고빈도 거래에 대한 스펙트럼 분석

조중리(아주대학교)
김용식(아주대학교)
현종석(에프앤자산평가)

초 록

본 연구는 금융시장의 고빈도 거래(High-Frequency Trading) 행태를 식별하는 하나의 spectral periodogram 방법을 제시한다. 일반적으로 거래자별 주문 전략을 분석할 수 있는 사적자료는 이용이 제한되므로, 대중적으로 접근이 가능한 자료만을 이용한 거래 식별 방법에 대해 수요가 많다. 본 연구에서 제안된 방법은 오직 대중적인 정보만을 이용하기 때문에 고빈도 거래에 대한 대용치(proxy)를 이용하는 방법에 비해 보다 직접적이다. 우리는 비정규적으로 관찰되는 시계열 자료를 분석하는데 적합한 것으로 알려진 Lomb(1976)와 Scargle(1982)의 방법을 이용하여 전통적인 계량경제 모형으로는 분석이 어려운 금융시장의 거래 자료로부터 spectral density를 추정하여 스펙트럼 분석을 하였다. 그 결과 빈도영역(frequency domain)에서 KOSPI200 지수 선물의 고빈도 거래로 여겨지는 강한 신호를 관찰하였다. 반면 우리는 한국 주식시장에서 가장 많은 비중을 차지하고 있는 한 종목에 대해서는 신뢰할 만한 고빈도 거래 신호를 발견하지 못하였다. 이러한 경험적 분석 결과는 고빈도 거래에 대한 시장의 통념과 기존의 문헌에서 간접적으로 관찰된 사실들을 잘 설명하는 것으로 보이며, 앞으로 고빈도 거래와 관련된 시장미시구조 연구 등에 활용할 수 있을 것으로 기대한다.

1 서론

전통적인 금융 모형이 상대적으로 장기적인 시각에서 거래의 균형과 같은 상황에 초점을 맞추었다면, 최근의 현대 금융 모형 연구에서는 시장의 미시구조(Market Microstructure)적인 측면에서 순간적인 거래의 불균형이나 거래의 흐름(order flow)에 많은 관심이 있다. 특히 기존에는 계량경제(Econometrics) 모형을 이용하여 거시경제의 분석이나 시장의 변동성을 추정하는 연구들이 활발히 진행되어 왔는데, 아쉽게도 많은 연구들이 사용하는 기존의 방법으로는 현대 금융시장의 특성 중 하나인 고빈도 거래(High-Frequency Trading)의 분석이 어려운 것으로 알려져 있다. 그 이유는 우선 기존 계량경제 모형에서 많이 쓰이는 시계열 모형, 즉 ARIMA와 같은 차분방정식 형태의 모형들은 기본적으로 분석 시계열 프로세스의 시구간이 등 간격 프로세스(equally-spaced process)라고 가정하기 때문이다. 거시경제 변수들의 경우 월별, 분기별 등의 간격으로 측정되기에 계량경제 모형이 거시적인 분석에는 유효하지만, 거래 구간이 매우 불규칙적인 주식시장이나 파생상품 시장에서는 그러한 기존의 방법론으로는 정확한 분석이 어렵다.

물론 근사적인 방법으로서 보간법(interpolation)을 이용하여 등 간격 자료를 가공하여 분석할 수도 있으나, 금융 거래의 경우 1초에도 수 천 건의 거래가 일어날 때가 있고 1시간에도 몇 건의 거래밖에 없을 때도 있기 때문에 보간법으로는 원래의 거래 자료가 아닌 인위적인 자료를 얻게 되므로 설득력이 매우 떨어진다. 또한 고빈도 거래 분석의 가장 큰 궁금증은 과연 그러한 고빈도 거래가 일어나는가 하는 식별(identification)의 문제인데, 시장에서 우리가 관찰할 수 있는 자료들은 일련의 시계열 거래 자료이기 때문에 거래자별 주문이 완벽하게 구분되어 있는 사적자료(proprietary data)가 아닌 이상 그러한 거래를 식별하기 어렵다. 따라서 대중적으로 접근이 가능한 자료만을 이용한 고빈도 거래 식별 방법에 대한 수요가 많다.

