

Sinusoid 패턴 인식을 위한 측도로서의 허스트 지수*

한양대학교 수학과 차경준
kjcha@hanyang.ac.kr

한양대학교 수학과 황선호
woogaoga@lycos.co.kr

본 연구에서 카오스 모형을 직접적으로 검정하기 위한 표준적인 기법 중의 하나인 R/S 분석(rescaled range statistical analysis)과 허스트 지수를 'sinusoid' 패턴 평가하는 데 적용하였다. 이는 다소 잡음이 섞여 있으면서 동시에 준주기 성향을 갖는 시계열자료에 대해서 허스트 지수가 이를 간접적으로 평가 할 수 있는 측도(measure)로 활용될 수 있음을 논하였다.

주제어: 허스트 지수, R/S 분석, 카오스 방법

0. 서론

정현파(sine waves)는 가장 기본적인 파형이라 할 수 있는데, 이는 조화로운 사인함수의 수학적 특성에 기인한다. 일찍이 사인함수는 1596년에 레티쿠스(Rheticus, 1514~1574)에서 기원한 삼각법과 관련된 논문인 "Opus Palatinum de Triangulis"가 출판되면서 본격적인 연구가 시작되었다. 이 논문에는 여섯 개의 삼각함수(sine, cosine, tangent, secant, cosecant, cotangent)가 골고루 사용되었는데, 삼각함수표도 함께 제시되어 있다([3], [5]).

사인 곡선(sine curve; sinusoid)의 응용은 정현파로 일반화되는데, 이는 다음과 같이 사인함수의 형태로 주어진다.

$$f(x) = a \sin(wx + c)$$

여기서 a 는 진폭(amplitude)이고 w 는 주기(period) $2\pi/w$ 와 관계된 주파수(frequency)이며 c 는 위상(phase)이다. 실질적으로 공학 분야에 응용되는 주 개념은 주파수와 주기로서 일반적으로 반복되는 신호로서의 주파수는 헤르츠(Hz)로 측정된다. Hz는 일

* 이 논문은 2002년 한양대학교 일반연구비 지원으로 연구되었음.

초에 반복되는 신호 파형의 수와 같으며, 반복되는 신호는 주기도 가지는데, 이것은 한 사이클을 완성하는데 걸리는 시간이다. 또한, 위상은 정현파에서 설명이 잘 되는데, 정현파는 360도의 순환파로서 한 사이클의 정현파는 주기가 얼마나 지나갔나를 설명하기 위해 정현파를 위상 각도로 표시하는데 활용된다. 실 예로서 전압(voltage)의 진폭은 통상 접지 또는 0볼트를 기준으로 측정되는 신호의 최대 전압을 나타내고, 벽에 붙어 있는 콘센트에서 나오는 전압이 정현파와 같이 변하며, 또한 신호발생기의 오실레이터(oscillator) 회로에서 나오는 테스트 신호도 정현파이다.

세상 모든 것이 주기적이라는 철학적 논리가 아니더라도, 우리가 접하는 RF(Radio Frequency)는 주기적 변화의 상징 그 자체이다. 주파수를 가지고 있는 어떤 주기적 성분은, 그 주파수가 2배, 3배, 4배쯤 되는 지점에서 뭔가 유사한 특성을 가지게 된다. 그리고 그 유사한 특성은 주파수의 배수가 높아질수록 약해진다.

주역의 음양오행설(陰陽五行說) 역시 주기적 변화의 상징으로서 주식투자에게까지 적용되어 우리의 삶에 많은 영향을 끼치고 있다. 실제로, 음양오행설은 12간지를 기준하면 12년 주기가 되고, 60갑자를 기준하면 60년 주기가 된다는 것으로 해석할 수 있다.

경제학에서의 경기변동 혹은 경제순환 이론에 의하면 불황은 나쁘고 호황은 좋은 것이 아니라 호황과 불황의 순환은 경제현상에서 필연적이라고 본다. 이런 경제현상과 관련하여 설명되는 순환에는 2~4년의 Kitchin 사이클, 7~11년의 Juglar 사이클, 15~20년의 Kuznets 사이클, 40~60년의 Kondratieff 사이클 등이 있다. 그러나 이러한 경제순환에 대한 반박도 여러 학자에 의해 제기되었다. 즉, 이러한 순환이 존재하는 것이 아니라 단순히 통계적 가공에 불과하다는 주장이다. 그들은 의미 없이 불규칙(random)하게 발생하는 자료에서도 순환현상을 만들어 낼 수 있는 것을 증명하였다. 주기에 대한 계속되는 논쟁은 그만큼 이 문제가 어렵다는 뜻이기도 하다. 이러한 순환 해석을 위해 AR(Auto-Regressive), MA(Moving Average), ARIMA(AR+I+MA)와 스펙트럼(spectrum)분석 등이 많이 이용되고 있으나, 결정론적 비선형 동적시스템으로부터 생성하는 잡음과 같은 현상으로 발생하는 불규칙한 천이현상의 해석은 매우 어렵다. 이러한 현상을 푸앵카레(Poincare, 1854~1912)는 혼란스러움 또는 혼돈을 나타내는 용어로서 카오스(chaos) 개념을 도입하여 해석하였다.

