

# 바이노미얼 확률분포를 이용한 지수 예측 방법에 관한 연구

고영훈  
 협성대학교 컴퓨터공학과  
 e-mail:tigerko@uhs.ac.kr

## A Study on Index Prediction Method by Binomial Distribution

Ko Young Hoon  
 Dept of Computer Engineering, Hyupsung University

7

### 요 약

주식시장에서 개별 종목의 등락을 예측하는 것은 불가능하다. 미래가 정해져 있다면 그것을 아는 순간 거래는 성립되지 않기 때문이다. 따라서 개별 종목의 등락은 기업의 가치뿐만 아니라 투자참여자의 수급에 의해서 결정되므로 등락 확률은 예측불가인 0.5에 가깝다. 따라서 개별종목의 총합인 종합지수 역시 예측이 불가능해도 확률적인 틀은 제시할 수 있다. 바이노미얼 분포를 사용하여 n을 충분히 증가시키면 가우시안 분포가 되고 이를 이동평균선으로 지표화한 **Bollinger Band**를 이용하는 것이다. 중심선에 480일선을 상하한폭을  $2\sigma$ ,  $4\sigma$ 로 하여 그 틀을 제시하고, 이를 주요 종합지수로 검증하였다.

### 1. 서론

자본주의의 꽃이라고 불리우는 주식시장은 기업의 가치를 거래할 수 있는 시장이다. 최초의 주식회사는 중세의 네덜란드에서 인도와의 무역거래를 위한 대규모 자본조달을 목적으로 시작되었다. 이후 주식시장은 자금조달의 창구로서 수익배분의 수단으로서 발전하다가 주식의 위험관리를 위한 파생시장이 추가되었다. 종합지수에 대한 선물과 옵션의 거래가 가능하게 되어 다양한 전략으로 주식을 관리할 수 있게 되었다.

하지만 주식과 파생시장의 규모가 커짐에 따라 기업의 가치에 투자하는 가치투자가 외에도 단기차익을 노리는 투기거래자가 양산되었는데, 특히 투자금이 적고 레버리지 가 높은 옵션매수에 많은 투기거래자가 몰리고 있다.

투기거래자 뿐만 아니라 가치투자자에게도 종합지수의 움직임은 매우 중요한 관심사이다. 하지만 이를 예측하는 것은 불가능한데, 왜냐하면 종합지수는 주식시장에 참여하는 모든 사람들에 의해 만들어지기 때문이다. 만약에 종합지수를 예측할 수 있다면 모두가 한쪽 방향으로의 거래만을 하려고 하게되어 거래가 성립할 수 없기 때문이다.

하지만 종합지수의 움직임을 크게 보면 확률적인 틀 안에서 움직이게 되는데, 이는 개별 종목의 등락을 확률적으로 접근할 수 있기 때문이다.

동전을 던졌을 때 앞면이 나올지 뒷면이 나올지 알 수 없지만, 동전이 떨어지면 어떤 면인지 결정된다. 우리는 동전의 구조에서 확률을 예측할 수 있다. 동전을 n번 던지게 되면 앞면이 r번 나올 확률을 생각해 볼 수 있고, 이

를 분포로 나타낸 것이 바이노미얼 분포이다.

즉, 개별 종목의 등락을 확률로 나타낸다면 종합지수는 바이노미얼 분포의 확률적 틀에서 움직인다고 볼 수 있다. 본 논문에서는 이를 수학적 모델로 제시하고 기존 종합지수의 변화를 통하여 입증하려고 한다.

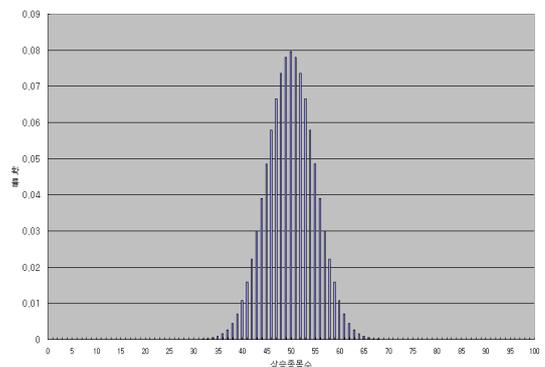
다음은 개별확률인 p인 사건을 n번 반복하였을 때 r번 발생할 확률은 나타낸다.

$$P(n, r) = {}_n C_r p^r q^{(n-r)}$$

만약 p가 0.5라면 다음과 같이 간략화할 수 있다.

$$P(n, r) = {}_n C_r p^n$$

바이노미얼 확률은 r이 n의 0.5 되는 지점에서 가장 큰 값을 갖는다. 만약에 n의 값이 매우 커진다면 바이노미얼 확률은 가우시안 분포와 같아진다.



(그림 1) 상승종목수에 따른 확률분포

따라서 이를 2000개가 넘는 종합지수에 적용한다면 개별 종목이 r개 이상 상승할 확률로 대치할 수 있고, 이는 CDF로 표시할 수 있다. 개별 종목의 등락은 투자자가 예측불가능한데, 매일매일의 등락은 기업가치뿐 아니라 수급에 의해서 좌우되기 때문에 예측불가능한 0.5로 통계적으로 나타난다.

$$CDF(n, m) = \sum_{r=0}^m C_r p^n$$

$$CDF\left(\frac{2m-n}{2n}\right) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{x^2}{2\sigma^2}}$$

가우시안 분포에서 중심값은 평균값으로 나타나며 종합지수의 이동평균선을 가우시안 분포의 중심값으로 대체하면 종합지수의 변동폭을 확률적인 관점으로 볼 수 있다.