본 연구에서는 금융시장의 고빈도 거래(High-Frequency Trading) 행태를 식별하는 하나의 스펙트럼 분석(spectral analysis) 방법을 제안한다. 우리가 제안하는 방법은 오직 대중적인 정보만을 이용하기 때문에 고빈도 거래에 대한 대용치(proxy)를 이용하는 등의 기존 방법들에 비해 보다 직접적이며, 따라서 시장의 원래 거래 자료를 재가공 없이 온전하게 이용한다는 장점이 있다. 우리는 비정규적으로 관찰되는 시계열 자료를 분석하는 데 적합한 것으로 알려진 Lomb(1976)와 Scargle(1982)의 방법(Lomb-Scargle Fourier Transform)을 이용하여 금융시장의 거래 자료로부터 spectral density를 추정하여 스펙트럼 분석을 하였다. 우리는 선물시장과 주식시장의 대표종목을 각각 하나씩 분석하였는데, 그 결과 빈도영역(frequency domain)에서 KOSPI200 지수 선물의 고빈도 거래로 여겨지는 강한 신호를 관찰하였다. 반면 한국 주식시장에서 가장 많은 비중을 차지하고 있는 한 종목에 대해서는 신뢰할 만한 고빈도 거래 신호를 발견하지 못하였다. 이러한 경험적 분석 결과는 고빈도 거래에 대한 시장의 통념과 기존의 문헌에서 간접적으로 관찰된 사실들을 잘 설명하는 것으로 보이며, 앞으로 고빈도 거래와 관련된 시장미시구조 연구 등에 성공적으로 활용할 수 있을 것으

로 기대한다.

논문의 구조는 다음과 같다. 두 번째 장에서 우리의 분석 도구인 *고속 푸리에 변환(Fast Fourier Transform)*과 *LSFT(Lomb-Scargle Fourier Transform)*에 대해 설명하고, 세 번째 장에서 분석 데이터의 설명과 주요 결과를 제시한다. 그리고 마지막 장에서 결론을 기술한다.

2 불규칙적 자료에 대한 고속 푸리에 변환

2.1 고속 푸리에 변환(Fast Fourier Transform)과 주기도(periodogram)

시계열 분석에서 하나의 중요한 문제는 바로 주기성(periodicity)을 찾는 것이다. 일반적으로 시계열 분석에서는 어떤 변수의 경로를 ARIMA와 같은 차분 방정식으로 모형화하고 미래의 흐름을 추정하는 시간 영역(time domain)에서의 분석방법이 있다. 반면에 주기성을 탐지하는 조금 더 효과적인 대안 방법은 스펙트럼 분석(spectral analysis)을 통해 그러한 시계열 자료를 공간 영역, 즉 빈도 영역(frequency domain)에서 분석하는 것이다.

스펙트럼 분석의 주요 목적은 어떠한 시계열 자료를 빈도별로 분해(decompose)하여 어떤 주기적인 신호(periodic signal)가 존재하는지 탐구하는 것이다. 예를 들어 $x_{t} = \alpha x_{t-1} + \varepsilon_t$과 같은 시계열 거래 자료를 가정해보자. 이때 $|\alpha| < 1$이고 $\{\varepsilon_t\} \sim \mathrm{N}(0,1)$이라고 가정한다. 만일 이 확률과정에 뚜렷한 주기성이 있다면 우리의 문제는 스펙트럼 분석을 통해 그 주기성을 찾아내는 것이다. 이러한 공간분석의 대표적인 방법이 바로 푸리에 변환(Fourier Transform)이다. 푸리에 변환은 직관적으로 어떤 샘플 데이터로부터 주기신호를 빈도별로 변환하는 방법으로서, 신호처리(Signal Processing) 등의 공학 분야에서도 많이 응용되는 방법이다. 우리가 일상생활에서 많이 이용하는 MP3나 Wifi와 관련된 기술에도 쓰이는 등 실로 다양한 분야에서 이 방법을 활용하고 있다. 금융공학 분야에서는 옵션 가격결정 (Carr and Madan, 1997), Affine 프로세스 (Duffie et al., 2003) 등에서 활용되고 있다.

