우 표준오차 값이 큰 차이를 보임으로 많은 경우 추정치의 평균적인 수준을 통제하고 단위에 영향을 받지 않는 평가지표로서 변이계수(CV)가 주로 이용된다(서울시 2011). 변이계수의 기본적인 산식은 표준편차를 평균으로 나눈 것으로 다음 식과 같다.

$$CV = \frac{\text{표준편차}}{\text{평균}} \tag{11}$$

그러나 가격지수의 경우 지수값이 절대적인 수준이 아니라 기준시점 대비 누적가격변동률의 성격을 지닌 상대적인 값으로 기준 시점을 달라지면 지수값이 달라지는 특성이 있다. 또한 지수값이 커질수록 동일한 표준오차를 가진 지수값의 경우 변이계수는 감소하는 문제점이 발생한다.

따라서 본 지수평가에서는 다음식과 같이 산정되는 지수화된 표준오차의 평균값(Mean of Standard Error of Index, MSEI)을 이용하기로 한다. 다만 추정계수 $\widehat{\alpha_t}$ 의 표준오차 se_t 가 대부분 0 에 가까운 값임으로 근사치로 추정계수 표준오차의 평균값을 이용해도 큰 무리가 없다.

$$MSEI = \sum_{t=2}^{T} e^{se_t} \times 100/T \tag{12}$$
$$\simeq \sum_{t=2}^{T} se_t \times 100/T$$

그러나 분위회귀의 경우 오차항의 분포에 대한 확률분포의 가정이 필요치 않음으로 인해 OLS와 같이 대칭적인 확률분포에 근거한 표준오차나 신뢰구간의 개념을 이용하는데 한계가 있다. 따라서 본 분석에서는 통계적인 신뢰도를 판단하기 위해 추정지수의 95%신뢰구간의 근사치로서 추정계수의 2.5%~97.5% 확률구간을 추가적으로 이용하도록 한다. 산정식은 I_t^{γ} 가 t 시점 가격지수의 γ 백분위수일 때 다음과 같이 산정된다.

$$CI = \frac{\sum_{t=2}^{T} (I_t^{97.5\%} - I_t^{2.5\%})}{T-1} \tag{13}$$

2) 지수의 안정성

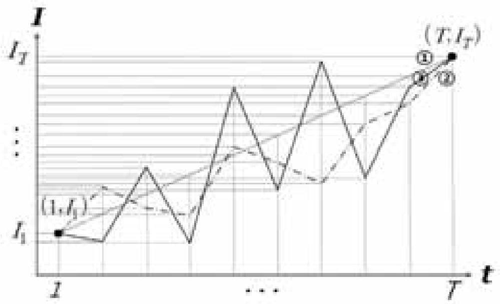
다음으로 지수추세의 안정성을 평가하는 지표인 안정성지수(Stability Index, SI)를 개발하여 적용하도록 한다. SI의 기본적인 개념은 시각적인 관측을 통한 판단을 계량화하는 것이다.

[그림 3]에서 선 ① 은 가상적으로 가장 안정적인 지수추세를 가정하여 나타낸 선이다. 이 경우 첫 시점 지수와 마지막 시점 지수간의 거리는 최소화된다. 현실에서 산정되는 지수는 선 ② 와 같이 시점별로 선 ①에서 벗어나 추세로 움직이게 된다. 선 ③ 은 선 ② 에 비해 좀 더 불안정한 추세를 보여주는 지수를 표현한다. 각 경우를 비교해보면 지수추세가 불안정해질수록 각 시점의 지수값을 연결하는 선의 길이가 길어지게 된

다. 특히 인접시점의 지수값의 차이가 클수록 총 이동거리는 길어지게 된다.

SI는 첫 시점과 마지막 시점의 직선거리를 모든 시점 간 이동거리의 총합으로 나누어 계산된다. 따라서 산정되는 안정성 지수 SI 는 0에서 1 사이의 값을 가지게 되고, 안정적일수록 1에 가까운 값을 갖게 된다.

