

부동산 경매 Big Data를 활용한 Chaos 분석

강준^a, 김지우^b, 이현준^c, 오경주^d

^a 연세대학교 투자정보공학 박사과정
(03722)서울특별시 서대문구 연세로 50,
Tel: +82-10-2780-1972, Fax: +82-2-718-0606, E-mail: kj@ggfund.co.kr

^b 연세대학교 산업공학과 석박사통합과정
(03722)서울특별시 서대문구 연세로 50,
Tel: +82-2-2123-7785, Fax: +82-2-364-7807, E-mail: jiwoo14@naver.com

^c 연세대학교 산업공학과 석박사통합과정
(03722)서울특별시 서대문구 연세로 50,
Tel: +82-2-2123-7785, Fax: +82-2-364-7807, E-mail: 2wisedeep@gmail.com

^d 연세대학교 산업공학과 교수
(03722)서울특별시 서대문구 연세로 50,
Tel: +82-2-2123-5720, Fax: +82-2-364-7807, E-mail: johanoh@yonsei.ac.kr

Abstract

빅 데이터 분석 및 활용이 확대 되는 현재 국내 부동산 영역에서의 연구는 타 영역에 비해 미진하다. 그러나 부동산 경매 분야는 데이터 축적이 계속되고 있고 그 형태도 정형화 되어 충분한 연구가 주목되는 분야이다. 국내에서 부동산 경매 낙찰가율 데이터를 활용한 Chaos 분석 연구는 전무하다. 본 연구에서는 Hurst 지수, correlation dimension, maximum Lyapunov 지수, 이 3가지 Chaos 분석기법을 활용하여 낙찰가율의 비선형 결정론적 동역학적 특성을 확인하고, Chaos 분석을 통하여 얻은 결과와 실무 데이터를 비교하여, 그 함의를 도출한다. 높은 Hurst 지수에 따르는 추세와, maximum Lyapunov 지수의 측정을 통한 지속성, 그리고 correlation dimension 분석의 결과에 따라 time lag가 개시결정일에서 낙찰일, 배당요구종기일에서 낙찰일까지와 일치하는 점으로부터, Chaos 분석이 낙찰가율의 움직임을 예측하는데 유용함을 확인할 수 있었다.

Keywords:

부동산 경매, Chaos 분석, Hurst 지수, correlation dimension, maximum Lyapunov 지수.

1. 서론

4차산업혁명 시대에 산업 각 영역에서 빅 데이터에 대한 연구가 활발히 진행되고 있다. 그 중에서 부동산 영역에 대한 빅 데이터 연구는 타 영역에

비해 속도가 더딘 것이 현실이다. 그 이유는 부동산 영역에서는 정형화된 데이터 정리 및 수집이 어려웠고 데이터를 이용한 산업 접목 보다는 부동산 개별적 문제해결에 집중해 왔기 때문이다. 그러나 부동산 경매 분야는 연간 20~55만건 이상의 경매가 진행 중이며 그에 대한 데이터 역시 정형화 되어 그 활용도가 커지고 있다. 2001.1. ~ 2016.9. 까지 약 17년간 전국 전체 물건의 경매물건을 보면 약 640만건의 경매가 진행되었으며 실제 낙찰된 건수도 190만건에 이르며 평균 낙찰가율이 감정가 대비 70%를 유지하고 있다. 이러한 데이터의 축적으로 향후 빅 데이터로서의 연구와 활용이 충분히 가능할 것으로 예상 된다.

Table 1 Real estate auction data

Period	Auction total number	Sold out total number	Sale price rate	Average bidder
2001.1. ~ 2016.9.	6,387,549	1,905,598	70.72%	3.52

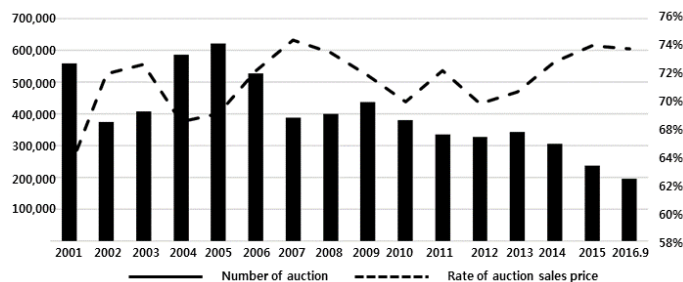


Figure 1 Auction sale price rate and number

이러한 부동산 경매 빅 데이터는 인공지능과

접목하여 연구할 단계에 이르렀음을 느끼게 된다. 본 논문을 통해 인공지능 적용에 앞서 Chaos 분석을 통해 데이터의 인공지능과 접목 가능성에 대한 구조를 분석해 보고자 한다.

부동산 시장에 대한 Chaos 분석 연구는 학계에서 거의 시도된 적이 없는 분야이다. 부동산 경매시장에 관한 연구로는 부동산 매매시장과 경매시장 사이의 가격 선도 및 지연에 대한 연구가 있었다. 전통적인 통계적 기법을 이용한 모형 비교 연구가 시도 되었으나 인공지능 기법 적용 부재와 통계적 분석방법의 한계가 있었다. 특히 부동산 영역 중 경매에 대한 Chaos 분석 연구는 지금까지 전무하다. 부동산 경매 분야는 다른 어느 부동산 영역보다 많은 거래가 이루어지기 때문에 데이터의 양이 방대하며 데이터가 실시간으로 이루어지고 있다. 따라서 빅데이터의 해석과 활용이 중요성을 더 해가고 있는 현재 통계적인 분석 한계를 넘어 인공지능을 활용한 분석이 최근 부동산 경매 분야에서 요구되고 있다.