특히, 1980년대 이후 시계열 내의 자기 의존성(auto-dependency)의 존재를 확인함으로써 랜덤워크(random walk) 가설의 기각을 함의하는 실증 분석 결과들이 계속 제기되고 있다. 이러한 실증 분석 중에서 주어진 자료를 가지고 자료 생성 모형(data generating process)이 chaos 또는 비선형 결정계인지 아닌지 검증할 수 있는 방법 중 하나인 R/S 분석(Rescaled range analysis)과 허스트 지수는 허스트(Hurst)가 나일강의 흐름에 관한 연구로부터 장기적 종속성(long-range dependence)이 존재함을 연구하는 과정에서 제안하였으며[6], 이러한 시계열의 장기적 종속성을 검증하기 위하여, 만델브로(Mandelbrot)는 허스트가 개발한 표준화한 재조정 범위(rescaled range)라는 통계량을 이용할 것을 제안했다[8]. 그리고 현재 이 비모수 방법인 R/S 분석과 허스

트 지수는 여러 분야에 걸쳐 시계열의 랜덤워크 가설 검정을 하는데 이용되고 있다. 본 논문에서는 R/S 통계량을 어느 한 시계열 ξ_t 와 평균 $\langle \xi \rangle$ 에 대하여 시계열의 누적편차를 다음과 같이 정의하였다.

$$X(t, \tau) = \sum_{u=1}^t (\xi_u - \langle \xi \rangle_\tau)$$

그리고 u 기의 데이터 ξ_u 와 총 데이터 양 τ 를 각각 정의함에 따라 R/S 분석을 실시하여, 삼각함수 중에서 사인 곡선([3], [5])에 대한 다양한 sinusoidal 패턴을 허스트 지수를 이용하여 추정, 관찰하였다.

1. 허스트 지수

카오스 모형의 실증 분석 방법 중에서 시계열의 생성과정이 카오스를 따르는지를 검정하기 위해서는 크게 직접적인 방법과 간접적인 방법으로 나눌 수 있는데, 직접적인 방법은 검증하고자 하는 시계열이 카오스적 특징을 갖고 있는지를 검정도구로 사용하는 것이며, 간접적인 방법은 카오스를 따르는 시계열인 경우 이를 가장 잘 예측할 수 있는 모형과 그렇지 못한 모형으로 각각 시계열에 대해서 예측한 후 예측 오차를 살펴봄으로써 검증하는 것이다.

R/S 분석과 허스트 지수는 카오스 모형을 직접적으로 검정하기 위한 표준적인 기법 중 하나이다. 이 비모수적 방법은 허스트가 나일강의 댐 프로젝트를 수행하던 중 허스트 과정(Hurst process)을 개발하였는데[6], R/S 분석은 시간의 흐름에 따라 평균 기준으로 저수지의 물의 양이 어떻게 변하는가를 측정하는 데에서 시작되었다. 즉, 매년 유입되는 강수량은 유동적이기 때문에 정확한 양을 알 수 없고, 방출량은 인위적으로 운용되고 있으므로 적절한 저수량 정책의 효과는 평균적 방출량에 대한 유입량의 누적편차(cumulative deviation)로서 측정될 수 있다. 이 측정값을 표준화하기 위하여 관측치의 표준 편차로 나누어주는데 이러한 방법을 재조정 범위(rescaled range) 분석이라고 한다.

어느 한 시계열 ξ_t 에 대한 R/S 통계량은 다음과 같이 정의된다([7], [8]).

$$X(t, \tau) = \sum_{u=1}^t (\xi_u - \langle \xi \rangle_\tau),$$

$$R(\tau) = \max_{1 \leq t \leq \tau} X(t, \tau) - \min_{1 \leq t \leq \tau} X(t, \tau),$$

$$S(\tau) = \left[\frac{1}{\tau} \sum_{t=1}^{\tau} (\xi_t - \langle \xi \rangle_\tau)^2 \right]^{\frac{1}{2}},$$

$$R/S(\tau) = R(\tau) / S(\tau)$$

일차원 랜덤 워크(random walk)에서 난수 ξ_t 들의 수열을 공간적인 증가로서 간주할 때, $\sum_{i=1}^{\tau} \xi_i$ 는 시간 τ 후의 보행자의 위치이다. 그리고 $X(t, \tau)$ 에서 ξ_t 의 기대값이 0이 아닐 때, 척도를 조정하기 위하여 시간지수 τ 에 걸친 평균 $\langle \xi \rangle_{\tau} = \frac{1}{\tau} \sum_{i=1}^{\tau} \xi_i$ 을 차감한다. $R(\tau)$ 는 τ 기간 중의 $X(t, \tau)$ 의 최대값과 최소값의 차이로서 시간지수 τ 에 대하여 보행자가 움직여간 거리를 의미하는 자기 조정 범위(self-adjusted range)이고 $R/S(\tau)$ 는 모형을 일반화하기 위하여 표준편차 $S(\tau)$ 로 나누어주었는데 이를 자기 재조정 범위(self-rescaled range)라고 한다. 이러한 재조정 범위는 시간이 증가함에 따라 증가하는 경향을 보이며, 허스트는 일반적인 형태로 다음과 같이 제시하고 있다.

