https://doi.org/10.15813/kmr.2021.22.4.007

# 카오스 이론 기반 시계열의 내재적 패턴분석: 룰렛과 KOSPI200 지수선물 데이터 대상

## Analysis of Intrinsic Patterns of Time Series Based on Chaos Theory: Focusing on Roulette and KOSPI200 Index Future

이희철 (HeeChul Lee) 연세대학교 공학대학원[1]
김홍곤 (HongGon Kim) 연세대학교 공학대학원[2]
김희웅 (Hee-Woong Kim) 연세대학교 정보대학원[3]

〈 국문초록 〉

각 산업에서 대량의 데이터가 생산되면서, 빠른 경영 의사결정을 위해 시계열 패턴 예측 연구가 수많이 진행되고 있다. 하지만 데이터에 내재된 불확실성으로 인해 비선형 시계열 데이터의 특정 패턴을 예측하는 데 한계가 존재하고, 기업경영의 전략적 의사결정 어려움이 존재한다. 또한, 최근 수십 년간 불규칙한 랜덤워크 모형의 시계열 데이터 예측을 위해 산업의 목적에 맞는 금융시장 데이터를 대상으로 다양한 연구가 진행되고 있지만, 특정 규칙을 예측하고 지속가능의 기업목적 달성 어려움이 있다. 본 연구에서는 룰렛 데이터와 금융시장 데이터를 Chaos 분석기법을 이용하여 예측 결과를 비교분석하고 유의미한 결과를 도출하였다. 그리고, 본 연구는 카오스 분석이 시계열 자료를 분석하는데 있어 새로운 방법을 모색하는데 유용함을 확인하였다. 룰렛 게임의 특성을 한국 주가지수 선물의 시계열과 비교 분석하여 추세가 확인되는 경우 예측력을 높일 수 있다는 점을 도출하였으며, 불확실성이 높고 랜덤워크가 존재하는 비선형 시계열 데이터가 특정한 패턴을 가지고 있는지 판단하는데 의의가 있다.

주제어: 빅데이터, 시계열 분석, 비선형데이터, 카오스 이론, 허스트 지수, 상관차원, 최대 리아프노프 지수

---

1) 제1저자, setno@daum.net
2) 제2저자, honggon.kim@yonsei.ac.kr
3) 교신저자, kimhw@yonsei.ac.kr

논문접수일 2021년 11월 08일 | 1차 수정 2021년 12월 06일 | 게재확정일 2021년 12월 09일

## 1. 서론

데이터 과학의 발전으로 금융, 물류, 의료, 교육 등의 산업뿐만 아니라, 레저 및 엔터테인먼트 산업까지 다양한 분야에서 빅데이터가 생산되고 있으며, 이를 활용한 연구가 최근에 활발히 진행되고 있다. 빅데이터는 데이터의 생성과정과 데이터 활용목적에 따라 정형, 반정형, 비정형 데이터로 구분(최봉 등 2019)되며, 데이터의 형태와 상관없이 무수히 많은 자료 속에서 의미가 있는 결론을 도출하는 것이 관건이라고 볼 수 있다.

빅데이터를 활용한 연구는 다양하다. 한가지 예로, 차별화된 고객 관계 관리(CRM) 전략을 위해 고객 군을 세분화 및 맞춤형 전략을 세우고(Run-Qing Liu et al. 2018) 소비자의 SNS 데이터를 분석해 기업의 브랜드 컨셉 개발에 응용(이주민 등 2020)하기도 한다. 뿐만 아니라 빅데이터를 통해 정부의 대응에 관한 대중의 견해를 파악하고 위기상황에서 정부와 관련 기관이 국민의 필요와 요구에 따른 정확한 대응책을 마련할 수 있도록 기초자료로서 제시(김진솔 등 2021)될 수 있다. 하지만 빅데이터 분석을 통해 신뢰성 있는 유의미한 패턴을 발견하기는 어렵다.

그럼에도 불구하고 빅데이터는 앞서 언급한 것과 같이 다양한 분야에서 활용되는데, 빅데이터에 대한 한가지 대표연구로 시계열 데이터 분석을 적용시켜 빅데이터의 패턴을 분석하는 것이다. 김미형(2010)의 연구에서는 시계열 데이터 분석을 통해 금융 시장의 미래 패턴을 예측하여 기업의 잠재적 수익성을 높일 수 있는 주가 예측 연구를 수행하였다. 또한, 빅데이터인 부동산 뉴스가 아파트 매매가격과 아파트거래량에 미치는 영향을 ‘상승’과 ‘하락’과 관련된 기사수를 변수로 벡터자기회귀모형(VAR)을 통해 분석하였으며, 빅데이터를 활용해 주택시장 예측프로그램을 구축할 수 있다는 단서를 제공(전해정, 2020)한 연구도 있다.

일반적으로 빅데이터 분석은 원인과 결과에 있어 그 결과가 비 선형적인 경우가 많다. 이러한 결과가 발생하는 가장 큰 이유는 시간의 흐름에 따라 천재지변, 정치, 사회, 문화 등에서 여러가지 우연한 사건이 발생하여 오차를 발생시킬 수 있기 때문이다. 또한 불규칙성을 가지는 시계열 데이터가 대량 생산되면서, 이에 따른 데이터에 내재된 불확실성이 증가하여 특정 패턴을 판별하는데 어려움이 존재(박거준 등 2017)하여, 의사결정 리스크가 발생하게 된다. 이러한 비선형의 결정적 과정을 ‘카오스’로 일컬으며, 결정적 체계를 가지면서도 확률적 체계가 만들어 낼 수 있는 것과 동일한 특성을 가지는 시계열을 생성해 낼 수 있는 특징을 갖는다(최영일, 2001).

본 연구에서는 빅데이터분석에 있어 시계열분석방법을 적용시킬 수 있다는 점과 그 결과가 대체로 비선형적인 모습으로 발현되는 것에 착안하여 빅데이터 분석에 있어 카오스이론을 적용하였다. 이를 위해, 시간이 지날수록 불확실성이 높은 룰렛과 금융 시장의 시계열 데이터를 선정하여 카오스이론 기반으로 시계열의 카오스적 성질을 분석하고, 시계열의 장기 기억 속성과 자기 유사성을 검정하는 연구를 진행했다.

관련 선행연구로는 비모수 회귀분석 기법 중 하나인 국부적 가중회귀를 사용하여 조건부 평균의 변화를 잘 탐색할 수 있는 카오스적 끌개를 모형화 한 연구(이상빈, 최우석, 1995)가 있으며, 주가가 어떤 형태의 경제 구조로부터 생성되었는지를 업종별 주가지수의 비선형 검정으로 검정한 연구(백웅기, 1997)가 있다.