푸리에분석(Fourier analysis)을 통해서 설명하자면, 어떤 일반적인 시계열 프로세스는 무한개의 삼각함수, 즉 sine과 cosine 같은 함수들의 합으로 근사할 수 있다. 예를 들어 일련의 거래 가격 자료 $x(t_j)$가 $j = 1, 2, \cdots, N$ 로서 유한개 관찰되었다고 가정하자. 이때 시간 간격은 $t_{j+1} - t_j = \Delta t$로서 모든 j에 대해 등 간격이다. 그렇다면 이 시계열자료는 다음과 같은 푸리에 급수로 나타낼 수 있다.

$$x(t) = \frac{a_0}{2} + \sum_{k=1}^{\infty} \left(a_k \cos(\omega_k t) + b_k \sin(\omega_k t) \right).$$

이때 각(角)주파수 $\omega_k = 2\pi f_k$는 단위 시간당 위상각의 변화로 측정되며 호도(radian)로 표현된다. 반면 f는 우리가 실제로 푸리에 변환 시 빈도영역에서 살펴볼 단위로서 단위 시간당

진동횟수로 측정되며 Hertz(빈도의 단위)로 표현된다. 만일 t가 밀리세컨드(millisecond; ms), 즉 1천 분의 1초로 표현된다면 100Hz라는 것은 10ms 당 1번을 반복 혹은 진동함을 의미한다. 예를 들어 시계의 초침은 1 Hz로 반복한다. 나머지 a_k와 b_k는 각각 다음과 같은 푸리에 계수(Fourier coefficient)를 의미한다.

$$a_{k=}\frac{1}{\pi}\int_{-\pi}^{\pi}f(x)\cos dx,$$

$$b_k=\frac{1}{\pi}\int_{-\pi}^{\pi}f(x)\sin dx.$$

이때 오일러의 식(Euler's formula)인

$$e^{i2\pi f}=\cos(2\pi f)+i\sin(2\pi f)$$

을 이용하면 $x(t)$는 다음과 같이 간단하게 표현될 수 있다.

$$\Rightarrow x(t)=\sum_{-\infty}^{\infty}c_k e^{i\omega_k t}.$$

이때 i는 허수 단위(imaginary unit)을 나타낸다.

푸리에 급수에는 여러 가지 변형(variation)이 있는데, 그 중 대표적인 것이 바로 푸리에 변환이다. 푸리에 변환은 연속적인 실수 공간에서 연속적인 빈도 공간으로 변환하는 일대일 연산자(operator)이다. 조금 더 명확하게 기술하면 빈도 f에서 거래변수 $x(t)$의 푸리에 변환은 다음과 같은 복소수로 표현할 수 있다.

$$FT(f)=\int_{-\infty}^{\infty}x(t)e^{-i2\pi ft}dt.$$

이러한 연속적인 형태의 푸리에 변환을 실제로 응용하는 하나의 기본적인 도구가 바로 이산 푸리에 변환(Discrete Fourier Transform)이다.

$$FT(\omega)=\sum_{n=1}^{N}x(t_n)\exp(-i\omega t_n).$$

이때 $\omega=2\pi f$이다.

빈도 영역에서 빈도함수의 강도 혹은 영향력을 측정하는 하나의 측도는 스펙트럴 밀도 함수(SDF; spectral density function)이다. 스펙트럴 밀도 함수는 각각의 빈도가, 즉 어떤 주기성이, 데이터를 얼마나 잘 설명하는지를 나타내는 측도로서 시간 영역에서의 자기상관 함수(autocorrelation function)와 비슷하다. 구체적으로 SDF의 추정(estimation)인 주기도(periodogram)는 다음과 같이 정의된다.

$$P_X(\omega)=\frac{1}{N\sigma^2}|FT_X(\omega)|^2=\frac{1}{N\sigma^2}\left|\sum_{n=1}^{N}X(t_n)\exp(-i\omega t_n)\right|^2$$

이때 $\sigma^2=(N-1)^{-1}\sum_{j=1}^{N}(X_j-\overline{X})^2$ 이고 $\overline{X}=N^{-1}\sum_{j=1}^{N}X_j$ 이다. 엄밀히 말해서 주기도라는 것은 연속적인 시간영역에서 정의된 파워 스펙트럼(power spectrum)의 유한한 버전이지만 이 논문에서는 두 용어 사용에 차이를 두지 않는다.