부동산 경매 낙찰가를 종속변수로 하고 부동산 경매에 영향을 주는 많은 입력변수들의 상호 관계를 연구하여 종속변수의 흐름을 예측한다면 여타 부동산 거래 이상으로 경매의 경우도 낙찰가가 시장의 흐름을 따르며, 이에 앞서 경매 낙찰가율이 예측될 수 있다면 그 변동성에 따르는 현금흐름변화와 시장 충격에 효과적으로 반응하는 포트폴리오를 구성할 수 있을 것으로 기대 된다.

본 연구를 통해 부동산 경매 데이터를 활용한 Chaos 분석을 통하여 낙찰가율의 시계열적 속성과 의미를 분석하고 그 결과를 정량적으로 분석해보고자 한다. 본 연구는 인공지능을 활용한 예측이 가능한 것인가를 확인하는 논문으로써 단순히 Chaos 분석에서 끝나는 것이 아니라 경매 실무적인 이론과 분석 자료가 충분히 일치하는지 확인하여 그 함의를 도출해보고자 한다.

본 논문은 다음과 같이 구성되어 있다. 2절에서는 부동산 분야에 대한 Chaos 분석을 위한 데이터 지표 및 선행연구를 살펴보고, 3절에서는 Chaos 분석을 위한 모델을 제안하며, 4절에서는 부동산 경매 낙찰가율을 월간, 주간, 일간별 데이터로 구분하여 Chaos 분석을 시도하여 각 기간 낙찰가율 데이터를 통해 Hurst 지수, correlation dimension, maximum Lyapunov 지수 등을 도출하였고 그 값을 통해 데이터의 속성을 분석하였다. 덧붙여서 각 용도별 부동산 경매 Chaos 분석을 통해 아파트 경매의 Chaos분석과 비교해 보았다. 마지막으로 결론에서는 앞서 제시한 3가지 분석의 결과와 연구의 기대효과 및 향후 연구에 대해 서술하였다.

2. 연구배경

지난 17년간 (2001.1. ~ 2016.9.) 전국 부동산 경매에

대한 용도별 분석을 살펴보면 주거용 부동산이 124만건 진행되었고 낙찰가율 75% 수준을 유지했다. 아파트 112만건 낙찰가율 83%, 상업용 부동산 100만건 낙찰가율 57%, 공업용 부동산 12만건 낙찰가율 66%, 토지 167만건 낙찰가율 70% 정도를 보이고 있다. 이중 아파트 경매는 입찰자가 평균 약 5명 정도로 활발하게 진행되고 있음을 알 수 있다.

Table 2 2001.1. ~2016.9. Real estate auction data for each purpose of use

Period	Auction total number	Sold out total number	Sale price rate	Average bidder
Residential	1,237,315	375,515	74.95%	3.6
Apartment	1,120,339	378,427	82.88%	4.95
Commercial	1,005,153	207,165	57.65%	2.38
Industrial	122,159	31,828	65.76%	2.59
Land	1,667,423	488,069	70.49%	2.08

특히 동 기간 동안 서울 아파트 경매의 평균 낙찰가율을 살펴 보면 부동산 급등 (2003.6.) 및 외환위기 시기 (2008.12.)를 제외하고는 약 감정가의 82.88% 수준에서 낙찰가가 형성되고 있어 일정한 투자의 패턴이 형성되고 있음을 가정할 수 있다.

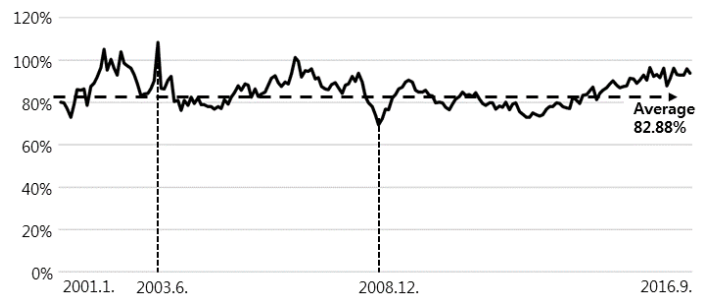


Figure 2 Seoul APT auction sale price rate

한편 자본시장에 대한 카오스 분석의 뿌리는 효율적 시장가설과 시장 수익률 밀도함수의 정량적 분석에 의한 모순에 기반을 둔다. 효율적 시장가설 (efficient market hypothesis)은 시장의 가격이 모든 정보를 반영하여 형성된다는 가설로, 결국 모든 정보가 이미 반영되어 있기 때문에 어떠한 방법을 통해서도 추가수익을 낼 수 없다는 내용이다. 이는 결국, 과거 정보는 이미 가격에 모두 반영되어서 근사적으로 정규분포를 이루고 독립적이어야 한다는 뜻이다 (Cootner와 Fama, 1964).

그러나 거듭된 연구결과 자본시장의 수익률 분포의 꼬리 부분이 정규분포보다 두터운 현상을 보이는 점을 알아냈다. 이에 자본시장의 수익률을 정규분포가 아닌 끝이 두터운 안정적 파레시안 분포라고 규정했다 (Osborne, 1964; Mandelbrot, 1964). 이후 수행된 정량적 실험들에 의해, 수익률이 정규분포를 따르지 않는다는 결론에 이른다 (Fama, 1965; Sharpe, 1970). 결국 시장은 임의적이지 않으며,

끊임없이 변화하는 일종의 질서가 있다는 것을 시사한다. 시장이 정적평형상태가 아닌 비선형 동역학계에서 움직이고 있다는 뜻이다.