주식시장에 비선형성이 존재하고, 주가는 정규분포를 따르면서 주식시장에 영향을 준 충격들이 남아있어 비선형성이 발생한다는 것을 보인 연구(이일균 1998), Whang Linton 방법으로 주식시장의 카오스존

재를 검증한 연구(백웅기 등 1999)등이 있다. 또한, 코스피(KOSPI) 일별 수익률과 주별 수익률로 비선형성과 카오스검증을 수행한 연구(장경천 등 2002)와 부동산 시장의 경매 낙찰가율에 대한 시계열 자료로 카오스분석을 진행한 선행연구(강준 등 2017)가 있다.

카오스이론을 빅데이터와 같이 예측이 불가능해 보이고, 일정한 규칙이 없어 보이는 룰렛게임과 KOSPI200 주가지수에 각각 적용시켜 봄으로써 향후, 빅 데이터 분석에 있어 카오스이론의 적용가능성에 대한 기초연구를 수행하고자 하였다. 큰 관련성이 없어 보이는 룰렛 게임과 KOSPI200 주가지수를 동일선상에 두고 비교분석을 수행한 이유는 룰렛 게임이 가지고 있는 일정한 규칙성과 KOSPI200 주가지수가 보이는 가격흐름의 변동성이 큰 규칙하에 움직이고 있기 때문이다.

특히, 본 연구의 차별성은 기존의 관련 선행연구에서 연구대상으로 고려된 적 없는 룰렛 게임을 빅 데이터처럼 분석에 활용함에 있다. 이와 함께 연구대상으로 자주 고려되었던 KOSPI200 주가지수를 함께 비교함으로써 향후, 규칙이 없어 보이는 다양한 연구대상을 분석하는데 있어 시계열분석과 카오스이론의 적용방법 및 가능성을 보여주고자 하였다. 만약 룰렛 게임과 금융 시장의 불규칙한 시계열 자료에 존재하는 문제를 기술적으로 해결할 수 있다면, 불확실성 문제를 해소하고 효율적인 경영 의사결정 체계를 갖추는데 큰 도움이 될 수 있을 것이다.

## 2. 개념적 배경

### 2.1. 카오스 이론

카오스 이론의 기원은 Lorenz가 1963년 기상예측을 위하여 컴퓨터를 활용한 시뮬레이션을 하는 과정에서 발견되었으며, 결정론적 현상계로 정의 내려지면서 매우 복잡한 불규칙이 존재하면서 불안정한 특징을 보인다는 이론(R. A et al. 1995)이다. 시계열분석에서의 카오스이론은 일반적인 시계열데이터의 비선형성을 해석하고자 연구되었다. 시계열 분석에서는 선형추계학 모형인 자동회귀/이동평균(ARMA)모형이 주로 사용(Salas et al., 1980)되었는데 이러한 선형모형은 비선형 구조를 가지는 시계열 자료를 해석함에 있어 자료의 비선형성을 고려하지 못하는 한계를 갖게 된다(최강수 등 2009). 이러한 비선형 구조를 해석하는 방법론 중 가장 대표적인 방법론이 카오스 이론이다.

카오스란 결정론적 비선형 시스템으로 표현되는 불규칙하고 예측이 어려운 현상으로 주가동향 분석뿐만 아니라 다양한 분야에서 응용된다(박대규, 조원철, 2003). 카오스는 다양한 분석방법으로 특성파악이 가능하다. 대표적으로 허스트 지수와 상관차원, 최대 리아푸노프 지수 방법이 있으며 본 연구에서도 각각의 분석방법을 룰렛 게임과 KOSPI200주가지수 데이터에 적용해 시계열데이터가 갖는 비선형성에 대한 패턴을 증명하였다.

### 2.2. 허스트 지수 (Hurst exponent)

수리학자인 Harold Edwin Hurst가 나일강의 저수지 수위를 관리하던 1907년부터 나일강 저수지의 저수량을 평균에서 벗어나는 움직임의 변동범위를 측정했고, 시계열의 long term memory의 척도로 사용되는 지표로 체계화 하여 시간의 흐름에 따라 표준화 하였다(Carbone et al. 2004). 즉, 저수량 관측치의 표준편차로 저수량의 변동하는 범위의 값을 나누어 무 차원이라는 개념을 도입해 비율값을 제시한 분석방법이다. 허스트 지수는 분석하고자 데이터가 지속적인 시계열 패턴인지 여부를 분석하는 근거를 제시하여, 시계열

의 지속성을 예측하고, 시계열 데이터로써 활용할 경우 데이터의 신뢰성을 높여주는 기초 분석이라 할 수 있다.

허스트 지수가 커질수록 주가 수익률의 무작위한 정도가 줄어들고, 추세가 지속적일 가능성이 높아진다. 즉, 예측가능성을 높이기 위해 과거의 데이터를 이용해 투자를 하며 위험은 감소하게 되는 반면 H가 0.5에 가까울수록 데이터의 생성과정은 임의적이며 예상할 수 없는 경로를 따르는 랜덤워크에 가까워지는 것으로 판단한다. 관측치의 표준편차로 데이터 변동범위의 값을 나눠서 무 차원 비율을 만들고, 시계열이 랜덤워크를 따르는 지 확인할 수 있는 지표로 장기 의존성 지수를 계산하여 판단하게 된다. 허스트 지수는 시계열에 있어서 각 사건들이 해당 시점 이후의 사건들과 상관성이 있는지를 판단하고 구분할 수 있는 기준이고, 시계열이 Random 이거나 결정론적(Deterministic) Chaotic인지를 판단할 수 있게 해 주는 매우 중요한 판단 지표이다. <표 1>은 허스트 지수의 계산 수식이다. N기간 동안의 누적 편차($X_{t,N}$)를 계산하고, 누적 편차의 범위($R$)를 도출한다. 이 누적 편차의 범위 값($R$)과 누적 편차의 표준편차($S$)를 <표 1>식에 대입하여 허스트 지수($H$)를 구한다.

<표 2>를 살펴보면, 허스트 지수 값이 0~0.5 사이에 있다면 반 지속적인(회귀적인) 시계열로 회귀적 시계열이라 할 수 있으며, 허스트 지수 값이 0.5라면 완전히 비 상관적인 시계열인 랜덤워크를 따르는 시계열로 판단한다. 허스트 지수 값이 값이 0.5보다 크고 1보다 작으면 추세가 있는 시계열로 추세강화 시계열이라고 판단하는 분석법이다.