2.2 불규칙적 자료에 대한 고속 푸리에 변환과 스펙트럼

지금까지 살펴보았던 고속 푸리에 변환을 이용하면 우리가 시간영역에서의 시계열 자료를 빈도영역에서 각 주기별로 분해하여 분석할 수 있게 된다. 그런데 이 변환의 전제로서 시간 자료가 등 간격이라는 제한이 있다. 만일 우리의 관심이 되는 금융거래 자료가 등 간격으로 기록된다면, 다시 말해 거래가 일정한 단위시간마다 기록된다면 기존의 이산 푸리에 변환을 이용하여 주기성을 관찰할 수도 있을 것이다. 하지만 주식시장이나 선물시장과 같은 역동적인 금융시장에서는 거래의 발생이나 기록시간이 매우 불규칙적이기 때문에 이산 푸리에 변환을 바로 적용할 수 없다. 하나의 가능한 방법은 원 데이터(raw data)가 등 간격이 되도록 보간(interpolation)을 하는 등 적절히 가공하여 변환할 수도 있으나, 이렇게 얻어진 데이터는 실제 거래가 일어나지 않은 시점에도 인위적으로 거래가 발생한 것처럼 '거짓' 데이터를 생성하게 되므로 옳은 방법인지에 대한 논란이 있을 수 있다. 한편 이러한 보간을 이용한 방법은 실제로 성과가 좋지 않다는 것이 알려져 있다.

이에 대해 Lomb (1976)가 최소자승(least-squares)을 이용한 방법을 개발했고, 이것을 Scargle(1982)이 정교하게 다듬어 불규칙적 자료를 다룰 수 있는 더욱 일반적인 형태의 파워 스펙트럼을 다음과 같이 제시했다.

$$P_X(\omega_k) = \frac{1}{2\sigma^2}\left\{ \frac{\left[\sum_{j=1}^{N}(X_j - \overline{X})\cos\omega_k(t_k-\tau)\right]^2}{\sum_{j=1}^{N}\cos^2\omega_k(t_j-\tau)} + \frac{\left[\sum_{j=1}^{N}(X_j - \overline{X})\sin\omega_k(t_j-\tau)\right]^2}{\sum_{j=1}^{N}\sin^2\omega_k(t_j-\tau)} \right\}$$

이때

$$\tau(\omega_k) = \frac{1}{2\omega_k} arctan\left(\frac{\sum_{j=1}^{N}\sin(2\omega_k t_{j)}}{\sum_{j=1}^{N}\cos(2\omega_k t_{j)}} \right)$$

이고 $f_k = \omega_k / 2\pi$ 는 zero frequency 근처부터 Nyquist frequency $f_{Nyquist} := \frac{1}{2\Delta}$ 까지 걸치는 빈도를 나타낸다. $T := \max_i(t_i) - \min_i(t_i)$는 전체 자료의 시간 길이이고 $\Delta := \frac{N}{T}$ 는 연속된 자료들의 샘플링구간(sampling interval)이다. 이 스펙트럼은 불규칙적으로 기록된 자료에 대한 스펙트럼의 근사로서 기존 주기도의 보다 일반화된 형태이며, $\Delta t \to 0$, $N \to \infty$로 갈수록 등 간격 샘플링의 주기도로 수렴한다. 보다 직관적으로 이 스펙트럼은 다음과 같이 데이터를 주기함수에 최소자승법으로 적합시킨 것이다.

$$X_f(t) = A\cos\omega t + B\sin\omega t.$$

2.3 거짓 확률(false-alarm probability)

푸리에 변환을 통해 얻어진 빈도영역의 주기신호는 사실 추정된 값이므로 스펙트럼 신호는 잡음(noise)으로 이루어진 변동(fluctuation)의해서도 발생할 수 있다 (Scargle, 1982). 따라서 각 빈도마다 나타나는 스펙트럼이 얼마나 신뢰할 만한 것인지에 대한 통계적 검증은 필수적인 단계이다.