카오스 (Chaos)는 비선형 결정론적 동역학계 (nonlinear deterministic dynamic system)에서 비롯되는 현상, 그 상태를 말한다. 우리는 시장의 비선형 결정론적 동역학계를 수학적으로 표현할 수 없다. 때문에 시장을 형성하는 비선형 결정론적 동역학계는 일반적인 분석에 많은 오차를 지니고 있어 규칙을 찾아낼 수 없을 뿐만 아니라 그 움직임을 예측하는 것에도 큰 어려움이 있다. 수익률이 정규분포로 발생하지 않는다는 사실을 우리는 알고 있으며, 수학적이고, 물리적인 함수를 활용하여 현상을 재현하지 못하는 카오스 현상을 띠는 이러한 데이터를 분석하기 위해서는 카오스 분석방법을 활용해야 한다.

Li (2014)는 부동산 개발업을 가정하여 Chaos에 대한 방정식을 수리적으로 접근하였다. 그러나 구체적인 실증분석이 없고 가정에 근거한 수리적 증명의 한계를 두고 있다. Ahn 등 (2012)은 거시적 경제 지표를 바탕으로 GA-ridge regression을 통한 감정평가 지수 예측을 시도했다. 회귀분석, 인공지능망 등과 성과비교 연구를 수행했다. 개별 가격 예측이 아닌 포괄적 지수 및 추세 예측의 한계가 있었다. Choi와 Kim (2013)은 울산시 개발제한구역 집단취락지구를 대상으로 ridge regression을 이용하여 가격 변동의 결정요인을 분석했다. 국소적 지역에 한정했다는 것과 인공지능 기법 적용 부재의 한계가 있었다. 한편, 경매낙찰가에 미치는 요인에 대한 분석이 있었다. Hedonic price model을 이용하여 부동산 경매 낙찰가에 영향을 미치는 요인들을 분석하였다. 양의 영향을 미치는 요인으로서는 감정가, 응찰자 수, 경과연도로 나타났으며, 음의 영향을 미치는 요인으로서는, 소비자 물가지수, 평형, 특수권리유무, 환율로 나타났다. 그러나 이는 단순 요인 분석에 의미가 있으며 예측 및 인공지능 기법 적용 부재의 한계가 있었다 (Jung과 Yoo, 2010).

부동산 분야에 대한 Chaos 분석 연구로는 ‘카오스 분석을 활용한 주택가격지수의 예측’이 유일하다 (Kim, 2007). 이 연구에서는 각 부동산별 Index를 데이터로써 사용하였다. 국내주택매매지수, 국내주택전세지수, 미국주택가격지수, 미국 Reits 지수 등을 사용하였다. 사용 방법론으로는 BDS test, Hurst 지수, maximum Lyapunov 지수와 같은 Chaos 분석을 활용하였다. 이러한 선행연구를 바탕으로 본 연구에서는 Hurst 지수, 상관 차원 (correlation dimension)과 maximum Lyapunov 지수가 분석 기법으로 사용되었다.

Kim (2007)의 연구는 거시적 부동산 지수를 비교하여 부동산 시장의 예측 가능성을 확인해 보는 데 의미가 있다. 이전 연구와 다르게 본 연구는

첫째, 서울 아파트 경매라는 세부적 영역에 초점을 맞추어 세밀한 분석을 하였으며 둘째, Chaos 분석 기법과 데이터 검증을 일간, 주간, 월간으로 나누어 비교하여 Chaos 분석의 신뢰도를 높였다. 이전 연구는 예측 결과를 비교하여 우수한 모델을 찾는 것에 집중한 반면, 본 연구는 Chaos분석 기법을 통하여 데이터의 특성을 상세히 파악하는 데 그 초점을 두었다는 것에 차별점이 있다.

3. 연구방법

Chaos 분석 방법론 세 가지를 통해 부동산 경매 데이터를 분석 해 보고자 한다. 모델의 구성은 다음과 같다.

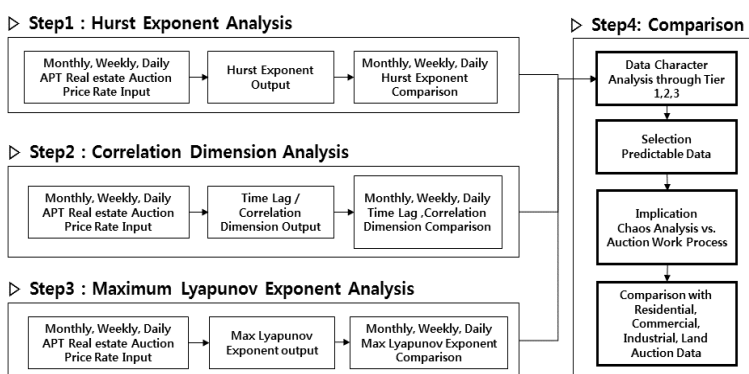


Figure 3 Model architecture

Step 1. Hurst 지수

Hurst는 1907년부터 나일강 저수지의 수위조절을 위하여 40년 동안 저수지의 저수량이 시간에 따라 평균수준에서 어떻게 변동하는지 측정하였고, 그 결과 저수량의 변동범위는 측정 기간의 길이에 따라 변화한다는 사실을 발견하였다. 그는 그 변동범위를 시간에 따라 표준화하기 위하여 관측치의 표준편차로 변동범위의 값을 나누어 나온 무차원 비율을 계산하였다. Brownian motion에서 Hurst 지수 값은 0.5로 Random walks를 따르며 지수가 1과 가까울수록 시계열은 지속적이며 추세 강화적인 형태를 보인다. 반대로 지수가 0에 가까울수록 시계열은 반지속적이며 회귀적인 형태를 보인다. 결국, 이는 시계열에 있어 각 사건들이 해당시점 이후의 사건들과 상관되어 있는지 여부를 구분할 수 있는지에 대한 방법론으로 시계열 데이터가 Random walks인지 Biased random walks인지를 판단한다 (Hurst, 1951; Mandelbrot, 1969). Hurst지수를 구하는 방법은 아래와 같다.