## 2.3. 상관 차원 (Correlation Dimension)

프랙털 차원(Fractal Dimension)은 1967년 영국 해안선의 길이를 측정하면서 Mandelbrot가 제시한 개념(Aringhaus, S. L, 1985)이다. 해안선 길이를 측정하는 측량자의 기본적인 길이를 ε 으로 가정하면, 측량자로 해안선의 길이를 측정하는 경우 측량자의 기본적인 길이 ε 을 축소시켜 감에 따라 증가해야 하는 측량자의

〈표 1〉 허스트 지수(Hurst exponent) 계산 수식

| Hurst Exponent | 계산 수식 설명 |
|---|---|
| $X_{t,N}=\sum_{u=1}^{t}(e_u-M_N)$<br>$R=Max(X_{t,N})-Min(X_{t,N})$ ,<br>$R/S=(C*N)^H$<br>$\log(R/S)=H*\log(N)+\log(C)$ | * R: 누적 편차의 범위 값<br>* S: 누적 편차의 표준편차<br>* R/S: rescaled range값<br>* C: 상수<br>* N: 관측치의 개수,<br>* H: Hurst Exponent<br>*$X_{t,N}$ : N기간 동안의 누적 편차,<br>*$e_u$ : U년도의 저수지 유입량,<br>*$M_N$ : N기간의 $e_u$ 평균을 의미 |

〈표 2〉 허스트 지수(Hurst Exponent)에 따른 추세 특성

| Hurst Exponent | Correlation | Meaning |
|---|---|---|
| 0 〈 H 〈0.5 | − | 반 지속적이며(Anti-persistent), 회귀적인(Mean reverting) 시계열 |
| H = 0.5 | 0 | 랜덤워크(Random Walk) |
| 0.5 〈 1 | + | 지속적이며(Persistent), 추세강화(Trend reinforcing) 시계열 |

증가율을 확정적인 D라고 가정한다. 일반적으로 확정적인 D는 정수는 아니며, Mandelbroit는 정수가 아닌 차원을 강조하여 프랙털 차원이라고 정의 내렸다. 그는 프랙털 차원을 추정하여 도출한 결과를 통해 시계열이 결정론적(Deterministic)인지 시계열이 확률적(Random) 자료인지 구별 가능하다고 했다. Mandelbrot가 제시한 프랙털 차원을 활용한 해안선의 차원 계산 방식은 실제 시계열 데이터를 활용해 차원을 계산하는 것이 기하학적 방법으로 어렵다는 이슈가 있어, 상관차원이라고 하는 차원을 추정하는 방법을 활용하여 도출하게 된다. <표 3>는 상관차원의 계산 수식이다. $C_m(R)$은 상관적분을 의미하며 끌개(attractor)에 놓인 한 쌍의 점이 거리 $R$안에 존재할 확률을 의미한다. 상관차원(D)는 주어진 차원(m)에서 R이 증가함에 따라 $R^D$의 비율로 증가하는 특성을 이용하여 구할 수 있다.

상관차원은 몇 개의 요인으로 시스템의 특성을 정의할 수 있는지를 도출해 내는 방법론으로써, 이러한 계산으로 어떤 계의 변수 개수를 알아내어 시계열 내에서 움직이는 구조와 영향을 주는 요인을 파악하게 해 준다. 프랙털(fractal) 도형의 의미는 자기 유사성이라는 특성을 내포하는 도형으로 부분과 전체가 유사한 모습으로 구성되어 이러한 도형을 확대 또는 축소하여도 동일한 형태를 지니는 도형을 의미한다. 상관차원은 프랙털 차원을 추정하기 위해 개발된 방법으로 한 시계열의 프랙털 차원이 3이라면 그 시계열은 3개의 변수에 다음 시계열의 특성이 설명될 수 있음을 의미한다.

## 2.4. 최대 리아푸노프 지수 (Maximum Lyapunov Exponent)

복잡계에서 최대 리아푸노프 지수(Maximum Lyapunov exponent)는 데이터가 무질서한 행태를 보이는지 임의적인 행태인지를 구별하는데 사용되는 다양한 방법 중 하나로, 카오스 특성의 가장 중요한 부분인 데이터의 초기 조건과 관련된 민감도의 수준을 의미한다(HolgerKantz, 1994). 리아푸노프 지수의 특징 측면에서 분석해 보면 양(+)의 리아푸노프 지수는 미래에 대한 예측력의 손실이라고 해석할 수 있고, 리아푸노프 지수 측정값이 0.01이라고 가정한다면 1단위 시간이 경과할 때마다 시계열 데이터는 0.01씩 초기값에 대한 특성을 잃어버려가고 있다는 의미이다.

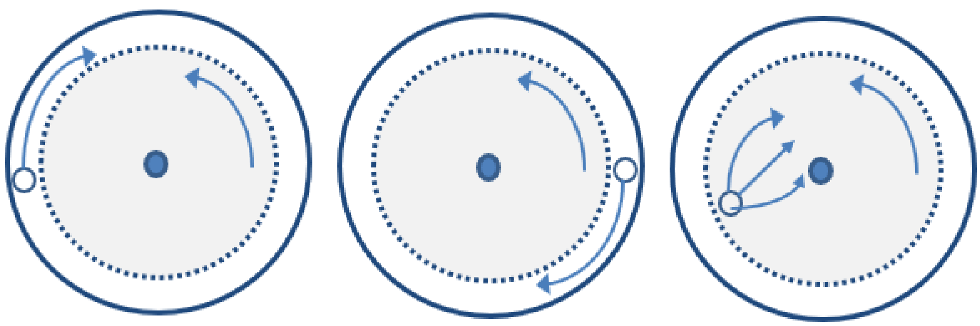
다시 말해, 1/0.01단위의 시간이 경과하면 데이터는 초기값의 특성을 완전히 소실하게 된다고 해석할 수 있으며, 리아푸노프 지수 측정값이 양(+)이라면 초기 조건에 민감한 의존성이 존재한다는 것을 의미하는 것이고 카오스적 끌개가 존재한다는 의미이다. 리아푸노프 지수가 음(-)이라면 안정적이며 확률적인 과정을 따른다는 의미이다. 최대 리아푸노프 지수는 동적 시스템의 예측가능성을 결정하며, 양(+)의 최대 리아푸노프 지수는 시스템이 Chaotic하다는 의미를 가진다는 뜻을 내포하게 된다. <표 4>는 최대 리아푸노프

〈표 3〉 상관차원(Correlation Dimension) 계산 수식

| Correlation Dimension | 계산 수식 설명 |
|---|---|
| $C_m(R) = (1/N^2) * \sum H(R - \|X_i - X_j\|)$<br>$C_m = R^D$<br>$\log(C_m) = D * \log(R) + a$ | * $C_m$ : 내재차원 m에서의 상관적분 값 Cm<br>* $R$ : 특정 거리<br>* $N$ : 시계열의 자료 수<br>* $H(X)$ : X〉0인 조건에서 1, X≤0에서 0인 Heaviside 함수<br>* $C_m = R^D$의 등식과 동일하게 given 차원 m에서 상관적분 $C_m$은 특정 거리 $R$이 증가함에 따라 $R^D$의 비율로 증가하는 특성을 이용함.<br>* $Lo0g(C_m) = D * log(R) + \alpha$ 와 같이 양변에 log를 취해 $D$를 계산함. |

〈표 4〉 최대 리아푸노프 지수(Maximum Lyapunov exponent) 계산 수식

| Maximum Lyapunov exponent | 계산 수식 설명 |
|---|---|
| $L_1 = \lim_{t\to\infty} \lim_{L'(t_0)\to 0} (1/t)\ln(\lvert L'(t)\rvert / \lvert L'(t_0)\rvert)$ | * $L_1$: Maximum Lyapunov exponent.<br>* $L'_{(t)}$ : t시점 위상 공간상 비선형 동역학적 갈라짐 상태.<br>* $L'_{(t0)}$ : 최초의 상태를 의미.<br>* 위상 공간상 최소한으로 정한 거리 내에서 나눠지고 발산해 가는지 행태를 나타내는 값.<br>* 최초 시작 벡터에 따라 발산과 갈라짐이 달라지는데, 모든 것을 확인하는 것은 현실적으로 불가능하기에. 선형 근사 극한값으로 도출된 maximum Lyapunov exponent를 활용함. |




〈그림 1〉 룰렛 구슬의 회전 3단계

지수 계산 수식이다. I 차원 $P_I(t)$에 대한 $I$ 번째 리아푸노프 지수 $L_I$를 나타낸다.