$$R/S(\tau) = (a \cdot \tau)^H$$

위의 식에서 a 는 상수이고, H 를 허스트 지수(Hurst exponent)라고 하는데, 다음과 같이 식의 양변에 로그를 취하고 회귀 분석을 실시하면 허스트 지수를 추정할 수 있다.

$$\log(R/S(\tau)) = H \cdot \log(\tau) + a$$

허스트 지수는 시계열에 있어서 각 사건들이 그 이후의 사건들과 상관되어 있는지 여부를 구분할 수 있는 기준을 제공하며, 혼돈이론에서의 Lyapunov 지수와 유사한 의미를 지닌다[10]. 또한, 펠러(Feller)는 유한의 변동을 갖는 어떤 독립적인 랜덤 워크(independent random process; Poisson process)에 대한 점근적인 반응(asymptotic behavior)은 다음에 의해 주어짐을 증명하였다[4].

$$R/S(\tau) = \left(\frac{\pi}{2} \tau\right)^{\frac{1}{2}}$$

즉, 허스트 지수가 $H=0.5$ 이면 장기 기억 효과가 없는 순수한 랜덤 워크며, 현재의 사건은 미래에 영향을 미치지 않음을 의미한다. 그러나 허스트 지수가 0.5가 아니면 관측값들은 독립적이지 않으며 장기 기억을 가지게 된다. 이는 현재의 사건이 미래의 사건에 영향을 미치며, 단순히 정보의 효과가 급속히 사라지는 연속적 상관관계가 아니라 장기적으로 영향이 지속되는 것을 의미한다. 만약 허스트 지수가 $0.5 < H < 1.0$ 인 경우, 그 시계열은 랜덤 워크 보다 긴 거리에 이르기 때문에 지속적인(persistent) 성향을 의미하고 양(+)의 상관관계를 갖는다. 즉, 이는 현재의 정보가 계속해서 그 정보를 획득한 미래의 시점에서 반영되는데 더욱 장기적인 영향의 지속을 의미하는 것이다. 역으로 허스트 지수가 $0.0 < H < 0.5$ 인 경우는 랜덤워크 보다 짧은 거리에 이르므로 비지속적인(anti-persistent) 평균 회귀적 시계열(mean reverting series)을 의미하고 음(-)의 상관관계가 존재한다. 즉, 허스트 지수가 0으로 접근해 가면서 현재의 정보가 미래의 시점에서 반전될 확률이 높음을 의미한다.

2. 사인 곡선 패턴에 대한 허스트 지수 추정

R/S 분석과 허스트 지수는 시계열의 랜덤워크 가설 검정으로 τ 기간 중의 누적편차 $X(t, \tau)$ 에서 ξ_u 는 u 기의 자료로 정의하였다. 본 장에서는, 차경준 등[1]이 제시한 것을 참조하여 자료의 길이와 lag를 각각 $2^{10} = 1,024$, $2^3 = 8$ 으로 설정하여 분석하였으며, 자료는 사인함수의 기본주기인 2π 를 1,024 등분하여 시계열을 구성하였다. 그리고 분석 대상 시계열 자료는 사인 곡선 $\sin(x)$ 를 기준으로 주기를 각각 $2\pi/2$, $2\pi/5$, $2\pi/10$, $2\pi/50$ 인 함수 $\sin(2x)$, $\sin(5x)$, $\sin(10x)$, $\sin(50x)$ 등과 가우시안 잡음(Gaussian noise) $N(0, \sigma^2)$ ($\sigma = 0.01, 0.1, 0.5, 1$)을 주어 잡음효과에 대한 'sinusoid'의 영향을 살펴보았다. 그리고 부가적으로 파워 스펙트럼 분석(power spectrum analysis)을 시행함으로써 다양한 사인곡선의 주기변화 패턴을 확인하였다. 파워 스펙트럼은 신호에 포함된 단위 시간당의 에너지양을 진동수의 합으로서 측정된 것으로서 계가 주기적인지, 준 주기적인지, 카오스적인지를 나타낸다[2]. 본 연구에서는 $N=1,024$ 개의 점(x_n)들을 푸리에 변환(discrete Fourier transformation, DFT)하였고, DFT는 고속푸리에 변환(fast Fourier transformation, FFT)를 사용하였다. 그리고 최종적으로 주기가 일정치 않으며 진폭이 발산하는 함수 $x \sin(1/x)$ 의 허스트 지수를 구하여 분석하였다.

본 연구에서의 모든 실험은 통계 패키지 S-PLUS 2000을 사용하여 분석을 하였다.

먼저, 사인곡선 $\sin(x)$ 에서 주기변화를 준 $\sin(2x)$, $\sin(5x)$, $\sin(10x)$, $\sin(50x)$ 등과 가우시안 잡음(Gaussian noise) 중에서 $N(0, \sigma^2)$ ($\sigma = 0.05$)에 대한 시계열 그림은 Fig. 1과 같다. 주기가 짧아짐에 따라 잡음의 효과가 사인곡선의 고유한 형태에 큰 영향을 끼치는 것으로 나타났으며, 특히, 주기가 $\frac{2\pi}{50}$ 에서 잡음효과가 더욱 두드러진다. 그러나 대부분 주기변화가 잡음효과보다는 사인곡선의 특성들이 유지되고 있음을 보여 주고 있다.