$$R/S = (C * N)^H \quad (1)$$

$$\log(R/S) = H * \log(N) + \log(C) \quad (2)$$

식 (1)의 R/S 는 rescaled range값으로 누적 편차의 범위인 R 값과 그것을 표준편차(S)로 나눈 것을

의미한다. C 는 상수를 의미하며, N 은 관측치의 개수, H 는 Hurst 지수를 의미한다.

$$R = \text{Max}(X_{t,N}) - \text{Min}(X_{t,N})$$

$$X_{t,N} = \sum_{u=1}^t (e_u - M_N)$$

$X_{t,N}$ 은 N 기간 동안의 누적 편차, e_u 는 u 년도의 유입량, M_N 은 N 기간의 e_u 평균을 의미한다. 기존의 식 (1)에 양변에 $\log$ 를 취하면, 식 (2)와 같은 $\log/\log$ 그래프의 기울기가 Hurst 지수가 된다.

Step 2. Correlation Dimension

Mandelbrot이 제시한 해안선의 차원을 계산하는 방법으로, 실제 기하학적인 형상에서 프랙탈 차원을 계산하는 것은 어렵지 않다. 그러나, 실제 시계열 데이터에서 차원을 구하는 것은 기하학적인 방법으로는 구할 수 없는 문제가 있다. 이런 시계열 데이터의 차원을 현실적으로 계산하기 어렵다는 점을 고려해, 차원을 추정하는 방법인 correlation dimension을 활용한다. 끌개 (attractor)에 놓인 한 쌍의 점이 특정 거리 R 안에 존재할 확률을 계산하는 상관적분 (correlation integral)을 구하고, Chaos 시스템에 주어진 내재차원 (embedding dimension)에 대해, 특정 거리 R 이 증가함에 따라 지수함수적으로 증가하는 correlation dimension값을 이용하여, 실제 프랙탈 차원의 근사값을 계산하게 된다. 시계열의 특성을 몇 개의 요인으로 정의할 수 있는가를 나타내는 방법으로, 결정론적 시계열인지 확률적 시계열 데이터인지 구별 가능하다. 즉 시계열 안에 흐르는 구조, 영향을 주는 요소를 파악한다 (Peter와 Itamar, 1983). 내재차원이 증가함에 따라 correlation dimension 값의 변화 양상을 관찰하고, 그래프의 값을 수치화하여, 얼마에 근사한다고 말할 수 있는지를 확인한다.

$$C_m(R) = (1/N^2) * \sum H(R - |X_i - X_j|) \quad (3)$$

식 (3)의 $H(X)$ 는 $X > 0$ 인 조건에서 1, 그렇지 않을 경우 0을 의미하는 Heaviside함수이다. N 은 시계열 자료의 수를 의미하고, R 은 특정 거리, C_m 은 내재차원 m 에서의 상관적분 값 m 은 내재차원을 의미한다.

$$C_m = R^D \quad (4)$$

$$\log(C_m) = D * \log(R) + a \quad (5)$$

식 (4)처럼 주어진 차원 m 에서의 상관적분 C_m 은 R 이 증가함에 따라 R^D 의 비율로 증가한다는 성질을 이용하여, 식 (5)처럼 양변에 $\log$ 를 취해 D 를 구할 수 있다. correlation dimension에 민감한 영향을 주는 Time lag의 경우 최소값인 1부터 데이터의

개수까지의 time lag를 대입해가며, 자기상관계수 (autocorrelation coefficient)를 계산하고, $1/\exp$ 값을 그 threshold로 하여 time lag를 결정한다.

Step 3. Maximum Lyapunov 지수

Lyapunov 지수는 단위 시간당 데이터의 정확성의 소실량을 뜻한다. 즉 예측능력의 손실 측정이 가능한 방법론이다. 결정론적 카오스에 대한 검증과 동시에, 초기 조건에 대한 민감도를 확인할 수 있다. 카오스과정을 따르는 시스템이라면, 초기 상태의 작은 오차로 인하여, 시간에 따라 다른 궤적에 다다르게 된다. 동작방정식 (The equation of motion)을 알 수 없는 현실에서는 Lyapunov 지수의 모든 스펙트럼을 계산할 수 없기 때문에 maximum Lyapunov 지수를 계산하고 활용한다 (Wolf et al., 1985). maximum Lyapunov 지수를 구하는 방법은 다음과 같다.

$$L_1 = \lim_{t \rightarrow \infty} \lim_{L'(t_0) \rightarrow 0} (1/t) \ln(|L'(t)|/|L'(t_0)|)$$

L_1 은 최대 리아푸노프 지수, $L'(t)$ 는 t 시점에 위상공간상 비선형동역학적 갈라짐 상태, $L'(t_0)$ 는 최초의 상태를 의미한다. 위상공간상 최소한으로 정한 거리 내에서 어떻게 나뉘고 발산해가는지의 행태를 나타내는 값이다. 최초의 시작 벡터에 따라서 발산과 갈라짐이 달라지는데, 이 모든 스펙트럼을 확인하는 것은 현실적으로 불가능하다. 따라서 선형근사 극한값을 통해 계산된 maximum Lyapunov 지수를 활용한다 (Rosenstein et al., 1993).