[그림 3] 안정성지수 SI 개념도



기준시점이 1, 마지막 시점이 T 이고, t 시점의 지수값을 I_t 라고 할 때, SI는 다음과 같이 산정된다.

$$SI = \frac{\sqrt{(T-1)^2 + (I_T - I_1)^2}}{\sum_{t=1}^{T-1} \sqrt{1^2 + (I_{t+1} - I_t)^2}} \qquad (14)$$

3) 지수의 민감도

마지막으로 지수의 민감도를 평가할 수 있는 지표로서, 신호-잡음비 (Signal-to-noise ratio, S/N ratio) 가 있다. 여기서 신호(signal)는 가격지수의 시점 간 변동률의 표준편차로서 모형에 의해 설명되는 변동량으로 시장의 변화를 나타내고, 잡음(noise)은 추정계수 표준오차의 평균으로 산정된 지수가 지닌 확률적 오차를 의미한다(Case and Shiller, 1987). 따라서 설명할 수 없는 오차에 비해 설명할 수 있는 변동량이 크다는 것은 상대적으로 산정된 지수의 통계적인 설명력이 크다는 것을 의미한다.

$$S/N = \frac{\text{지수변동률의 표준편차}}{\text{추정계수 표준오차의 평균}} \qquad (15)$$

그러나 시점 간 지수변동률의 표준편차가 크다는 것은 지수의 지수추세가 불안정하다는 의미도 내포되어 있으며, 추정계수의 표준오차가 클수록 산정된 지수가 불안정하게 추정되어 인접시점 간 변동률이 크게 산출될 가능성도 있다.

따라서 S/N 비는 통계적인 신뢰도를 의미하기 보다는 지수의 민감도를 표현하는 지표로서 이해할 수 있다. 지표의 성격 상 현실적인 적용에 있어 해석이 애매한 경우가 자주 발생하여 본 분석에서 해석에 있어 큰 비중을 두지 않기로 한다.

2. 평가결과

앞에서 두 지수 산정 방법을 평가하기 위한 방안으로 3가지 평가지표를 소개하였다. 통계적 신뢰도를 측정하는 지수 표준오차의 평균(MSEI)과 지수추세의 안정성을 평가하는 안정성지수(SI), 지수의 민감도를 측정하는 신호-잡음비(S/N)이 그 것이다. 이 3가지 지표와 95%신뢰구간의 거리를 측정한 결과를 주택규모에 따라 정리하였다.

각 지표의 값은 각 주택유형의 관측치 개수와 밀접한 연관성을 갖고 있었다. 소형과 중소형 주택은 비교적 많은 관측치가 확보되어 있는 반면, 대형주택의 경우에는 약 140개 관측치로 지수가 산정되었다.

〈표 2〉 주택규모별 지수 비교지표 결과

구분	비교지표	OLS	Quantile
소형	MSEI	0.351	0.317
	95% CI	1.397	1.261
	SI	0.489	0.493
	S/N ratio	6.132	6.480
중소형	MSEI	0.334	0.333
	95% CI	1.329	1.323
	SI	0.481	0.484
	S/N ratio	6.322	6.560
중대형	MSEI	0.669	0.669
	95% CI	2.661	2.661
	SI	0.443	0.452
	S/N ratio	3.411	3.404
대형	MSEI	1.673	1.733
	95% CI	6.649	6.893
	SI	0.345	0.428
	S/N ratio	1.799	1.483

먼저 어느 지수가 더 통계적인 신뢰도가 높은지 알아보기 위하여, 두 지수의 MSEI와 95%신뢰구간의 크기를 비교해보도록 하자. 소형과 중 중소형 주택 사이에는 두 지표 사이에 큰 차이는 없으나 Quantile 지수가 상대적으로 낮은 MSEI 값과 짧은 신뢰구간을 가지므로 통계적 신뢰도가 높은 것으로 나타났다. 반면 중대형주택의 경우는 두 지표사이의 차가 없었고, 대형주택의 경우에는 OLS지수가 더 높은 통계적 신뢰도를 갖는 것으로 나타났다.

그러나 SI의 경우 모든 주택규모에서 Quantile 방식으로 지수 산정시 안정성이 높게 나타난다. 특히 관측치의 개수가 적어질수록 그 격차가 더 벌어져 대형의 경우 OLS 지수는 0.345인 반면 Quantile 지수는 0.428로 24% 이상 향상된 SI 지표값이 산출되었다. 이와 같이 지표를 통해 살펴본 지수의 안정성은 앞에서 지수추세의 시각적인 관측을 통해 분석된 결과에 부합하는 경향성을 보여준다.