또한, 위의 사인곡선에 대한 스펙트럼 분석을 각각 시행하였던 바(Fig. 2), 그 결과를 살펴보면, 주기변화에 따라 잡음이 없었던 사인곡선의 Log(frequency) vs. Log(spectral density) plot이 단순한 형태를 띠었으며, spectral density의 피크지점이 주기가 짧아질수록 주파수의 증가영역으로 이동하는 것을 알 수 있다. 또한, 곡선 $\sin(x)$ 의 β (power-law) 기울기가 -1.932에서 주기가 짧아짐에 따라 -1.895 ~ -1.216으로 증가되었다(Fig. 2; 왼쪽). 반면 가우시안 잡음 $N(0, \sigma^2)$ ($\sigma = 0.05$)이 포함된 경우에는 β 기울기 변화가 $\sin(5x)$ 에서 0.083으로 가장 크게 나타났으나 전체적으로 0.037 ~ 0.083 정도로 작은 변화를 보였다(Fig. 2; 오른쪽).

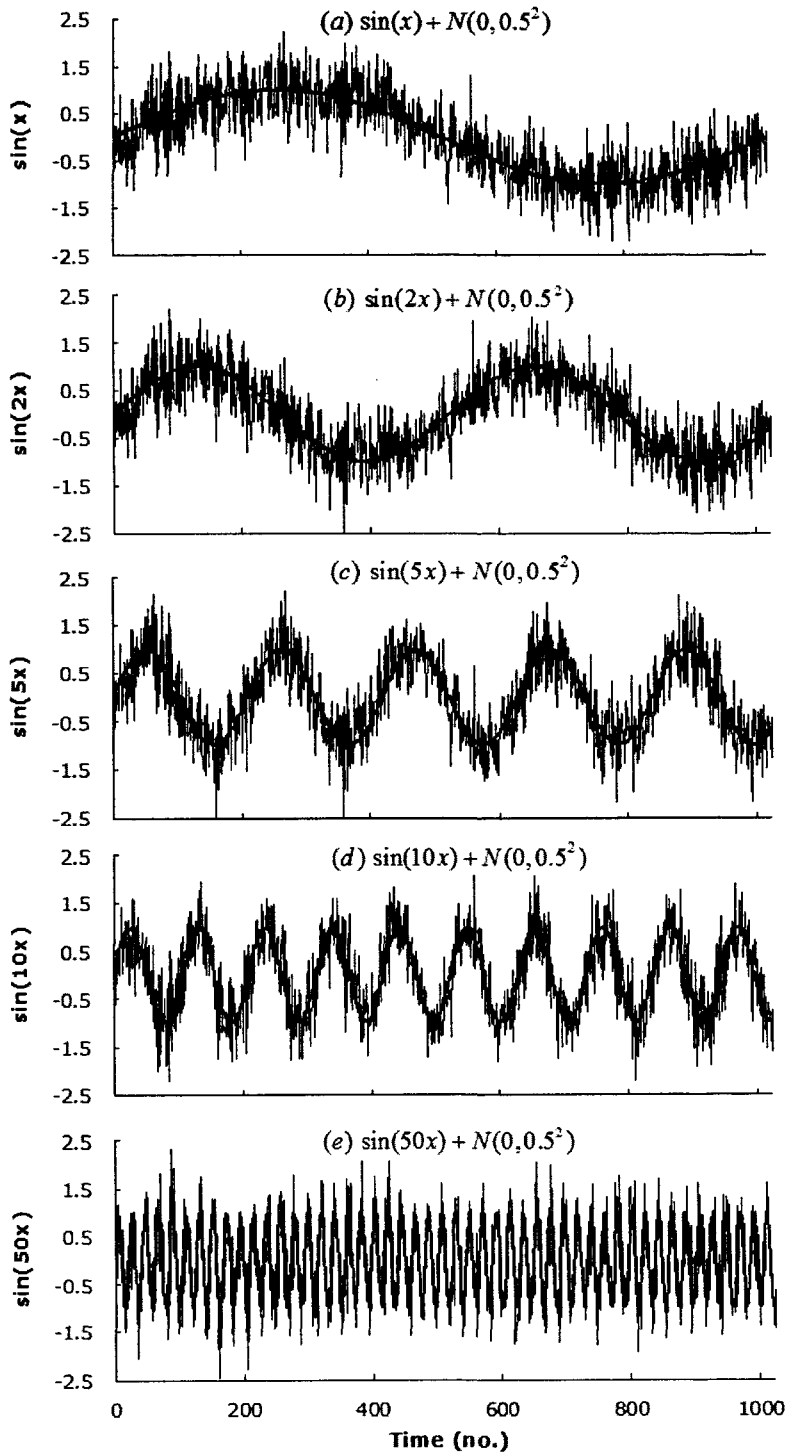


Fig. 1. The plot of sine curves add on Gaussian noise $N(0, 0.5^2)$:
(a) $\sin(x)$, (b) $\sin(2x)$, (c) $\sin(5x)$, (d) $\sin(10x)$, (e) $\sin(50x)$