Step 4. 비교/분석

앞선 분석을 통한, Data 속성 분석을 통해 예측력이 높은 데이터 기간을 선별하고, Chaos 분석 비교와 경매실무 데이터 비교를 통한 함의를 도출한다. 특히 아파트 외 주거용, 상업용, 공업용, 토지 부동산의 경매 낙찰가율에 대한 Chaos 분석을 통해 비교 분석을 시도한다.

4. 실증 분석

데이터는 서울 아파트 경매 낙찰가율 데이터를 활용했다. 월별의 경우 2001.1. ~ 2016.9.까지 약 17년 동안의 데이터를 활용하며 그 개수는 201개, 주별의 경우 2001.1.2. ~ 2016.9.30.까지 약 17년 동안 818개, 일별의 경우 2001.1.2. ~ 2016.9.30.까지 약 17년 동안 2,847개의 데이터를 사용한다.

Table 3 Empirical data

Data	Period	Number
① Monthly	2001.1. ~ 2016.9.	201
② Weekly	2001.1.2. ~ 2016.9.30.	818
③ Daily	2001.1.2. ~ 2016.9.30.	2,847

월별 낙찰가율은 0.80~0.92에 대부분 분포되어 있으며, 주별 낙찰가율은 0.81~0.84 인근에 집중되어 있다. 또한 일별 낙찰가율은 거의 정규분포곡선을 따르며 0.82~0.84를 중심으로 데이터가 형성된다.

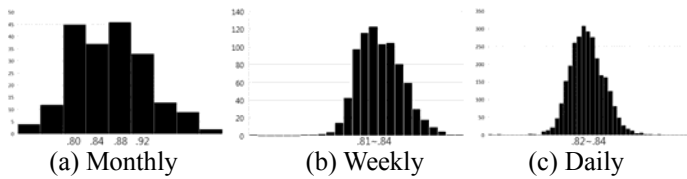


Figure 4 Auction data histogram

4.1 Hurst 지수

각각의 데이터에 대해 Hurst 지수를 측정한 결과, Monthly 시계열 분석의 경우 Hurst 지수가 0.8994의 결과를 보였다. 지수 값이 0.5를 넘는 상태이므로 지속적 추세 강화 시계열의 패턴이 형성되었다. 이는 아파트 경매 최근 낙찰가율에 대한 패턴이 다음 사건에서도 유사하게 형성될 가능성이 높으며 그 확률은 약 89.94% 라는 의미이다. 즉, 과거의 정보를 이용하여 아파트 경매를 참여할 때 예측값의 유의성이 높고 반대로 위험은 줄어든다고 볼 수 있다. 마찬가지로 weekly 시계열 분석의 경우 Hurst 지수가 0.8405, daily 시계열 분석의 경우 Hurst 지수가 0.9581이 도출되었다. Hurst 지수는 장기 의존성이 존재하는 random walk에 대한 상태를 지수화 시킨 것이다. 변동범위를 시간에 따라 표준화하기 위하여 관측치의 표준편차로 변동범위의 값을 나누어 나온 무차원 비율을 R/S라고 표현하였고, 이 R/S분석 우변의 지수가 Hurst 지수가 된다. R/S분석결과 Figure 5의 기울기 값이 Hurst 지수를 의미한다. 일별 데이터의 경우 가장 높은 Hurst 지수 값을 보였으며, 월별, 주별, 일별 세 데이터 모두 지속적 추세 강화적 데이터로 높은 예측 가능성을 갖는다.

Table 4 Monthly, Weekly, Daily Hurst exponent comparison

Monthly	Weekly	Daily
0.8994	0.8405	0.9581

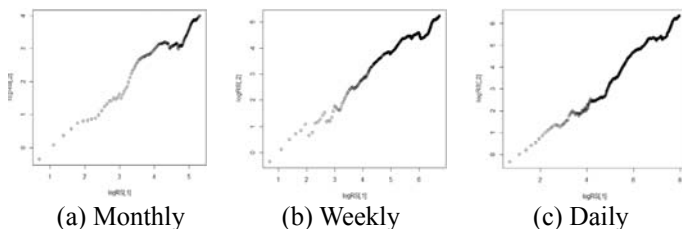


Figure 5 Hurst Exponent

4.2 Correlation Dimension 분석

각각의 데이터에 대해 correlation dimension을

계산했다. 월별 시계열 분석의 경우, time lag 9, embedding dimension 10이 가장 적합한 것으로 확인되었다. time lag가 9라는 의미는 지난 9개월의 영향을 받아 10개월째부터 fractal 현상이 보였음을 의미한다. 앞선 조건으로부터 correlation dimension은 2.795 값이 도출되었다. Correlation dimension은 embedding dimension 10에 대한 기울기 값을 의미하며 fractal dimension을 추정하기 위해 사용되는 방법이다. 이로부터 2.795 차원의 계(系)를 확인할 수 있었으며, 영향을 주는 변수는 2.795보다 큰 첫 정수인 3개로 해석된다. 즉, 서울 아파트 경매 월별 낙찰가율은 9개월의 영향을 받아 10개월째부터 일정한 패턴이 나타나고 약 3개의 변수가 낙찰가율에 영향을 미친다고 볼 수 있다. 주별 시계열 분석의 경우 time lag 23, embedding dimension 24 값이 도출되었고 correlation dimension의 경우 0.418이 도출되었다. 즉, 서울 아파트 경매 주별 낙찰가율을 보면 23주의 영향을 받아 24주째부터 일정한 패턴이 나타나고 약 1개의 변수가 낙찰가율에 영향을 미쳤다. 일별 시계열 분석의 경우, time lag 28, embedding dimension 29 값이 도출되었고, correlation dimension은 0.266 었다. 즉, 서울 아파트 경매 일별 낙찰가율을 보면 28일의 영향을 받아 29일째부터 일정한 패턴이 나타나고, correlation dimension에 의해 약 1개의 변수가 낙찰가율에 영향을 미쳤다.