신호대 잡음비의 경우 소형과 중소형에서는 Quantile지수가, 중대형과 중형에서는 OLS지수가 더 높은 값을 가지고 있었다. 앞에서 언급한 것처럼 지수의 변동성이 크면 S/N비가 높게 산정되는 경향이 있다. 대형 OLS 지수의 경우 산정된 지수의 불안정성이 지수변동률의 표준오차를 크게하여 S/N 비가 높게 산출되었다.

V. 결론

본 연구는 표본 수의 제약으로 인해 발생하는 지수의 불안정성의 문제를 해결하기 위해 최근 해외에서 논의되고 있는 분위회귀를 이용하여 반복매매지수를 산정하였다. 기본적으로 분위회귀는 OLS와 달리 중위수를 이용하여 지수를 산정하는 방법으로 비교적 안정적인 지수의 추이를 볼 수 있으며, 이상치의 영향을 적게 받는 장점을 가지고 있어 표본 부족으로 인한 이상치 문제를 해결할 수 있는 하나의 대안이 될 수 있다.

본 연구에서는 분위회귀와 OLS를 이용한 반복매매지수의 특성을 비교하기 위해 2006년 1월부터 2011년 12월까지 거래된 서울시 실거래가 자료를 이용하여 규모별로 지수를 추정하고, 3가지의 지표를 통해 신뢰도와 안정성과 민감도를 비교해보았다.

분석 결과를 통해 분위회귀를 이용하여 지수를 산정할 경우, 전체적으로 지수변동의 피크가 둔화되는 경향성을 보인다. 이러한 경향성은 표본수가 적은 대형주택의 경우 좀 더 명확하게 관측된다. 이러한 특징을 좀 더 지표화된 형태로 비교하기 위해 통계적 신뢰도 지표(MSEI), 안정성 지표(CI), 민감도 지표(S/N Ratio)를 이용해

비교하였다.

통계적 신뢰도의 지표인 MSEI와 신뢰구간의 경우에는 표본 수가 많은 소형 및 중소형의 경우 Quantile 지수가 우수하나 표본수가 적은 대형의 경우는 OLS가 우수하여 그 차이는 크지 않다. 그러나 안정성에 있어서는 모든 주택규모에서 Quantile 지수가 우수하며, 특히 대형의 경우는 그 차이가 크게 발생하였다.

이와 같은 분석결과를 바탕으로 판단할 때 대형 아파트와 같이 빈번한 반복거래가 관측되지 못하는 경우 지수의 불안정성을 해결하는 방안으로 통계적인 신뢰도는 크게 훼손하지 않으면서 시점 간 지수변동의 불안정성을 줄일 수 있는 대안으로 Quantile 지수의 산정은 유효한 선택이라고 판단된다. 다만 표본수가 극히 적은 경우 추정치의 안정성에 근본적인 한계가 있음으로 Quantile 지수의 적용에도 적정 수준의 표본수가 요구됨을 다시 밝힌다.

논문접수일 : 2013년 7월 1일
논문심사일 : 2013년 7월 15일
게재확정일 : 2013년 8월 1일

참고문헌

1. 류강민·박유미·이창무, “비선형 회귀분석을 이용한 산술평균 반복매매지수 산정방법에 관한 연구”, 「주택연구」, 제17권 제4호, 한국주택학회, 2009, pp.259-278
2. 류강민·이창무, “반복매매지수 지수변화 보정에 관한 연구”, 「주택연구」, 제19권 제2호, 한국주택학회, 2011, pp.5-22
3. 이창무·김동근·안건혁, “아파트 월세지수 산정에 관한 연구”, 「국토계획」, 제38권 제6호, 대한국토·도시계획학회, 2003, pp.47-60
4. 이창무·김병욱·이현, “반복매매모형을 이용한 아파트 매매가격 지수”, 「부동산학연구」, 제8권 제2호, 한국부동산분석학회, 2002, pp.1-19
5. 이창무·김용경·배익민, “반복매매모형을 이용한 아파트 실거래지수 운영특성 분석”, 「부동산학연구」, 제13권 제2호, 한국부동산분석학회, 2007, pp.21-40
6. 이창무·김진유·이상영, “공동주택 실거래가 지수 산정에 관한 연구”, 「국토계획」, 제40권 제4호, 대한국토·도시계획학회, 2005, pp.121-134
7. 이창무·배익민, “시세가격을 활용한 아파트 실거래가 반복매매지수 산정”, 「부동산학연구」, 제14집 제2호, 한국부동산분석학회, 2008, pp.21-37
8. Bailey, J. M. and R. F. Muth and H. O. Nourse, “A Regression Method for Real Estimate Price Index Construction”, Journal of the American Statistical Association, Vol. 58, 1963, pp. 933-942.
9. Case, K. E., “The Market for Single-Family Homes in Boston Area”, New England