## 2. Bollinger Band와 종합지수

이러한 확률적 틀에 가장 적합한 지표로 Bollinger Band가 있다. 1950년에 태어난 John Bollinger는 미국의 경제분석가로서 1980년대에 유명한 Bollinger Band를 발표한다. 중심선으로 이동평균선과 변동폭으로서의 표준편차의 n배수 상하한선으로 구성되어 있다. 일반적인 파라메타 값은 20선을 이동평균선으로 2σ를 상하한선으로 잡는다.

하지만 종합지수의 흐름을 보기 위해서는 한달간의 이동평균선인 20선은 적절하지 않다. 기업의 영업주기는 1년으로 1년마다 사업보고서를 제출하고 배당을 실시한다. 따라서 기업의 주기는 1년인 240일선을 봐야하며, 파동은 음과 양의 2배수로 이루어지므로 480일선을 중심선으로 보아야 한다. 상하한선은 두가지가 있는데 2σ와 4σ다. 일반적으로는 2σ를 넘지않지만 만약 넘는다면 불안정한 구역으로 진입하게 되어 4σ에 접근하면서 강한 저항에 부딪쳐 회귀압력을 받게 된다.

다음은 대한민국의 종합지수인 KOSPI지수의 Bollinger Band를 도표로 나타내었다. 480이동평균선을 중심선으로 하여 2σ선을 파란색으로 4σ선을 빨간색으로 표시하였다. 대부분의 경우에는 중심선과 파란색 안쪽에 분포하지만 간혹 파란선을 넘어서서 빨간선까지 움직이기도 한다. 2007년의 대세 상승장과 2008년의 금융위기가 그 대표적인 경우다.



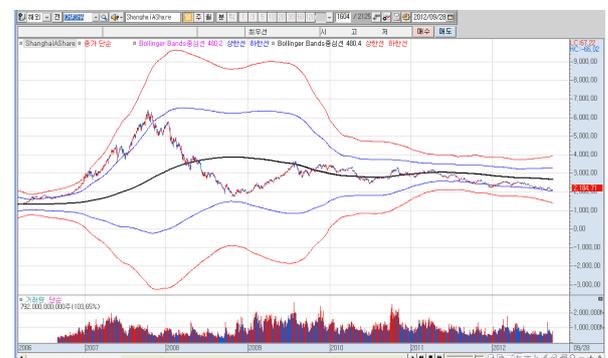
(그림 2) KOSPI의 Bollinger Band

미국의 다우산업지수도 같은 파라메타를 적용해서 나타내었다. KOSPI 지수의 경우와 비슷한 패턴을 나타내며 금융위기때 리먼 파산의 여파로 4σ선이 순간적으로 깨지는 강력한 충격파가 보인다.



(그림 3) 다우산업지수의 Bollinger Band

상해A지수를 보면 2007년까지 2σ를 넘어 장기상승하여 거품을 유도하였고, 이후 2008년에 지속적으로 하락하였다. 중국증시는 파생시장이 형성되기 이전이라 단조파동만 관찰되는데 이러한 현상은 전세계 증시에 광범위한 영향을 미치게 된다.



(그림 4) 상해A지수의 Bollinger Band

### 3. 결론

자본주의 사회라면 사회가치가 집약된 기업의 가치를 종합한 종합지수가 한 나라의 가치를 좌우한다고 해도 과언이 아니다. 이러한 종합지수를 예측한다는 것은 불가능하지만 확실적인 틀이 있다면, 매우 유용할 것이다. 본 논문에서는 개별 종목의 등락확률에서 시작하여 바이노미알 분포와 가우시안 분포까지 확장하여 이를 Bollinger Band로 실제 종합지수에 적용하여 확실적인 시각을 제시한다.

### 참고문헌

- [1] 고영훈, “MultiCharts의 포트폴리오를 위한 다중 진입 전략의 시그널 변환 시스템 설계”, 소프트웨어공학소사이어티 논문지, 제22권, 제1호, 2009, pp. 44-52.
- [2] 고영훈, 김윤상, “멀티차트를 사용한 종목군 계단식 매매 전략에 대한 성능 분석”, 디지털산업정보학회 논문지, 제6권, 제2호, 2010, pp. 225-231.
- [3] 고영훈, “옵션시장에서 풋취폴 전략의 성능분석”, 한국정보처리학회 추계학술발표대회 논문집, 제 17권, 제 1호, 2010, pp. 1051-1054.
- [4] 고영훈, 김윤상, “시스템 트레이딩에서 진입시점과 델타에 따른 스트래들매도의 성능 분석”, 디지털산업정보학회 논문지, 제6권, 제1호, 2010, pp. 151-157.
- [5] Ko, Young Hoon, "Analysis of Straddle trades using open interest of stock index futures in Korea option market", PPBRI 2010 Summer International Conference at CSUSB, 29 July 2010, pp. 18.