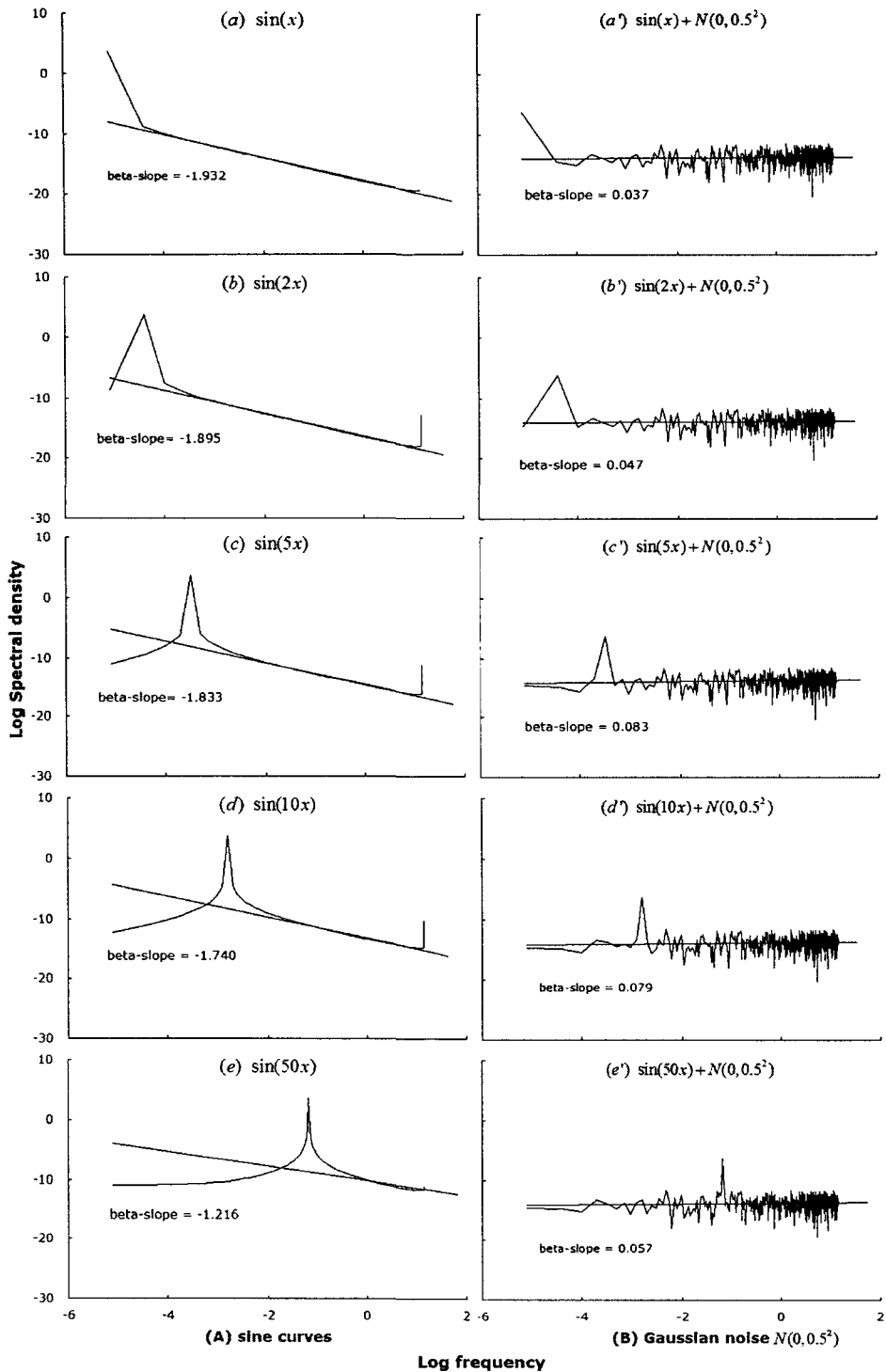


Fig. 2. The corresponding β (power-law) slope to the original sine curves (left) vs. add on Gaussian noise $N(0, 0.5^2)$ (right) in Fig. 1.

상기한 여러 가지 사인곡선에 대해서 R/S 분석을 시행한 결과(Fig. 3), 먼저 $\sin(x)$ 를 기준으로 주기를 변화시켰을 때에는 $\text{Log}(R/S)$ vs. $\text{Log}(T)$ 의 분포가 주기가 작아질수록 주름이 많아지는 형태를 띠었으며, 회귀직선의 기울기(허스트 지수; H)는 0.935 에서 시작하여 주기가 $2\pi/2, 2\pi/5, 2\pi/10, 2\pi/50$ 으로 변함에 따라 $H=0.839$ 에서 0.074 까지 점점 낮아지는 경향을 보였다(Fig. 3; 왼쪽). 또한, 잡음이 포함된 경우에서도 기준점 ($H=0.954$) 으로부터 잡음이 커질수록 다소 높은 점에