Economic Review, September/October, 1986, pp. 38-48
10. Case, K. E. and R. J. Shiller, “Price of Single Family Homes since 1970; New Indexes of Four Cities”, New England Economic Review, September/October, 1987, pp. 46-56
11. Clapp, John. M. and Carmelo Giaccotto, “Estimating Price Indices for Residential Property: A Comparison of Repeat Sales and Assessed Value Methods”, Journal of Real Estate Finance and Economics, Vol. 5, 1992, pp. 357-374
12. Coulson, N. E. and Daniel P. McMillen, “The Dynamics of Intraurban Quantile House Price Indexes”, Urban Studies, Vol. 44 No. 8, 2007, pp. 1517-1537
13. Epley, D., “A better method to estimate price change in single family housing: A test of median-to-median compared to repeat sales”, International Journal of Housing Markets and Analysis, Vol. 5 No. 3, 2012, pp. 377-391
14. Goetzmann, W. N., “The Accuracy of Real Estate Indices: Repeat Sales Estimators”, Journal of Real Estate Finances and Economics, Vol. 5, 1992, pp. 5-53
15. Goetzmann, W. N. and Liang Peng, “The Bias of the RSR Estimator and the Accuracy of Some Alternatives”, Real Estate Economics, Vol. 30 No. 1, 2002, pp. 13-19
16. Liang Peng, “GMM Repeat Sales Price Indices”, Real Estate Economics, Vol. 30 No. 2, 2002, pp. 239-261
17. McMillen, D and P. Thorsnes, “Housing Renovations and the Quantile Repeat Sales Price Index”, Real Estate Economics Vol. 34, 2006, pp. 567–587
18. Shiller, R. J., “Arithmetic Repeat Sales Price Estimators”, Journal of Housing Economics, Vol. 1, 1991, pp. 110-216

부 록 : OLS와 중위회귀의 지수 추정 결과

		소형		중소형		중대형		대형	
연도	월	OLS	Quantile	OLS	Quantile	OLS	Quantile	OLS	Quantile
2006	1	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
	2	101.0	100.7	101.8	101.6	102.7	102.8	103.6	102.4
	3	102.3	101.8	104.6	104.1	107.1	106.8	108.0	107.2
	4	103.1	102.5	107.5	106.8	108.4	108.8	111.7	110.0
	5	103.2	102.7	108.3	107.5	109.4	109.5	107.7	108.3
	6	103.6	102.8	108.5	107.9	110.3	111.5	107.2	108.3
	7	103.8	103.0	108.6	108.1	109.7	110.7	109.0	109.8
	8	104.1	103.6	110.3	109.6	110.7	112.1	108.0	109.1
	9	105.7	105.1	112.9	112.1	113.9	114.5	112.9	112.4
	10	110.6	109.5	119.3	118.0	121.7	122.1	117.7	116.8
	11	116.6	115.4	125.0	124.4	126.4	127.9	119.1	119.5
	12	123.0	121.3	130.0	129.3	130.8	131.9	121.9	119.1
2007	1	125.0	124.0	131.2	131.2	129.9	131.9	118.9	118.8
	2	125.7	124.8	132.0	131.5	128.3	131.5	118.9	119.1
	3	127.0	126.7	132.0	131.5	131.0	131.5	120.6	118.5
	4	127.7	127.3	132.2	131.5	130.6	131.4	119.7	118.5
	5	127.0	126.9	129.9	129.8	129.1	129.1	117.4	115.9
	6	129.0	128.5	132.7	131.8	129.6	131.0	116.4	117.2
	7	130.3	130.1	134.3	133.3	131.3	131.9	120.2	117.6
	8	132.3	131.9	135.3	134.3	130.9	131.4	118.6	118.8
	9	133.0	133.1	135.4	134.6	130.4	131.4	117.6	118.8
	10	134.8	134.6	136.1	134.8	130.8	131.4	121.7	118.8
	11	136.2	136.0	135.8	134.8	130.6	131.4	119.7	116.8
	12	137.4	137.2	136.2	135.4	129.8	130.6	117.3	119.5
2008	1	140.2	139.6	137.4	136.7	130.6	131.4	123.0	118.8
	2	144.7	143.2	139.6	138.6	134.2	134.1	119.1	118.8
	3	151.2	148.9	142.1	141.0	133.3	134.1	121.1	118.8
	4	157.0	154.6	145.7	144.1	135.4	135.4	123.2	120.0
	5	160.0	158.0	146.9	145.8	136.0	135.7	120.5	118.8
	6	162.4	160.9	148.2	146.8	135.6	135.8	119.2	119.3
	7	162.5	161.1	148.2	146.8	135.8	135.4	118.9	119.5
	8	161.4	160.9	146.0	145.6	132.3	131.1	114.3	113.2
	9	161.3	159.6	144.4	143.3	129.4	130.8	113.7	113.8
	10	157.5	157.6	143.6	142.2	128.4	130.0	122.3	118.9
	11	146.6	146.5	134.0	133.8	117.6	118.0	108.2	108.8
	12	134.0	132.8	121.5	119.3	107.5	106.7	101.9	98.5