Scargle(1982)의 스펙트럼은 지수 확률 분포를 따른다는 특성이 있는데, 하나의 스펙트럼이 어떤 양수의 z와 $z+dz$사이에 있을 확률은

$$\Pr(z < Z < z+dz) = \exp(-z)dz$$

가 되며, 0에서부터 z까지의 누적확률분포는

$$\Pr\{Z < z\} = \int_0^z \exp(-z')dz' = 1-\exp(-z)$$

가 된다. 조금 더 유용한 통계량으로는 스펙트럼의 값이 z보다 클 확률, 즉 푸리에 변환을 통해서 얻어진 스펙트럼 신호가 거짓일 확률(false-alarm probability)은

$$\Pr\{Z > z\} = \exp(-z)$$

가 된다. 이 거짓확률은 사실상 통계학의 *p-value*와 같은 역할을 하는데, 낮은 거짓확률은 그 스펙트럼 신호가 가우시안 확률 값(Gaussian random value)이라는 귀무가설을 기각시킬 수 있음을 의미한다.

3 고빈도 거래 자료의 스펙트럼 분석

3.1 데이터 설명

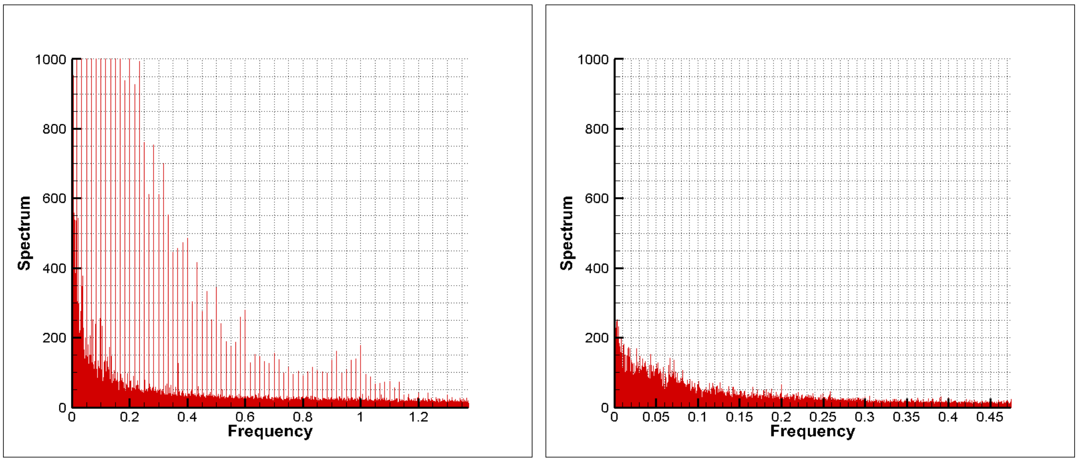
이 연구의 목적은 파생상품 시장이나 주식시장과 같은 금융시장에서 고빈도 거래의 신호를 식별하는 것이다. 따라서 우리는 각 시장을 대표한다고 여겨지는 대표상품을 다음과 같이 선정하였다. 파생상품 시장에 대해서는 KOSPI200 지수 선물의 최근월물(상품코드 : KR4101G30006)을, 주식시장에 대해서는 삼성전자 주식(상품코드 : KR7005930003)을 분석한다. 분석 기간은 2012년 1월 2일부터 2012년 3월 8일 선물만기일까지이며, 코스콤의 KRDS에서 제공하는 체결데이터를 원 자료로 이용한다. 이 연구에서 푸리에 변환 시 이용하는 입력 데이터는 밀리세컨드 단위로 기록된 시간과 거래가 체결된 가격이 된다.

일반적으로 선물시장은 유동성이 세계적으로 풍부하며 거래량이 매우 많다. 여러 선물상품 중에서 시장을 대표할 수 있는 KOSPI200 지수 선물의 경우 알고리즘 트레이딩(algorithmic trading)이나 고빈도 거래가 활발한 것으로 알려져 있는 반면, 주식시장의 경우 시가 총액이 가장 많은 삼성전자의 경우라도 상대적으로 미미한 것으로 짐작진다.