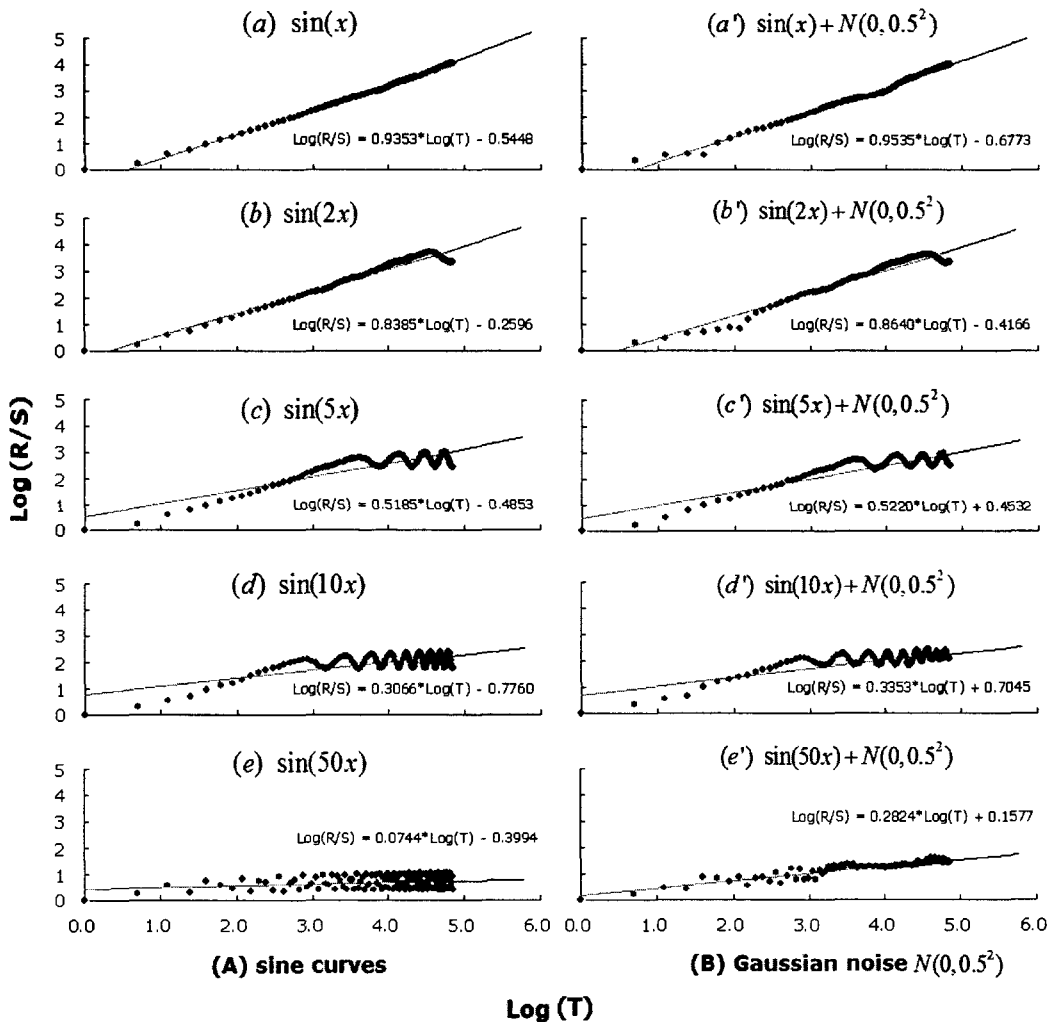


Fig. 3. The plot and regression line of rescaled range($\log(R/S)$) vs. Frequency ($\log(\tau)$) with time series data generated by sine function:

- (A) sine curves; (a) $\sin(x)$, (b) $\sin(2x)$, (c) $\sin(5x)$, (d) $\sin(10x)$, (e) $\sin(50x)$
- (B) sine curves + Gaussian noise $N(0, 0.5^2)$; (a') ~ (e')

서 출발하였으나 잡음이 없었던 경우와 비슷하게 점점 작아지는 경향을 띠었다(Fig. 3; 오른쪽). 또한, 각각의 가우시안 잡음에 대해서 대부분 원래의 사인곡선 고유의 허스트 지수에 크게 영향을 주지는 않았으나, $\sin(50x)$ 에 대해 잡음 $N(0, 1^2)$ 의 효과는 다소 큰 오차(0.2)를 보였다.

위의 사인곡선에 대한 허스트지수 H 의 신뢰성 있는 결과를 위하여 가우시안 잡음의 생성과정을 100번 반복실험을 하여 평균과 표준편차를 각각 구하여 비교하였다(Fig. 4). 그 결과 $\sin(x)$, $\sin(2x)$, $\sin(5x)$, $\sin(10x)$ 에 대해서는 잡음에 크게 영향을 받지 않으면서 일관성 있는 지수분포를 보였으나, $\sin(50x)$ 에서의 다소 큰 가우시안 잡음 $N(0, 1^2)$ 에 대해서는 표준편차가 0.056 으로서 신뢰성이 있는 허스트지수를 보여 주지 못하였다(Fig. 4).