결국, 리아푸노프 지수는 단위 시간당 정확성의 소실량을 의미하며, 결정론적인 카오스적 특징에 대한 검증 및 초기 조건에 대한 민감도를 확인하는 지표로써, 예측능력의 손실 측정이 가능한 방법론이다. 예를 들어, 위상공간에서 리아푸노프 지수가 0.0039일 때 초기에 1bits 정밀도를 갖는다면 1/0.0039=254 로 계산되는데, 의미는 약 254 unit 후에는 현재의 정보로 예측력이 떨어진다는 것이다.

## 3. 실증 분석

### 3.1. 분석 프로세스

룰렛의 궤적은 3단계로 분리가 가능하다. 룰렛 게임 시 게임 딜러가 룰렛판을 돌리면, 최초에 룰렛 구슬은 룰렛판의 끝자리에 붙어서 회전하고, 룰렛판은 반대방향으로 회전하게 된다. 룰렛 구슬에는 구슬이 룰렛판의 가장자리를 떠나지 않게 하는 구심력과 룰렛 구슬을 아래로 당기는 중력의 2가지 힘이 작용하게 된다.

룰렛 게임 시 구슬이 회전하다가 멈추는 룰렛판의 숫자는 0~36 사이이며, 룰렛 게임 시 구슬이 멈추는 숫자판에 대한 룰렛 데이터는 별도로 제공 해주는 기관이나 공개된 자료가 없어 실제로 마카오 카지노 5곳과 룰렛 게임기 10개에서 4,614회의 게임을 실행하여 데이터 수집을 하였고, 데이터 전처리를 하여 주가의 등락 발생처럼 룰렛 게임 구슬의 회전판 착지번호의 등락이 발생하는 것처럼 시계열화 처리해 주었다.

룰렛 게임 시 구슬이 착지하는 번호판의 숫자와 주가지수 선물 수익률과의 특징분석은 카오스 분석을

통해 실행하였고, 룰렛 구슬의 경로와 번호판 착지지점에 대한 예측이 아닌 룰렛 알의 경로가 완전히 무작위가 되는 경우에 중점을 두어 카오스 이론의 분석방법을 활용하였다.

## 3.2. 데이터 전처리

룰렛 게임 결과 나온 구슬이 착지지점 번호판의 자료는 R로, KOSPI 200 주가지수 선물 자료는 K로 표기하고, 가공하지 않은 원본 자료는 1로 데이터 마이닝한 자료는 2로 표기하여 분류하였다. 즉, 원본 룰렛 구슬의 룰렛판 착지 번호의 자료는 R1 시계열, 데이터 마이닝을 한 자료는 R2로 표기하고, KOSPI 200 주가지수 선물의 가공하지 않은 자료는 K1 시계열, 데이터 마이닝을 한 자료는 K2로 표기하여 시계열의 속성이 다른 부분을 비교 분석을 위하여 비슷한 속성을 가지도록 처리해 주었다.

룰렛 게임 결과 나온 구슬이 착지지점 번호판인 R1 시계열은 0과 36사이의 숫자로 이루어져 있으며, K1 시계열은 증가 추세를 지닌 시계열 자료이므로 전처리를 통해 유사한 속성을 지닐 수 있도록 변환해 준 것이다. R1 자료에서 중간 값인 18을 차감해 -18~+18의 값을 갖도록 변환하여 K1의 시계열과 유사속성으로 변환시킨 것이 R2 시계열이며, K1 자료의 시계열마다 현재 밸류 값에서 그 다음 밸류 값을 차감해 폭을 제한하여 R1 시계열과 유사속성으로 변환시킨 값이 K2 시계열이라고 하겠다.

R1 시계열은 0에서 36의 숫자로만 구성된 룰렛 게임의 시계열이나, K1 시계열은 주가지수 선물 가격의 상승과 하락폭이 일별로 발생하고 추세를 가진 시계열이다. R1 시계열을 KOSPI200 주가지수 선물의 K1 시계열과 유사한 속성을 만들기 위해 룰렛 게임의 구슬 착지번호의 중간 값 18을 기준으로 룰렛 게임 전체 데이터에서 각 18을 차감하여 음수와 양수가 발생하도록 0~+36을 -18~+18의 값으로 변환해 주었다. 전처리의 가장 큰 이유는 1차 전 처리 후 룰렛 게임의 룰렛 구슬의 착지 번호 자료를 누적시키면 단순히 계속 양의 부호로 커지는 자료가 발생하여 우 상향하기 때문이다. 따라서, 룰렛 게임의 구슬의 착지번호가 양의 방향과 음의 방향의 변동폭을 모두 갖게 되는 구조로 변환시켜 시계열화 처리했다.

이 자료들을 위해 룰렛 게임의 구슬 착지번호의 R2 시계열이라 정의하며 <표 5>에 간단하게 예시를 표현했다. <표 5>의 2열은 위해 룰렛 게임의 구슬 착지번호의 R1 시계열, 3열은 위해 룰렛 게임의 구슬 착지번호의 R1 시계열 각 시점의 값에서 18을 모두 뺀 것, 4열은 B열의 값을 누적한 변환 값이다.