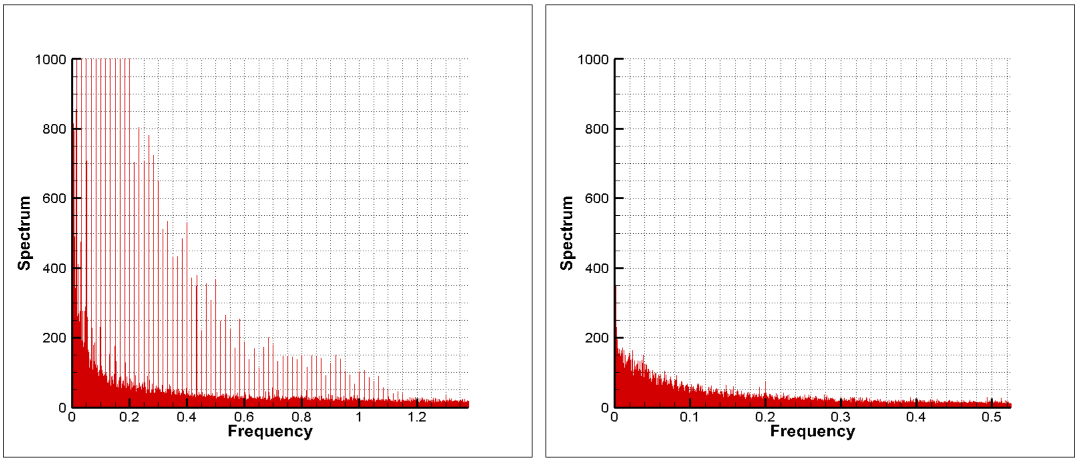
3.2 분석결과

본 연구에서 푸리에 변환의 결과는 2차원 플롯(plot)으로 표현하는데, 가로축은 빈도를,

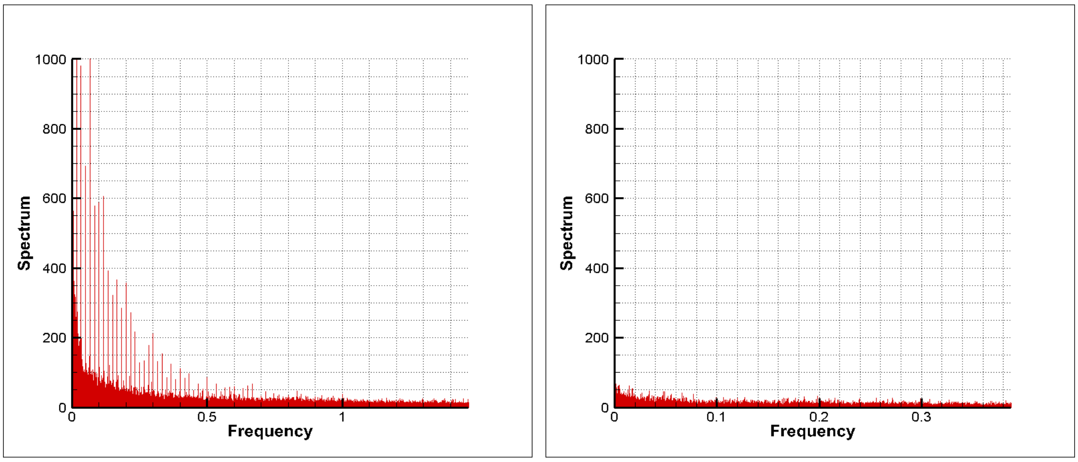
세로축은 스펙트럼을 나타낸다. 계산은 일별(daily), 주별(weekly), 월별(monthly)로 실시하였으며, 월별 분석을 비교 분석한 결과는 다음과 같다.



<그림 1> 2012년 1월 선물시장과 주식시장의 비교(전체 영역)