(부록 계속)

		소형		중소형		중대형		대형	
연도	월	OLS	Quantile	OLS	Quantile	OLS	Quantile	OLS	Quantile
2009	1	137.7	137.6	124	123.4	111	112.0	102	100.9
	2	142.9	141.9	130	129.0	116	116.2	105	103.8
	3	144.0	142.9	131	130.5	117	117.7	107	104.9
	4	147.4	146.6	134	133.7	121	120.6	111	109.9
	5	150.2	148.6	137	136.1	124	124.1	111	111.4
	6	153.8	152.2	140	138.7	126	126.1	114	113.3
	7	157.4	155.5	142	141.4	127	127.8	116	115.2
	8	160.9	159.1	145	144.1	131	130.8	117	116.3
	9	163.4	161.1	147	145.7	133	132.6	118	117.6
	10	161.2	159.2	147	145.1	132	131.2	119	117.8
	11	159.6	157.5	144	142.9	130	129.7	116	114.0
	12	159.7	157.4	143	142.3	129	129.1	115	114.1
2010	1	161.1	159.2	146	144.1	130	129.9	118	115.8
	2	161.7	159.7	146	144.4	130	129.5	116	115.9
	3	161.1	159.2	145	143.1	128	127.7	115	115.7
	4	157.1	155.2	141	139.5	127	125.7	114	110.6
	5	154.3	152.4	139	136.7	123	123.5	107	105.3
	6	153.4	151.0	137	135.8	122	121.0	111	106.3
	7	151.1	149.1	136	134.5	121	120.7	104	103.1
	8	149.8	148.0	136	133.8	120	119.5	110	107.5
	9	151.1	149.1	136	134.3	121	120.6	108	106.7
	10	151.6	149.5	137	135.1	120	120.2	110	108.3
	11	153.1	151.0	138	136.5	122	121.4	109	107.9
	12	154.9	153.1	140	137.9	124	123.3	111	108.6
2011	1	158.1	155.9	142	140.0	125	124.4	111	108.7
	2	160.0	158.0	144	141.3	126	125.0	110	108.7
	3	160.0	157.8	143	141.2	125	124.4	110	108.7
	4	158.8	156.5	142	140.1	124	123.0	111	108.7
	5	157.6	155.5	141	139.0	124	122.8	111	108.6
	6	157.0	155.0	140	138.3	122	122.3	107	108.6
	7	156.9	154.8	141	138.7	122	121.1	107	106.1
	8	157.8	155.8	141	138.7	121	120.5	108	106.0
	9	158.1	156.3	141	138.6	121	120.2	111	108.7
	10	157.1	155.1	140	137.4	121	120.2	105	102.8
	11	155.3	153.3	138	136.3	118	117.6	105	102.7
	12	154.6	153.1	137	134.9	117	116.1	104	103.2