룰렛 게임의 구슬 착지번호의 R1 시계열은 0~36의

〈표 5〉 R 시계열 전처리 과정

| Index | 원지수 (R1) | 원지수-18 (R2) | 누적지수 |
|---|---|---|---|
| 1 | 27 | 9 | 9 |
| 2 | 24 | 6 | 15 |
| 3 | 31 | 13 | 28 |
| 4 | 36 | 18 | 46 |
| 5 | 16 | −2 | 44 |
| 6 | 29 | 11 | 55 |
| 7 | 33 | 15 | 70 |
| · · · · · | | | |
| Data 총 4614개 | | | |

〈표 6〉 K 시계열 전처리 과정

| Index | 날짜 | 원지수 (K1) | 차분지수 (K2) |
|---|---|---|---|
| 1 | 20000104 | 136.2 | 9.9 |
| 2 | 20000105 | 126.3 | 5.1 |
| 3 | 20000106 | 121.2 | −0.8 |
| 4 | 20000107 | 122 | −4 |
| 5 | 20000110 | 126 | 1.6 |
| 6 | 20000111 | 124.4 | 2.8 |
| 7 | 20000112 | 121.6 | −1.9 |
| · · · · · | | | |
| Data 총 4614개 | | | |

일정한 1 단위의 변화폭 안에서 움직이는 시계열이고, KOSPI200 주가지수 선물 K1 시계열을 R1 시계열과 유사한 속성을 갖도록 변환했다. KOSPI200 주가지수 선물의 K1 시계열 현재 지수 값에서 그 다음 지수 값을 차감하면 어느 정도 등락이 있는 변화율이 제한된 시계열이 된다. 이러한 KOSPI200 주가지수 선물의 데이터 마이닝 된 값을 K2 시계열이라 정의했으며, <표 6>에 간단한 예시를 정리했다. 2열은 KOSPI200 주가지수 선물의 생성일, 3열은 KOSPI200 주가지수의 K1 시계열, 4열은 현재 날짜에서 다음 시점의 값을 뺀 K1 시계열이다.

### 3.3. 분석 단계

분석은 크게 두 가지 단계로 구성되며, 그 구조는 <그림 2>에 나타냈다. 첫 번째는 룰렛 게임의 구슬 착지번호 자료와 KOSPI 200 주가지수 선물 자료를 수집하는 과정이다. 두 번째 단계는 R1, R2, K1, K2 시계열에 대해 카오스 분석을 하는 단계이다. 카오스 분석은 첫 번째 분석으로 허스트 지수를 도출하고 계산하여 추세성을 확인했다. 다음으로는 룰렛 게임의 구슬 착지번호 자료와 KOSPI 200 주가지수 선물 자료의 시계열의 상관차원을 도출하였고, 마지막 단계는 각 시계열에 대한 최대 리아푸노프 지수를 계산하고 데이터를 활용해 다음에 실행될 게임의 결과를 예상할 수 있는 과정을 포함하여 분석했다.

## 4. 실증분석결과

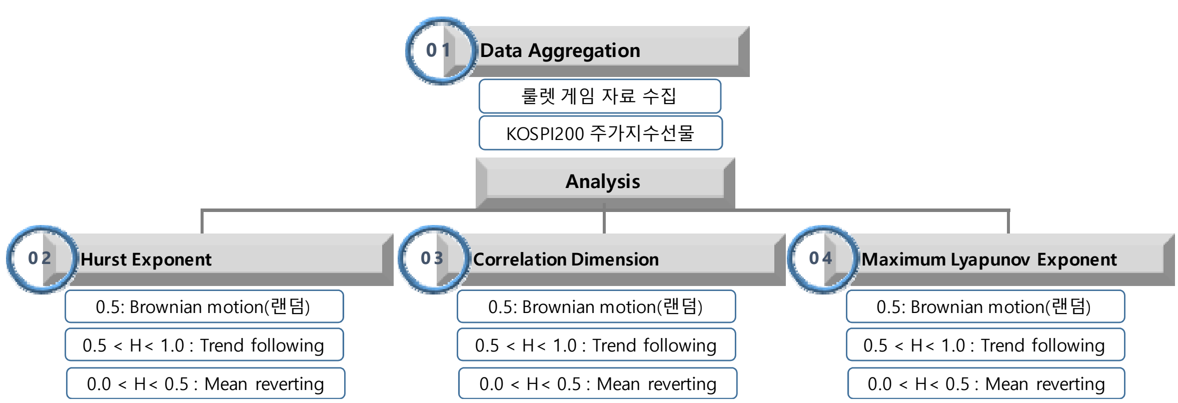
### 4.1. 허스트 지수(Hurst exponent) 분석

룰렛 게임 결과 나온 구슬의 착지지점 번호판인 R 계열 시계열과 KOSPI200 주가지수 선물의 K 계열 시계열에 대해 분석 자료의 수를 점점 증가시켜 가며 허스트 지수를 측정한 결과를 그래프로 출력했다. 그래프에서 알 수 있듯이 분석 자료의 개수가 적을 경우에 허스트 지수 값이 요동치는 것을 볼 수 있는데, 이 허스트 지수 값은 신뢰할 수 없는 값이라고 할 수 있다.

〈표 7〉 허스트 지수(Hurst Exponent) 요약표

| | R | K |
|---|---|---|
| 1 | 0.460287 | 0.884157 |
| 2 | 0.880239 | 0.543248 |

<표 7>을 살펴보면, R1의 허스트 지수 값은 약 0.46으로 0과 0.5 사이에서 형성되는 값이며, 이는 평균에 회귀하려는 시계열이라고 판단할 수 있다. 그러나, R2의 허스트 지수 값은 약 0.88로 추세강화 시계열이라



〈그림 2〉 연구 프로세스 및 방법론

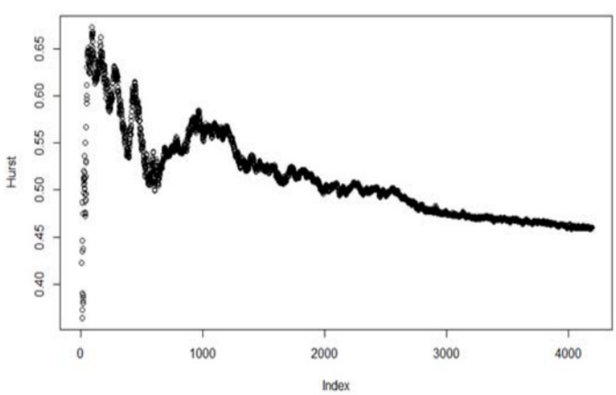
고 판단할 수 있다. 아래 <그림 3>, <그림 4>는 R 시계열의 허스트 지수를 나타내는 시각화이다.

R2 시계열에서는 원래 구슬의 착지 번호에서 -18을 하여 등호의 등락이 발생하게 전처리 하였기에, 0~17의 구슬 착지번호는 전 처리 후 음(-)의 값, 구슬 착지번호 18은 0으로 처리되면, 구슬 착지번호 19~36사이의 값은 전 처리 후 양(+)의 값을 나타내기 때문에, 룰렛 게임 결과의 시계열은 0~17이 나온 이후에 0~17이 다시 나올 확률이 높고, 19~36이 나온 이후에 19~36이 나올 확률이 높은 시계열로 추세가 있다고 판단할 수 있다.