Table 5 Monthly, Weekly, Daily correlation dimension comparison

	Monthly	Weekly	Daily
Corr. dim.	2.795	0.418	0.266
Time lag	9	23	28

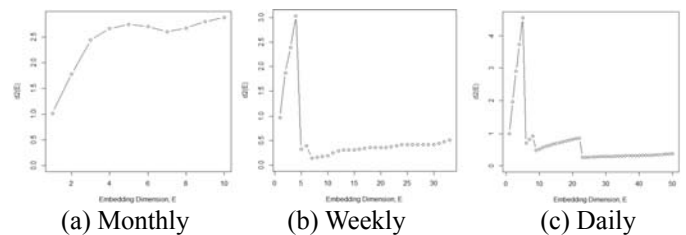


Figure 6 Correlation dimension

4.3 Maximum Lyapunov 지수 분석

각각의 데이터에 대해 maximum Lyapunov 지수 분석을 계산했다. 월별 시계열의 maximum Lyapunov 지수는 0 bit/month가 나왔다. 이는 정보가 손실되지 않고, 과거의 정보가 무한한 미래에도 영향을 미친다는 의미이다. 즉, maximum Lyapunov 지수가 0, Hurst 지수가 높게 유지된다면 정보가 손실되지 않아 예측능력의 손실이 없다. 주별 시계열의 maximum Lyapunov 지수는 0.006076443 bit/week로 1주 당 약 0.0061 bit의 정확성이 소실됨과 1bit의 정확성이 164.57 week/bit로 계산된다. 일정 시간이

흐르면서 현재 정보의 영향력이 해당 지수만큼 소멸됨을 의미하며, 약 165주 이후에는 지금의 조건을 이용해서 미래를 예측하는 것은 불가능하다. 일별 (daily)의 경우, maximum Lyapunov 지수가 0.003940475 bit/day로 1일 당 약 0.0039 bit의 정확성이 소실되며, 1bit의 정확성이 253.77 day/bit임을 알 수 있다. 약 254일 이후에는 지금의 조건을 이용해서 미래를 예측할 수 없다.

정확성 소실률은 monthly가 가장 적고, weekly가 소실률이 가장 크다. 따라서 bit당 정확성 역시 monthly가 가장 양호하다. 실제 시계열 자료는 노이즈가 포함된 한정된 수의 자료이며 nonlinear dynamic system은 장기 기억과정을 가지므로, 단순한 자료의 양이 아닌 장기간 자료가 분석 효과를 높이는 것을 확인했다. 실험결과 경매 monthly data가 경매 weekly, daily data 보다 낮은 maximum Lyapunov 지수를 보여 정보가 거의 소실되지 않아 우수한 예측능력이 있을 수 있음을 알 수 있다.

Table 6 Monthly, Weekly, Daily maximum Lyapunov exponent comparison

	Monthly	Weekly	Daily
Exponent	0	0.00607	0.00394
Accuracy per bit	∞	165	254

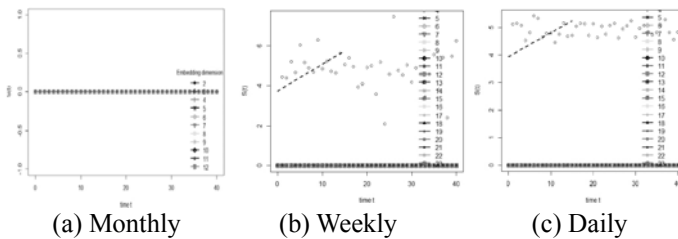


Figure 7 Maximum Lyapunov exponent

4.4 Chaos 분석을 통한 경매 실증 분석

서울 아파트 경매의 경우 낙찰이 되기까지 소요 기간은 Table 7과 같이 정리될 수 있다. Table 7에서 볼 수 있듯이 개시결정일 (approval auction start day)로부터 낙찰일 (bidding day)까지는 약 271일 (9.06개월), 배당요구종기일 (last request day of allocation)로부터 낙찰일 (bidding day)까지는 160일 (22.9주) 정도가 일반적으로 소요된다.

Correlation dimension은 시계열 안에 흐르는 구조와 영향을 주는 요소를 파악할 수 있다. 특히 이중 time lag는 fractal 현상이 나타나기까지의 필요한 기간으로 볼 수 있는데 본 연구에서 경매 낙찰가율의 fractal 현상이 나타나기까지는 Table 8과 같은 time lag가 나타나고 있다.

Table 7, Table 8 그리고 Figure 8를 볼 때 경매 실무상 개시결정일부터 낙찰일까지 약 271일 소요되고 이것은 correlation dimension상의 monthly time lag 9개월 (270일)과 일치한다고 볼 수 있다. 또한 경매 실무상 배당요구종기일부터 낙찰일까지 약 160일

소요되고 correlation dimension상의 weekly time lag 23주 (161일)와 일치한다.

Table 7 2001.1. ~ 2016.9. Seoul APT auction period

	Total period	Sales number	Average period
Approval auction start day -Bid	9,546,634	35,133	271.73
Last request day of allocation-Bid	4,844,220	30,122	160.82

Table 8 Correlation dimension time lag

	Monthly	Weekly
Time lag	9 months(270 days)	23weeks(161 days)

즉, correlation dimension상의 monthly time lag를 통해 개시결정일로부터 낙찰일까지의 데이터로 fractal 현상이 나타나 그 이후의 예측이 가능하다고 주장할 수 있다. 마찬가지로 correlation dimension상의 weekly time lag를 통해 fractal 현상이 검증되어 그 이후의 예측이 가능하다고 주장할 수 있다. 결국, Chaos 분석의 결과는 경매의 fractal 현상이 나타나는 기간을 검증하여 그 기간 이후에 일정한 패턴이 나타나고 예측이 가능하다는 구조적 설명이 된다.