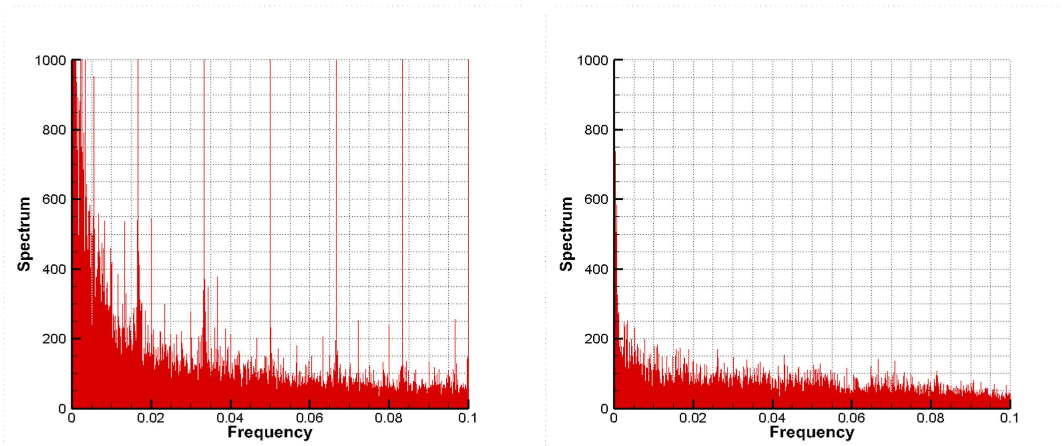


<그림 2> 2012년 2월 선물시장과 주식시장의 비교(전체 영역)

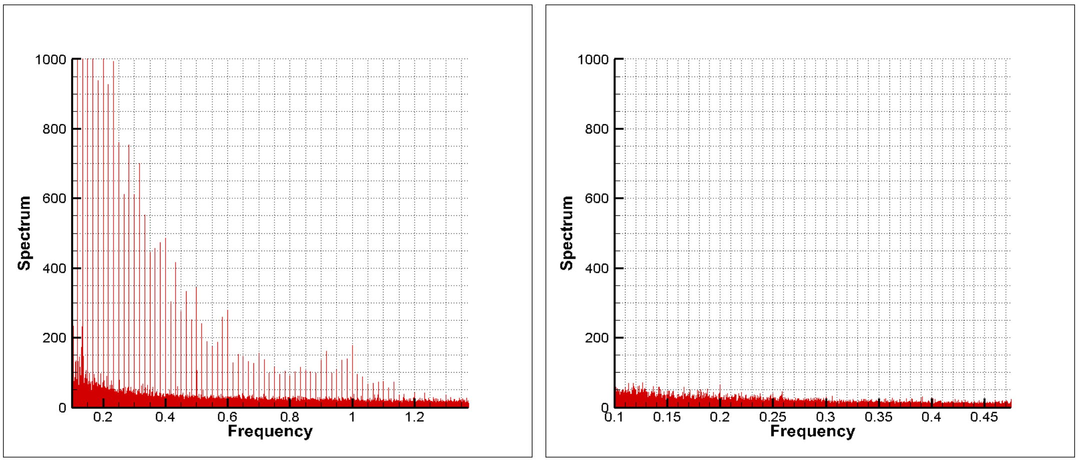


<그림 3> 2012년 3월 선물시장과 주식시장의 비교(전체 영역)

<그림 1>, <그림 2>, <그림 3>을 보면 우선 전반적으로 선물시장의 경우 특정 빈도마다 강한 신호가 시각적으로 확인된다. 이것은 그 빈도별로 알고리즘 트레이딩이 일어나고 있다고 해석할 수 있다. 기본적으로 관찰하는 빈도영역은 zero frequency 다음부터 최대 Nyquist frequency까지 인데, 이는 자료의 개수와 총 관찰 구간에 따라 달라지므로 전체 빈도 영역의 구간이 달라질 수 있다. 한편 0.1 Hz 라는 것은 초당 0.1회, 즉 10초에 1회 반복된다는 것이고, 거래자가 직접 손으로 주문하는 경우 통상 10초의 시간이 필요하다고 가정한다면 전체 빈도 영역을 0.1 Hz를 기준으로 저빈도 영역과 고빈도 영역으로 나누어 두 시장을 비교해 볼 수 있다.