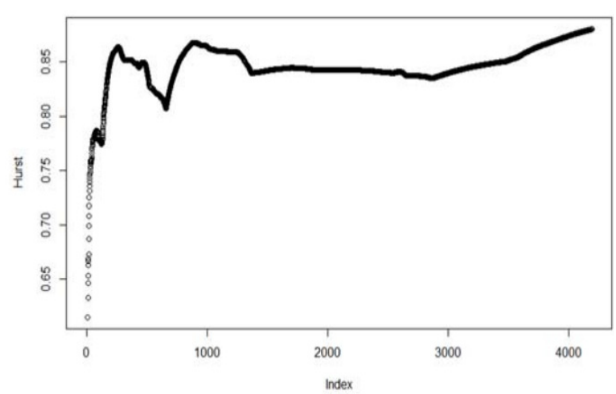
<표 7>에서 K1의 허스트 지수 값은 약 0.90으로 추세강화 시계열이고, K2 허스트 지수 값은 약 0.49로 랜덤워크를 따르는 시계열로 추세가 없는 시계열이라 고 판단 수 있다. KOSPI 200 주가지수 선물 시계열은 2000년 초반부터 글로벌 주식시장의 상승세와 동반 상승해 왔기에 종가만 판단할 경우는 허스트 지수 값이 거의 1에 근사한 값이 나왔지만, KOSPI 200 주가지수 선물의 전일과 당일의 종가(Close price) 차이 값(K2)은 예측할 수 없는 무작위성의 랜덤워크를 따른다고 할 수 있다. 아래 <그림 5>, <그림 6>은 K 시계열의 허스트 지수를 나타내는 시각화이다.

### 4.2. 상관 차원(Correlation Dimension) 분석

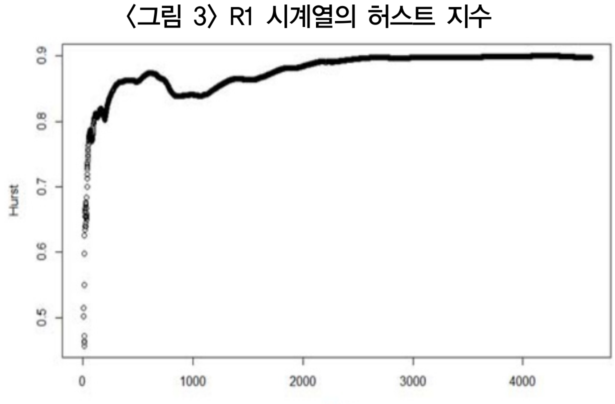
룰렛 게임 결과 나온 구슬의 착지지점 번호판인 R 계열 시계열의 R1의 상관 차원 값은 8로 도출되었고, R2의 상관 차원 값은 10으로 도출되었다. KOSPI200 주가지수 선물의 K 계열 K1의 상관 차원 값은 9, K2



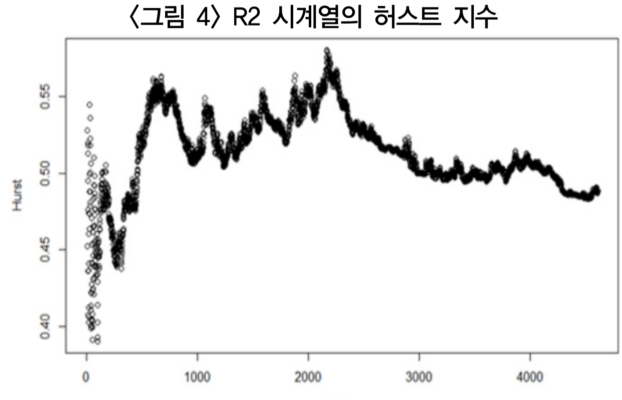

〈그림 3〉 R1 시계열의 허스트 지수




〈그림 4〉 R2 시계열의 허스트 지수




〈그림 5〉 K1 시계열의 허스트 지수




〈그림 6〉 K2 시계열의 허스트 지수

의 상관 차원 값은 12로 도출되었다. 상관(Correlation) 값이 8이라는 것은 8개의 변수가 R1 시계열에 영향을 준다는 의미이며, R1 시계열은 8개의 변수로 설명할 수 있는 시계열이라고 판단할 수 있다. 각 시계열에 대한 결과 값은 <표 8>에 몇 개의 변수로 설명력이 가능한지, 즉 Long-term memory의 소실이 발생하는 변수의 개수라고 할 수 있다.

〈표 8〉 상관 차원(Correlation Dimension) 요약표

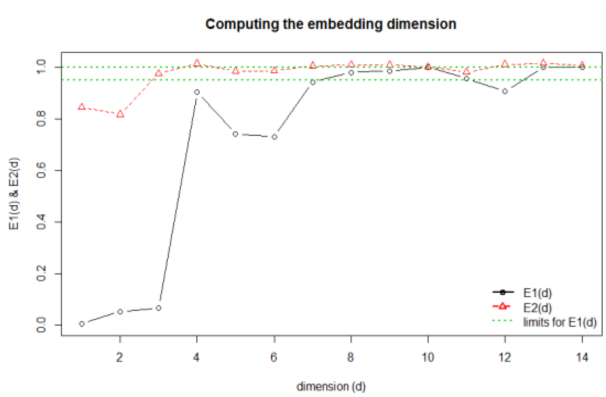
| | R | K |
|---|---|---|
| 1 | 8 | 9 |
| 2 | 10 | 11 |

아래 그림은 상관차원을 분석한 그래프이다. 이 그래프는 분석하고자 하는 비선형 데이터가 몇 차원에 존재하는지를 정량적으로 분석하는 방법이며, 값이 상승하다가 수평으로 수렴하는 그래프가 나올수록 성능이 좋다고 판단할 수 있다. <표 8>의 상관차원 요약표와 같이 <그림 7>은 값 8, <그림 8>은 값 10, <그림 9>은 값 9, <그림 10>은 값 11에서 그래프가 수렴하는 것을 볼 수 있다.

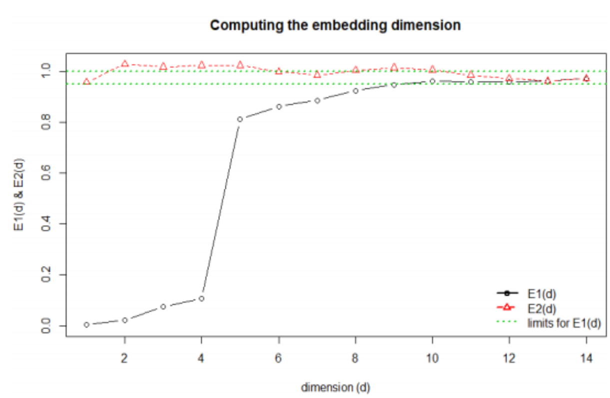
다시 말하자면, K1의 상관차원 값이 9라는 것은 룰렛 게임을 9번 실행한 결과에 영향을 받아 10번째부터 프랙탈 현상이 나타나는 결정론적 구조라고 정리할 수 있다