그리고 최종적으로 주기 및 진폭이 발산하는 $x\sin(1/x)$ 의 함수에 대해서 R/S 분석을 시행하였는데, 이런 곡선 예는 주기 및 준주기 등의 개념과 지속 및 반지속 성향을 동시에 갖기 때문에 스펙트럼 분석으로 해석하는데 한계를 갖고 있기 때문이다.

곡선 $x\sin(1/x)$ 의 시계열은 변역 $[0, 2\pi]$ 에서 1,024 ($lag=2^3$) 개의 자료를 선택하였으며, 가우시안 잡음효과 역시 상기한 방법과 동일하게 100회 반복실험 하였다. 원래의 곡선 $x\sin(1/x)$ 의 허스트지수 H 는 0.825 로서 반지속성보다는 지속성향이 있는 함수로 나타났으며, 다소 큰 잡음 $N(0, 1^2)$ 를 제외하면 대부분 표준편차 내에 존재하였다.

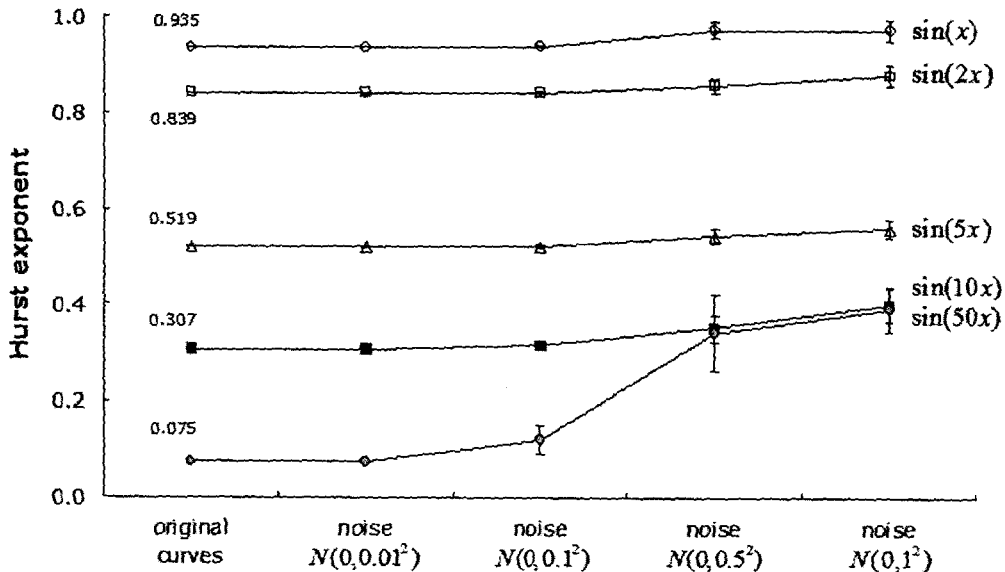


Fig. 4. The plot of Hurst exponents of sine curves $\sin(cx)$ ($c=1, 2, 5, 10, 50$) according to Gaussian noise $N(0, \sigma^2)$ ($\sigma=0.01, 0.1, 0.5, 1$).

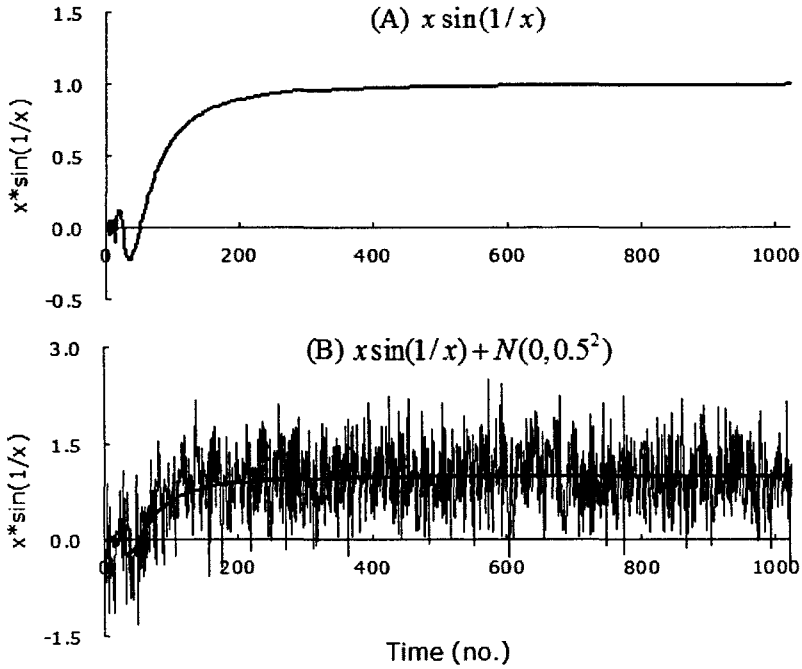


Fig. 5. The plot of (A) $x \sin(1/x)$ and (B) add on Gaussian noise $N(0, 0.5^2)$; (Time no. = 1,024 points, $x \in [0, 2\pi]$)