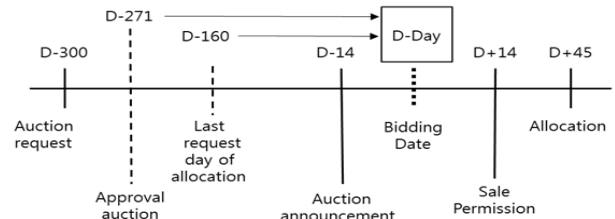


Figure 8 Real estate auction process

4.5 용도별 경매 부동산의 Chaos 분석 비교

지금까지 서울 아파트 경매물건에 대한 Chaos 분석을 시도 했다. 이에 대한 비교 분석을 시도하고자 동일한 기간(2001.1.~2016.9.) 서울 지역 주거용, 상업용, 공업용, 토지 부동산 경매에 대한 Chaos 분석을 추가로 진행하여 비교해 보고자 한다. 우선 Hurst 지수의 경우 주거용 부동산은 0.8805, 상업용 부동산은 0.7619, 공업용 부동산은 0.8518, 토지는 0.9664를 보였다. 모두 0.5를 넘는 값으로 Figure 9와 같이 지속적 추세 강화 시계열의 패턴이 형성되었다.

또한 correlation dimension 상에서 time lag는 주거용 부동산은 13, 상업용 부동산은 2, 공업용 부동산은 11, 토지는 2를 보였다. 주거용과 공업용의 경우 실증적인 값과 유사한 면을 보이나 상업용과 토지의 경우는 차이가 있는 것으로 보였다.

Maximum Lyapunov exponent의 경우 모두 0 값을 보여 정보가 거의 소실되지 않는 우수한 예측능력이 있음을 알 수 있다.

서울 아파트 경매 물건과 기타 용도별 부동산에

대한 Chaos 분석 비교를 통해서 Hurst 지수 및 Maximum Lyapunov exponent는 모두 지속적 추세 강화 및 정보의 무손실을 보인다는 측면에서 동일한 면을 보이나, correlation dimension 상의 time lag는 아파트 경매가 타용도에 비해 실증분석 결과가 보다 정확하다는 것을 확인 할 수 있었다. 이를 종합해보면 서울 지역의 아파트 경매 Chaos 분석 실증분석 결과가 우수한 것으로 판명 되었다.

Table 9 Chaos analysis for auction of each purpose used

Purpose of use	Hurst exponent	Correlation dimension time lag	Maximum Lyapunov exponent
Residential	0.8805	13	0
Commercial	0.7619	2	0
Industrial	0.8518	11	0
Land	0.9664	2	0

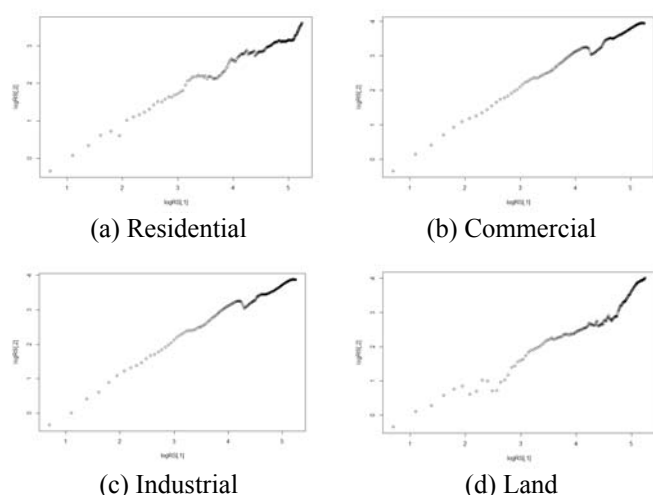


Figure 9 Hurst exponent for auction sale price rate of each purpose used

4.6 Chaos 분석의 경매 실무에 대한 유용성

Chaos 분석은 인공지능 기법으로 만들어진 모델링에 대한 기초가 되는 분석이다. 데이터에 포함된 노이즈로 인해 결정론적 카오스 형태를 띠는 경우, Chaos 분석은 전통적인 통계방법으로 해결할 수 없었던 데이터에 대하여, 인공지능을 통한 예측을 시도할 때 신뢰성을 높여줄 수 있는 근거가 된다. 데이터의 기억효과와 그 강도를 이용하여, 실제 부동산 경매 낙찰가율의 예측을 위한 시도에서, 인공지능을 통한 학습 및 예측에 활용할 기간을 결정할 수 있다. 또한 경매개시결정과 배당요구종기가 경매 낙찰가를 형성하는데 중요한 시작점이 되고 있다는 것을 Chaos 분석을 통해 확인한 바, 이 기간을 단축시키는 것이 경매 진행 과정의 효과적인 개선의 근거가 될 수 있다는 의미로, 이에 대한 최근 각종 규제 개선의 정책에도 부합되는 논리적 근거를 제공할 수 있다.

5. 결론 및 제언

Hurst 지수는 시계열에 있어서 각 사건들이 해당시점 이후의 사건들과 상관되어 있는지를 구분할 수 있는 기준이므로, 시계열이 random walks인지 biased random walks인지 판단하는 지표가 된다. Hurst 지수 값이 0.84 ~ 0.95까지 나타난 것은 아파트 경매 최근 낙찰가율에 대한 패턴이 다음 사건에서도 유사하게 형성될 가능성이 높으며 그 확률은 약 85% 이상이라는 의미로 과거의 정보를 이용하여 아파트 경매를 참여할 때 예측값의 유의성이 높고, 위험이 줄어든다는 의미이다.