### 4.3. 최대 리아푸노프 지수(Maximum Lyapunov Exponent) 분석

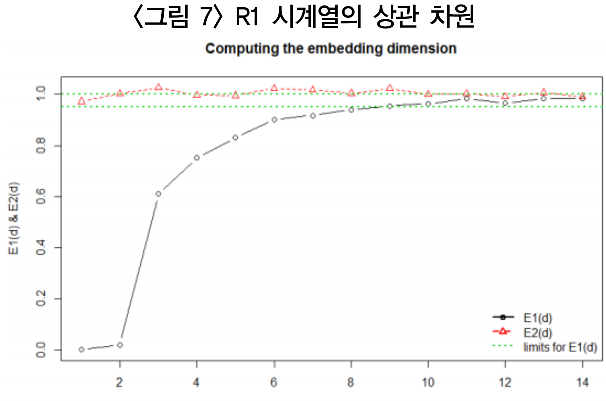
최대 리아푸노프 지수(Maximum Lyapunov exponent)



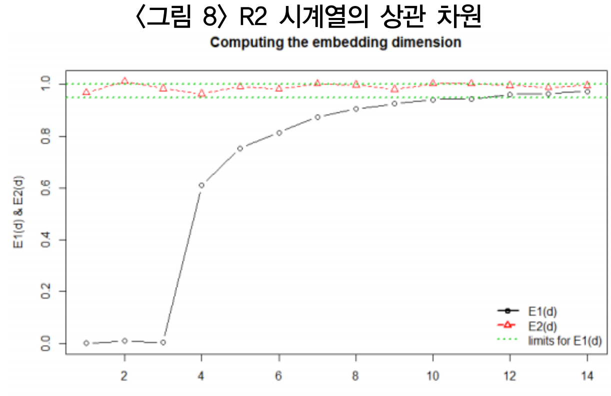

〈그림 7〉 R1 시계열의 상관 차원




〈그림 8〉 R2 시계열의 상관 차원




〈그림 9〉 K1 시계열의 상관 차원




〈그림 10〉 K2 시계열의 상관 차원

가 커질수록 해당 시계열의 추세가 오래 기억되지 않는다는 것을 의미한다. 결국, 과거의 정보로 미래를 예측하기 어렵다는 의미로 해석 된다. R1과 K1 시계열의 최대 리아푸노프 지수는 거의 0에 근사한 작은 값으로 도출 되었으며, 이러한 결과는 정보의 소실량이 거의 없어 예측 가능한 데이터임을 의미한다. R2와 K2 시계열의 최대 리아푸노프 지수는 약간 양의 값을 가지며 이는 해당 시계열이 chaotic하다는 것을 의미한다. 아래 <표 9>는 R과 K의 최대 리아푸노프 지수 값을 나타낸다.

〈표 9〉 최대 리아푸노프 지수(Maximum Lyapunov exponent) 요약표

| | R | K |
|---|---|---|
| 1 | 0 | 0 |
| 2 | 0.033328 | 0.0029842 |

## 5. 결론 및 시사점

### 5.1. 연구 결과

본 연구에서는 카오스이론의 시계열 빅데이터 적용 가능성을 룰렛 게임 결과에서 나온 구슬의 착지지점 번호 데이터와 KOSPI200 주가지수 선물 데이터를 대상으로 확인하였다. 룰렛 게임의 시계열데이터는 R, KOSPI200 주가지수의 시계열 데이터는 K로 명명하여 분석을 수행하였다. 이때 분석에 사용한 R1과 K1은 각 데이터의 기본적인 성질이 그대로 나타나 있는 데이터이며, R2는 K1의 데이터와 K2는 R1의 데이터와 유사속성을 가질 수 있도록 변화시킨 데이터이다.

허스트 지수분석 결과, R1 시계열은 평균에 회귀하려는 시계열이었으며 R2는 추세가 강한 시계열로 분석되었다. K1 시계열의 경우에는 추세가 강한 시계열이었으며, K2 시계열은 랜덤워크를 따르는 시계열로 추세가 없는 시계열인것으로 확인되었다. 따라서 룰렛의 R2와 주가지수 선물 K1은 추세가 있는 예측 가능한 데이터라는 것을 알 수 있었다. 이는 시계열 빅데이터를 분석할 시, 카오스이론을 적용할 경우 비선형적인 데이터에서도 선형의 결과값을 도출 할 수 있는 가능성을 확인한 것이라고 생각된다.

상관차원분석을 통해서는 분석하고자 하는 비선형 데이터가 몇 차원에 존재하는지 정량적으로 분석할 수 있었다. 마지막으로 최대 리아푸노프 지수분석을 함으로써 R1과 K1시계열 데이터는 정보 소실량이 거의 없는 예측가능한 데이터로 확인하였으며, R2와 K2의 결과는 해당 시계열 데이터가 Chaotic하다는 결과로 확인되었다.

모든 분석 결과를 종합 할 때, 룰렛 게임에서 발생한 구슬의 착지지점 번호 데이터에서 추세를 확인할 수 있었다. 룰렛 게임 결과 나온 구슬 착지지점 번호 1~18과 19~36이 군집되어 등장하는 시계열이면서 일정부분 예측 가능한 데이터라는 것을 검증할 수 있었기 때문이다. 허스트 지수의 결과값을 통해 R1과 K2가 유사하고 R2와 K1데이터가 유사한것으로 확인되었기 때문에, 카오스이론의 시계열 빅데이터에 대한 적용가능성을 증명한것으로 생각된다.

### 5.2. 시사점 및 한계점

본 연구에서는 카오스 이론을 시계열 빅데이터에 적용하여 비 선형적인 특성을 갖는 데이터의 패턴을 도출한 것에 의의가 있다. 서론에서 언급한 것과 같이 룰렛 게임 데이터는 일반적으로 분석대상이 되기 어려운 마이너한 주제이며, KOSPI200 주가지수는 주된 분석대상으로 활용되는 메이저한 주제이다. 두 가지의 관련성 없는 데이터를 통해 비선형 시계열 빅데이

터 분석 시 카오스이론의 활용가능성을 확인하였다. 이러한 연구를 바탕으로 산업에서 생산된 비선형 데이터 연구 진행 시, 빅데이터에 내재된 불확실성 판별 및 신뢰성 검증을 기대하고, 향후 경영 목적에 맞는 예측모델을 고도화하여 빠른 의사결정을 지원할 수 있다고 생각한다.

향후 연구에서는 다양한 디노이징 필터를 적용하여 시계열 데이터의 노이즈를 제거하고, 예측 모델이 편향적으로 학습되는 것을 보완한다면 더 유용한 시사점을 제시할 수 있을 것으로 생각된다.