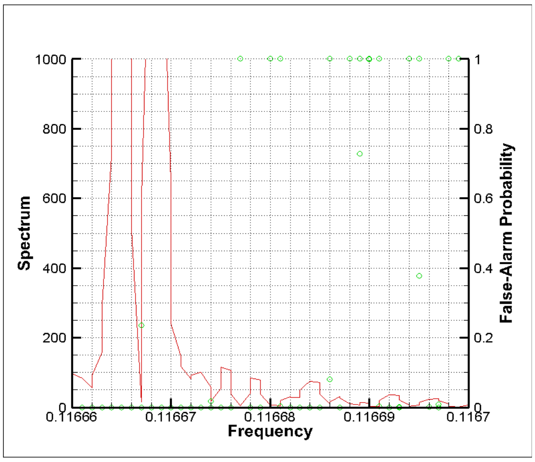


<그림 4> 2012년 1월 선물시장과 주식시장의 비교(저빈도 영역)



<그림 5> 2012년 1월 선물시장과 주식시장의 비교(고빈도 영역)

예를 들어 <그림 4>를 보면 선물시장의 경우 저빈도 영역, 즉 인간이 직접 매매가 가능한 영역에서 선물시장은 강한 알고리즘 트레이딩의 신호가 보인다. 또한 <그림 5>에서 보듯이 0.1 Hz 이상의 영역은 인간이 수동으로 직접 주문을 내기 어려운 빈도이므로 컴퓨터 시스템으로 자동화된 고빈도 거래로 여겨지며, 선물시장의 경우 주식시장과는 다르게 활발한 거래가 일어난다는 것이 확연히 드러난다. 이러한 분석 결과는 기존의 시장 통념에 잘 부합하는 것으로 보인다. 일별 결과와 주별 결과를 따로 첨부하지는 않았으나 그 결과는 월별 결과와 동일하게 해석된다.



<그림 6> 2012 1월 선물시장의 스펙트럼과 거짓확률 예시
(파워 스펙트럼-빨간색 실선, 거짓확률-연두색 동그라미)

한편 이러한 신호들이 유의미한지를 판단하는 기준인 거짓확률(false-alarm probability)의 경우 <그림 6>의 예에서 보는 것처럼 강한 신호로 여겨지는 스펙트럼에 대해서는 0에 가까운 확률이, 그 외의 잡음(noise)으로 보이는 신호에 대해서는 1에 가까운 확률이 나오므로

위의 결과에 대한 해석은 통계적으로 유의미한 것으로 볼 수 있다.

5. 결론

우리는 이 연구를 통해 선물시장과 주식시장의 거래 자료로부터 고빈도 거래의 식별 분석을 수행하였다. 금융 거래 자료의 경우 데이터의 간격이 불규칙하기 때문에 일반적으로 스펙트럼 분석에 많이 이용되는 전통적인 이산 푸리에 변환을 적용할 수 없다는 문제가 있다. 우리는 이러한 문제를 해결하기 위해 Lomb과 Scargle의 방법을 도입하여 KOSPI200 지수 선물 시장과 삼성전자 주식 시장에 대해서 스펙트럼 분석을 실시하였다. 그 결과 전반적으로 선물시장에서는 알고리즘 트레이딩 및 고빈도 거래가 활발하며 주식시장의 경우 그러한 거래가 거의 없는 것으로 드러났다. 이러한 분석 결과는 기존의 시장 통념과 잘 부합하는 것으로 여겨진다.

참고문헌

[1] Carr, P., and Madan, D.B. (1999), *Option valuation using the fast Fourier transform*, Journal of Computational Finance 2, 61-73.

[2] Duffie, D., Filipovic, D., and Schachermayer, W. (2003), *Affine processes and applications in finance*, Annals of Applied Probability 13, 984-1053.

[3] Lomb, N. R. (1976), *Least-Squares Frequency Analysis of Unequally Spaced Data*, Astrophysics and Space Science 39, 447-462.

[4] Scargle, Jeffrey D. (1982), *Studies in astronomical time series analysis. II. Statistical aspects of spectral analysis of unevenly spaced data*, The Astrophysical Journal 263, 835-853.

[5] William H. Press, Saul A. Teukolsky, William T. Vetterling, and Brian P. Flannery, *Numerical Recipes in C: The Art of Scientic Computing*, Second Edition, Cambridge University Press, 575-584.

[6] Hendershott, Terrence, Charles M. Jones, and Albert J. Menkveld (2011), *Does algorithmic trading improve liquidity?*", Journal of Finance 116, 1-32.

[7] Iacopo Giampaoli, Wing Lon Ng, Nick Constantinou (2009), Analysis of ultra-high-frequency financial data using advanced Fourier transforms, Finance Research Letters 6, 47-53