Maximum Lyapunov 지수는 거의 0에 가까운 값을 나타내는데 maximum Lyapunov 지수는 단위 시간당 정확성 소실량 (예측능력의 손실 측정)을 나타내며, 0이라는 값은 정보가 거의 손실되지 않음을 의미한다. 즉 Lyapunov 지수가 0에 가깝고, Hurst 지수가 85% 이상 유지된다면 정보가 손실되지 않아 예측능력의 손실이 거의 없다. 실험결과에서처럼 correlation dimension 상의 time lag가 경매 실무와 거의 일치하고 Hurst 지수로 추세가 형성됨과 maximum Lyapunov 지수로부터 예측 시 정보의 손실이 거의 없다는 결과로부터, Chaos 분석을 통한 부동산 경매 및 부동산 시장에서의 예측이 유용함을 확인했다.

현재의 값에 영향을 주는 시간 (time lag)은 28일, 23주, 9개월 등으로 나타남을 볼 때 결정론적 (deterministic) 시계열로 구분된다. 단위 시간당 정확성 소실량을 보면, 월별 데이터를 볼 때 정보 손실이 되지 않아 (Lyapunov 지수) 예측능력의 손실이 거의 없음을 볼 수 있으며 주별, 일별 데이터에서도 253일, 164주 이상의 정보 정확성이 유지되어 예측에 사용할 자료로써의 신뢰성을 확인할 수 있었다.

또한 주거용, 상업용, 공업용, 토지 부동산에 대해서도 Hurst 지수와 Maximum Lyapunov 지수는 아파트 경매와 동일하게 지속적 추세 및 무손실성을 보이고 있다. 반면, correlation dimension 상의 time lag는 아파트에 대한 분석이 각 용도별 분석에 비해 보다 정확하게 나타나고 있다.

본 연구는 경매 낙찰가율에 영향을 주는 요소를 배제한 채 낙찰가율의 시계열적 특성만을 가지고 분석을 한 것에 대한 한계가 있다. 향후 각 경매 개별 건에 대해서 낙찰가율에 영향을 주는 변수들을 종합하여 예측의 정확성을 높이고자 하는 인공지능망과 유전자알고리즘 등의 인공지능 기법을 연구해보고자 한다.

Reference

Ahn, J. J., Byun, H. W., Oh, K. J. and Kim, T. Y. (2012). Using ridge regression with genetic algorithm to enhance real estate appraisal forecasting. *Expert Systems with Applications*, **39**, 8369- 8379.

- Bachelier, L. (1964). Theory of speculation. *The random character of stock market prices*, P. Cootner, ed., MIT Press, Massachusetts.
- Choi, Y., Kim, H. J. and Lee, G. E. (2013). A study on the factor of the determination to the land value change using ridge regression. *Korea Real Estate Academy*, **52**, 313-326.
- Cootner, P. (1964). Comments on the variation of certain speculative prices. *The random character of stock market prices*, P. Cootner, ed., MIT Press, Massachusetts.
- Fama, E. F. (1964). Mandelbrot and the stable Paretian hypothesis. *The random character of stock market prices*, P. Cootner, ed., MIT Press, Massachusetts.
- Fama, E. F. (1965). Portfolio analysis in a stable Paretian market. *Management Science* **11**, 404-419.
- Fama, E. F. (1965). The behavior of stock market prices. *Journal of Business* **38**, 34-105.
- Grassberger, P. and Procaccia, I. (1983). Measuring the strangeness of strange attractors. *Physica D: Nonlinear Phenomena*, **9**, 189-208.
- Hurst, H. E. (1951). Long-term storage capacity of reservoirs. *Transactions of the American Society of Civil Engineers*, **116**, 770-799.
- Jung, G. Y. and Yoo, S. J. (2010). A study on decision factors of the apartment bid price in real estate auction market. *Housing Studies Review*, **18**, 29-50.
- Kim, D. U. (2007). *Using Chaos analysis to forecast real estate indexes*, Master Thesis, Yonsei University, Seoul.
- Kim, H. J. (2016). Study on equity of taxation for non-residential property by analysis of actual transaction price. *Journal of the Korean Data & Information Science Society*, **27**, 639- 651.
- Li, H. (2014). Cellular neural networks model on real estate investment systems and its stability and Chaos synchronization. *The Open Automation and Control Systems Journal*, **6**, 200-202.
- Mandelbrot, B. B. and Wallis, J. R. (1969). Robustness of the rescaled range R/S in the measurement of noncyclic long run statistical dependence. *Water Resources Research*, **5**, 967- 988.
- Mandelbrot, B. B. (1964). The variation of certain speculative prices. *The random character of stock market prices*, P. Cootner, ed., MIT Press, Massachusetts.
- Osborne, M. F. M. (1964). Brownian motion in the stock market. *The random character of stock market prices*, P. Cootner, ed., MIT Press, Massachusetts.
- Rosenstein, M. T., Collins, J. J. and De Luca, C. J. (1993). A practical method for calculating largest Lyapunov exponents from small data sets, *Physica D*, **65**, 117-134.
- Sharpe, W. F. (1970). *Portfolio theory and capital markets*, McGraw-Hill, New York.
- Wolf, A., Swift, J. B., Swinney, H. L. and Vastano, J. A. (1985). Determining Lyapunov exponents from a time series. *Physica D*, **16**, 285-317.