

# 새로운 패러다임, 데이터 그리고 데이터 분석

권불십년 화무십일홍(權不十年 花無十日紅)이라고 했다.  
어떤 일이든 영원히 지속되는 것이 없듯이 수년간 로우 레이터시로  
대변되던 해외 자본시장IT에도 최근 데이터 분석이라는  
새로운 변화의 바람이 불기 시작했다.



그동안 해외 자본시장IT는 로우 레이턴시로 대변되었다. 미국내 거래소들이 코로케이션Colocation과 DMADirect Market Access 서비스를 제공하면서 속도가 중요한 이슈로 부각된 것이다. 그러나 최근 주식시장을 살펴보면 이 이슈가 2009년과 2010년에 정점을 찍고 서서히 사라지고 있는 추세라는 점을 알 수 있다. 해외의 우수 자본시장IT 전문가들의 블로그나 뉴스에서 로우 레이턴시로 대변되는 속도에 대한 얘기가 점차 줄어들고 있는 중이다.

자본시장IT 기술 변화의 초석을 담당하던 로우 레이턴시에 대한 관심이 사라지는 이유는 크게 세 가지로 꼽을 수 있다. 첫째는 경쟁이다. 고빈도매매로 높은 수익에 도전하는 회사들이 늘어나면서 수익율이 낮아졌고 이 때문에 새로운 시장으로 옮기거나 전략을 바꾸는 흐름이 나타났다.

둘째는 기술적 비용의 증가 때문이다. 1마이크로세컨드를 줄이는데 들어가는 비용은 급속히 늘어나는데 비해 수익이 그만큼 늘어나지 않고 있다.

셋째는 규제다. 독일을 비롯하여 프랑스 및 미국도 고빈도매매를 규제하고 있거나 규제를 할 예정이다.

이런 현실적 상황에 봉착한 자본시장의 IT산업은 새로운 돌파구를 찾으려는 노력 중이다. 그 돌파구는 크게 4가지로 축약될 수 있다.

첫째는 금융공학을 기반으로 한 '새로운 발상', '새로운 위험 관리전략', '새로운 모델 찾기' 등이다. 예를 들면 다양한 종목간Cross-asset, 다양한 지역의 시간간Cross-regional/multi-temporal, 비대칭 및 대칭거래asymmetric versus symmetric trades다.

둘째는 자동매매다. 알고리즘 트레이딩이 고빈도매매를 거치면서 자동화는 알고리즘 트레이딩의 기본으로 정착되고 있다.

셋째는 고성능컴퓨터High Performance compute 기술을 기반으로 하는 새로운 트렌드의 형성이다. 로우 레이턴시와 고성능컴퓨터 기술이 결합한 기술이 새로운 흐름을 뒷받침하고 있는 것이다.

그리고 마지막으로 살펴볼 것은 빅데이터다. 2007년 하루 281억사바이트Exabyte이던 디지털 정보가 2011년 1800억사바이트로 급격히 늘어났다. 데이터의 증가도 증가지만 비정형 데이터의 비중은 더 크게 늘어났다. 자본시장에서는 어마어마한 데이터를 통해 자본시장의 흐름을 이해하고 예측하는 행위를 도전하고 있다. 이상 네 가지의 특징을 요약하면, 가장 빠른Fastest 매매에서 최선Best의 매매로의 전환이 현재 자본시장IT에서 나타나는 트렌드인 셈이다.





### 왜 데이터 분석인가

앞서 말한 자동판매와 고성능컴퓨터 기술이 고빈도매매의 유산이라고 할 수 있다. 하지만 새로운 알고리즘이나 모델은 빅데이터와 맞물리며 이전과 다른 양상을 보이고 있다. 이것을 과거와 구별하기 위해 세계적인 사모펀드인 콜버그크래비스로버츠KKR의 경영위험전문관리원Chief Risk Officer인 아틸리오 무치는 '데이터 중심 수치의 조직적 전략'이라고 부른다. 그렇다면 아틸리오 무치는 왜 데이터 중심이라고 불렀을까. 그 이유는 간단하다. 넘치는 데이터만 정확히 분류하고 분석할 수 있어도 다른 기업들과 차별화되는 새로운 전략을 창출할 수 있기 때문이다. 일반적으로 전략은 시세를 기반으로 한다. 물론 시세라고 하더라도 종류에 따라 중요성이 달라지기 마련인데 투자전략을 삼을 때 가장 중요한 시세는 초당 데이터다. 이 초당 데이터의 양은 상당히 많다. 미국의 OPRA Options Price Reporting Authority는 향후 초당 6백만 건 이상의 데

이터를 처리해야 한다고 발표했다. 당장 한국거래소에서 나오는 자료만 해도 초당 만 건에 가깝고 향후 더 늘어날 전망이다.

자본시장에서 중요한 초당 데이터만으로도 엄청난 자료를 실시간으로 처리할 수 있어야 하는 것이다. 여기에 더해 자본시장에서 활용하는 데이터에는 정부에서 발표하는 정책 혹은 공시도 들어간다.

금융투자업계가 다루어야 하는 정보는 이뿐만 아니다. 소셜 미디어나 뉴스를 통한 비정형 데이터도 다루어야 한다. 이런 데이터는 투자자에게 심리적인 영향을 주기 마련이다. 분명히 인터넷과 SNS로 유통되는 데이터로 시장의 경향을 분석하여 예측을 하는 곳도 점차 늘어나고 있다.