## 〈참고문헌〉


### [국내 문헌]

1. Run-Qing Liu, Young-Chan Lee, Hong-Lei Mu (2018). 군집분석과 연관규칙을 활용한 고객 분류 및 장바구니 분석: 소매 유통 빅데이터를 중심으로. **지식경영연구, 19**(4), 59-76.
2. 강준, 김지우, 이현준, 오경주 (2017). 부동산 경매 낙찰가율 시계열의 Chaos 분석. **한국데이터정보과학회지, 28**(2), 371-381.
3. 김미형 (2010). 재무제표분석을 이용한 장부시장가비율이 낮은 기업의 주가수익률 예측력 분석. **국제회계연구, 30**(1), 1-18.
4. 김진솔, 신동훈, 김희웅 (2021). 비정형 빅데이터를 이용한 COVID-19 주요 이슈 분석. **지식경영연구, 22**(2), 145-165.
5. 박거준, 김용갑, 황근창 (2017). 불확실성을 고려한 퍼지 클러스터링 기반 퍼지뉴럴네트워크 설계. **한국인터넷방송통신학회 논문지, 17**(1), 173-181.
6. 박대규, 조원철 (2003). 카오스 특성을 갖는 일유출량 자료의 비선형성 예측. **대한토목학회논문집, 23**(6B), 479-487.
7. 백웅기 (1997). 업종별 Chaos (카오스) 주가지수의 및 Chaos (카오스) 검정 및 비선형예측. **재무관리연구, 14**(1), 171-205.
8. 유다솜, 윤승렬, 조성진, 조현재 (2021). 시계열 분석을 통한 일별 매출 예측 모델 개발. **대한산업공학회 춘계공동학술대회 논문집**, 5872-5872.
9. 이일균 (1998). 자본시장의 가격형성 메카니즘. **한국증권학회지, 23**(1), 1-60.
10. 이주민, 방정혜 (2020). 화장품 회사의 빅데이터분석을 통한 브랜드컨셉 개발 사례분석. **지식경영연구, 21**(3), 215-228.
11. 장경천, 김현식 (2002). Chaos 이론을 이용한 증권시장 특성에 관한연구. **한국증권학회지, 25**(1), 263-302.
12. 전해정 (2020). 부동산 뉴스와 아파트 매매가격과 거래량 간의 관계에 대한 빅데이터 시계열 분석. **부동산법학, 24**(2), 53-69.
13. 최강수, 경민수, 김수전, 김형수 (2009). BDS 통계와 DVS 알고리즘을 이용한 수문시계열의 비선형성 분석. **대한토목학회논문집 B, 29**(2B), 163-171.
14. 최봉, 윤종진, 엄태휘 (2019). 서울시 공공빅데이터 활성화 방안 연구. **지식경영연구, 20**(3), 73-89.
15. 최영일 (2001). **카오스이론과 비선형성에 관한 서베이 연구**. 석사학위논문, 연세대학교 대학원, 서울.

### [국외 문헌]

16. Arlinghaus, S. L. (1985). Fractals take a central place. ***Geografiska Annaler, 67B***, 83-88.
17. Carbone, A., Castelli, G., & Stanley, H. E. (2004). Time-dependent Hurst exponent in financial time series. ***Physica A: Statistical Mechanics and its Applications, 344***(1-2), 267-271.
18. Ethier, S. N. (1982). Testing for favorable numbers on a roulette wheel. ***Journal of the American Statistical Association, 77***(379), 660-665.
19. Gut, A., & Holst, L. (1984). On the waiting time in a generalized roulette game. ***Statistics & Probability Letters, 2***(4), 229-239.
20. Kantz, H. (1994). A robust method to estimate the maximal Lyapunov exponent of a time series. ***Physics Letter A, 185***(1), 77-87.
21. Salas, J. D., Delleur, J. W., Yevjevich, V., & Lane, W. L. (1980). ***Applied modelling of hydrologic time series***. Water Resources Publications, Littleton, Colorado.
22. Small, M., & Tse, C. K. (2012). Predicting the outcome of roulette. ***Chaos, 22***(3), 1-10.
23. Thietart, R. A., & Forgues, B. (1995). Chaos theory and organization. ***Organization Science, 6***, 19-31.

## 저 자 소 개

### 이 희 철 (HeeChul Lee)

연세대학교 공학대학원 투자정보공학 박사과정에 재학 중에 있다. 현대증권과 한국투자증권에서 파생상품을 운용 했으며, 홍콩 CFSG 글로벌 마켓의 헤드 트레이더를 역임 하였다. 현재는 벤처기업 VIX9의 대표이다. 주요 관심분야는 메타버스를 접목한 트레이딩과 겜블링, 카오스 이론, 옵션 변동성 매매전략 등이다.

### 김 홍 곤 (HongGon Kim)

현재 연세대학교 공학대학원 겸임교수로 재직 중이다. Allianz Global Investors에서 글로벌 헷지 펀드와 해외 주식운용을 총괄하는 Global Head를 역임하였고 현재 KB 자산운용의 퀀트 운용본부장으로 재직 중이다. 2020년, 2019년 연속 Hong Kong의 Asia Asset Management로부터 CIO of the year in Korea 상을 수상하였고, 2018년에는 Best of the Best Equity Fund Manger로 선정되었다. 주요 관심분야는 지식경영시스템, 지식공유, A.I. in Finance, 블록체인 경제 생태계 등이다.

### 김 희 웅 (Hee-Woong Kim)

National University of Singapore 정보시스템학과에서 근무한 후, 현재 연세대학교 정보대학원 교수로 재직 중이다. 주요 연구분야는 디지털 비즈니스, 정보시스템 관리 및 활용 등이다. 관련 연구들은 MIS Quarterly, Information Systems Research, Journal of Management Information Systems 등에 50여 편의 논문이 게재되었다. JAIS, IEEE TEM의 편집위원으로 활동했고, KrAIS 회장을 역임했다

## 〈 Abstract 〉

# Analysis of Intrinsic Patterns of Time Series Based on Chaos Theory: Focusing on Roulette and KOSPI200 Index Future

HeeChul Lee*, HongGon Kim**, Hee-Woong Kim***

As a large amount of data is produced in each industry, a number of time series pattern prediction studies are being conducted to make quick business decisions. However, there is a limit to predicting specific patterns in nonlinear time series data due to the uncertainty inherent in the data, and there are difficulties in making strategic decisions in corporate management. In addition, in recent decades, various studies have been conducted on data such as demand/supply and financial markets that are suitable for industrial purposes to predict time series data of irregular random walk models, but predict specific rules and achieve sustainable corporate objectives There are difficulties. In this study, the prediction results were compared and analyzed using the Chaos analysis method for roulette data and financial market data, and meaningful results were derived. And, this study confirmed that chaos analysis is useful for finding a new method in analyzing time series data. By comparing and analyzing the characteristics of roulette games with the time series of Korean stock index future, it was derived that predictive power can be improved if the trend is confirmed, and it is meaningful in determining whether nonlinear time series data with high uncertainty have a specific pattern.

Key Words: Big Data, Time Series Analysis, Non-linear Data, Chaos Theory, Hurst Exponent, Maximum Lyapunov Exponent, Correlation Dimension

* Yonsei University, Graduate School of Engineering
** Yonsei University, Graduate School of Engineering
*** Yonsei University, Graduate School of Information