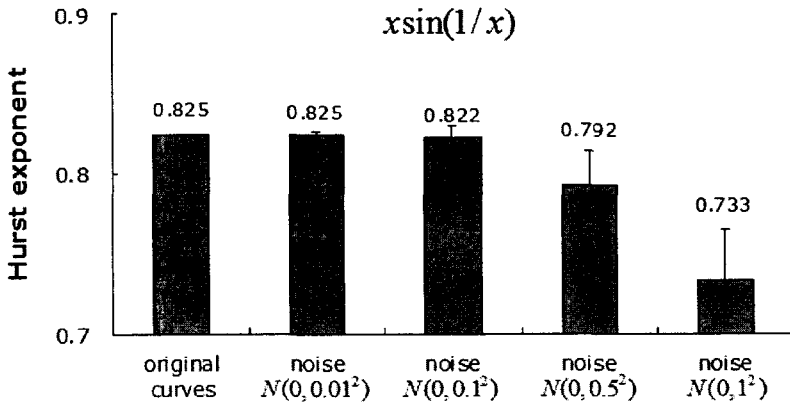


Fig. 6. The plot of Hurst exponents of $x \sin(1/x)$ according to Gaussian noise $N(0, \sigma^2)$ ($\sigma = 0.01, 0.1, 0.5, 1$).

3. 결론

R/S 분석과 허스트 지수는 시계열의 랜덤워크 가설 검정 및 자료의 패턴에서의 지속성과 반지속성을 평가하는데 활용되는 등 매우 응용범위가 넓은 분석기법이다.

본 연구에서는 자료의 길이와 *lag*를 각각 $2^{10} = 1,024$, $2^3 = 8$ 으로 설정하여 R/S 분석과 허스트 지수를 함수의 'sinusoid' 패턴을 측정하는데 적용하였다. 분석 자료는 사인함수의 기본주기인 2π 를 1,024 등분하여 사인곡선 $\sin(x)$ 의 함수값을 이용하였으며, 곡선 $\sin(x)$ 를 기준으로 주기를 증가시키면서 잡음에 영향을 받는지를 평가하여 허스트 지수의 신뢰성을 간접적으로 검증하였다. 그 결과, 가우시안 잡음(Gaussian noise)이 큰 경우와 주기가 매우 작은 경우에는 허스트 지수의 편차가 존재하였으나 대체로 오차범위 내의 값을 보였다. 그리고 준주기 성향을 갖는 동시에 국부적인 반지속성을 갖는 곡선 $x\sin(1/x)$ 에 대해서 허스트 지수를 구하였던 바, $0.5 < H = 0.825 < 1$ 로서 장기적인 영향을 받는 지속과정 시스템 신호로 판단할 수 있었다.

따라서 다소 잡음이 섞여 있으면서 동시에 준주기 성향을 갖는 시계열 자료에 대해서 스펙트럼분석으로 해석이 어려운 경우, 허스트 지수가 복합된 신호의 주기성을 측정할 수 있는 측도(measure)로 활용될 수 있음을 시사한다고 하겠다.

참고 문헌

1. 차경준 · 진정언 · 황선호, "허스트지수의 특성에 관한 연구," 한양대학교 자연과학 논문집 21(2001), 47-55.
2. Akselrod, S. · Gordon, D. · Ubel, F.A. · Shanon, D.C. · Barger, A.C. · Cohen, R.J., "Power spectrum analysis of heart rate fluctuation. Aquantitative probe of beat-to-beat cardiovascular control," *Science* 213(1981), 220-222.
3. Beyer, W.H., *CRC Standard Mathematical Tables* 28th ed., Boca Raton, FL: CRC Press, 1987, 225.
4. Feller, W., "The asymptotic distribution of the range of sums of independent random variables," *Annals of Statistics* 14(1981), 517-532.
5. Jeffrey, A., "Trigonometric Identities," §2.4 in *Handbook of Mathematical Formulas and Integrals* 2nd ed. Orlando, FL: Academic Press, 2000, 111-117.
6. Hurst, H., "Long term storage capacity of reservoirs," *Transactions of the American Society of Civil Engineers* 166(1951), 770-799.
7. Lo, A.W., "Long-term memory in stock market prices," *Econometrica* 59, 5

- (1991), 1279-1313.
8. Mandelbrot, B., "When can price be arbitrated efficiently? Alimited to the validity of the random walk and martingale models," *Review of Economics and Statistics* 53(1971), 225-236.
 9. Mandelbort, B. · J. Van Ness, "Fractional Brownian motion, fractional noise and applications," *SIAM. Review* 10(1968), 422-437.
 10. Peters, E.E., *Fractal market analysis: Applying chaos theory to investment and economics* 2nd ed., John Wiley & Sons Inc. 1994.

A Hurst Exponent as the Measure for a Sinusoid Pattern Recognition

Professor, Dept. of Mathematics, Hanyang Univ. **Kyung-Joon Cha**
MSC, Dept. of Mathematics, Hanyang Univ. **Sun-Ho Hwang**

The Rescaled range statistical analysis and Hurst exponent which are standard methods to test the chaotic model are used to examine sinusoid pattern. We notice that the Hurst exponent can be used as a measure to examine the time series data that show semi-cyclic trend with noise.

Key words : Hurst exponent, R/S analysis, chaotic method