위에서 언급한 자본시장에 영향을 주는 데이터들을 규모 Volume, 속도Velocity, 다양성Variety의 관점으로 보면 요즘 말하는 빅데이터가 된다. 그리고 이 빅데이터를 정확히 분석하면 과

거의 표본방식의 데이터 분석과 달리 정확도를 높일 수 있다. 더욱이 데이터를 처리하기 위하여 지불하여야 하는 비용도 상대적으로 낮아지고 있다. 즉 막대한 양의 데이터를 빠르게 분석해낼 수 있는 기술과 인프라가 구축되면서 이제 데이터 분석Data Analytics은 트레이딩의 선택이 아니라 필수가 되고 있는 중인 셈이다.

최근 이렇게 빅데이터를 기반으로 컴퓨터가 처리할 수 있는 환경을 만들고 이를 바탕으로 경쟁자보다 빨리 처리하도록 하는 트레이딩을 스마트 트레이딩Smart Trading 혹은 인텔리전트 트레이딩Intelligent trading이라고도 부른다. 비즈니스 통합 솔루션 개발 업체인 톱코가 제공한 리포트인 '인텔리전트 트레이딩의 아키텍처 - 이익 증가를 위한 증가하는 빅데이터의 활용' An Architecture For Intelligent Trading - Leveraging Big Data in Motion For Increased Profits을 보면 빅데이터와 결합한 인텔리전트 트레이딩으로 멀티-레그 전략Multi-leg strategy,마켓 트렌드 예견, 경제 지표/지수 추척 등의 사례를 소개하고 있다. 모두 데이터 분석을 기반으로 하고 있다.

### 데이터 분석이 필요한 이유

그렇다면 정확한 데이터 분석을 위해서는 무엇이 필요할까. 우선 정형 데이터와 비정형 데이터뿐 아니라 과거 데이터 및 실시간 데이터를 수집해야 한다. 이처럼 자본시장에 영향을 줄 수 있는 모든 데이터를 수집하고 분석해 결과를 도출하는 과정을 시스템화하는 것이 데이터 분석의 시작이다. 하지만 여기서 끝이 아니다. 여기에 인텔리전스를 결합 하여야만 차별점을 얻을 수 있다.

데이터 전문 기업인 엑셀레이트Exelate는 보고서를 통해 빅데이터와 인텔리전스가 결합하여 얻을 수 있는 것을 스마트 데이터라고 했다. 빅데이터와 스마트 데이터는 세 가지 3A Accurate, Actionable, Agile로 구별할 수 있다. 첫째 정확성으로 스마트 데이터는 빅데이터의 노이즈로부터 정확하고 양질의 정보를 전달할 수 있어야 한다.

둘째 행동성으로 스마트 데이터는 조직이 바로 행동으로 옮길 수 있는 특성을 보유함으로써 가치를 창출하는 원동력으로서 작동해야 한다. 셋째 민첩성으로 급변하는 비즈니스 환경에서 실시간으로 데이터 분석이 가능하여야 한다. 자본시장에서의 데이터 분석은 어떤 역할을 할까? 일반적으로 시장위험Market Risk, 실행위험Execution Risk 등과 같은 다양한 위험에 대한 노출을 줄일 수 있고 변화하는 시장조건에 따라 시의적절하게 혹은 실시간으로 대응할 수 있도록 한다. 또한 다양

어떤 보고서는 빅데이터를 스마트 데이터로 바꿀 수 있는 데이터 리더십이 시장에서 승리할 것이라고 했다. 지금 필요한 것은 데이터와 데이터 분석력인 것이다.

한 정보들이 결합한 알고리즘을 이용하여 최대 결과를 얻을 수 있도록 도움을 준다.

좀더 구체적으로 들어가면 첫째 연구 분석 및 시험에서 강력한 경쟁력을 갖도록 한다. 다양한 데이터와 이를 이용할 수 있는 환경은 알고리즘을 개발하고 운용으로 전환하는 시간을 줄여준다.

둘째 줄어든 의사결정 시간만큼 시장에서 성과를 얻을 수 있는 가능성이 그만큼 커진다. 예를 들어 고빈도매매를 초기에 도입한 회사들이 뒤늦게 참여한 회사들에 비해 훨씬 높은 수익을 얻은 것과 같은 이치다. 셋째는 과거 데이터와 실시간 데이터를 통하여 운용전략을 효율화하여 경쟁력을 높일 수 있다.

### 데이터 분석의 미래

데이터 분석은 컴퓨터공학, 수학(금융공학, 통계학) 및 금융 지식의 융합을 바탕으로 한다. 즉 세 가지 영역의 발전은 데이터 분석의 미래에 영향을 주게 된다는 것이다. 물론 이 중에서도 금융이나 수학의 발전 흐름을 볼 때 컴퓨터공학의 발전이 가장 큰 영향을 주리라 생각된다.

데이터 (분석)를 기반으로 트레이딩은 어떻게 진화할까? 명확한 지향을 내놓는 것은 불가능하여도 최소한의 경향을 예측할 수 있다. 첫째 실시간 분석이다. 현재 빠르게 데이터를 처리할 수 있는 기술, 예를 들어 FPGA와 같은 기술이나 SSD를 이용한 메모리기술이 나오고 있다. 가격은 점점 낮아지고 있다. 배치처리를 하였던 분석도 실시간 분석으로 전환되고 있다. 둘째 머신학습이다. 도이치 뱅크Deutsche Bank가 내놓은 보고서인 머신의 부상The Rise Of Machine에서는 머신학습을 중요한 변화로 소개하고 있다. 이제 시작 단계인 머신학습이 트레이딩과의 접점을 넓혀 갈 것으로 예상된다. 어떤 보고서는 빅데이터를 스마트 데이터로 바꿀 수 있는 데이터 리더십이 시장에서 승리할 것이라고 했다. 지금 필요한 것은 데이터와 데이터 분석력인 것이다. 18