Decision Tree를 이용한 주가의 본연 가치로부터의

괴리도 추정과 이를 이용한 거래 전략의 평가

2022.06.18

김현준, 박민규, 성시현

Report for Financial Machine Learning Course, SSU

초록

투자자가 자산이 가지는 불확실성에 대해 보이는 반응은 흔히 비합리적인 것으로 알려져 있다. 이러한 비합리적인 반응 중 하나는 이미 시장에 반영된 과거 정보를 이용한 투자 전략을 구사하는 것 등이 있을 수 있다. 본 보고서는 그러한 비합리적 투자자의 행태를 포착하기 위해 기업의 과거 재무제표 정보와 주가 수익률 정보를 이용한 의사결정나무 모형을 사용하였으며, 그 결과 실제 수익률에서 비합리적 투자자들의 행태로 인한 본연 가치로부터의 괴리 수준을 추출해낼 수 있었다. 또한, 주식 평균 수익률이 기대수익률로 회귀한다는 가정 하에서 괴리 수준에 대한 공매도를 허용하지 않는 콘트래리언 태도를 취한 전략 포트폴리오의 수익률이 통계적으로 유의한 초과수익을 발생시킴을 확인하였다. 본 보고서의 연구는 효율적 시장 가설에 부합하면서도 초과수익을 추구할 수 있는 통계적 방법론을 제시하였음에 의의가 있다.

1. 서론

오랫동안 금융연구 분야에선 위험을 가진 자산엔 장기적으로 균형 기대수익이 존재할 것으로 기대되어 왔고, 이러한 기대에 부합하는 이론을 만들거나 기존 이론을 보완하기 위해 자본자산결정모형, Fama French 3 factor model, arbitrage pricing theory와 같은 모형들이 개발되었다. (Fama & French, 1993; Ross, 2013; Sharpe, 1964) 그러나 비록 이러한 모형들의 sample 내 설명력은 높으나, out of sample 예측력은 강건하지 못한 것으로 알려져 있다. 이러한 현상이 일어나는 잠재적 이유 중 하나는 새로운 정보에 의해서 동적으로 변화하는 금융시장의 성질 상 현재의 가격결정 방식이 먼 미래까지 유지되진 못하기 때문이다. 이와 연관된 사실로, 고빈도 데이터를 이용한 자본자산결정모형은 미래 가격결정 방식을 예측하는데 더 유리하다는 선행연구가 있다. (Hollstein et al., 2020)

하지만 결정적으로 모형의 미래 예측 능력이 떨어지는 대표적인 이유는 가격결정방식과는 관계없는 시장 잡음이 과도하게 크기 때문으로 볼 수 있다. 이러하게 기업의 본연 가치와는 연관되지 않은 초과 변동성에 대한 증거는 Shiller(1987, 1990) 등에 의해 발견되었다. (Shiller, 1987; Shiller, 1990). 이후, 이러한 초과 변동성이 시장의 불확실성, 투자자의 attention 및 overconfidence 등과 같은 심리적 요소, 그리고 이 둘 간의 상호작용으로 인한 attention bias 등의 합성적 효과 등이 연관성이 높다는 점이 발견되었다. 먼저, 시장 불확실성과 자산에 대한 투자자의 attention은 시장 변동성에 대해 선형적이지 않은 영향을 준다고 알려져 있다. (Andrei & Hasler, 2015) 또한, 투자자의 attention은 제한된 자원이라는 점에서, 투자자들이 종목 하나하나에 대한 관심보다는 종목들을 분류하여 나타나는 category에 대해서 관심을 집중한다는 것이 밝혀졌으며, 이는 투자자의 overconfidence와 연결되어 시장의 초과 변동성에 기여한다. (Peng & Xiong, 2006) 마지막으로, 시장 불확실성이 커지면, 소형주 등에서 과거 기초적 지표에 대한 attention bias가 발생함 또한 보고되었다. (Hong & Wu, 2016)

이러한 선행연구들은 미래 예측을 통한 수익창출을 위해 사용될 가격결정모형은 계속 계수를 새로운 데이터에 대해서 갱신하는 동적인 모형이면서, 과거 데이터로 인한 excess volatility를 감지하고 이를 추정 오차에서 제거해내는 능력이 필요할 것을 시사한다. 본 보고서는 그러한 모형을 Decision Tree의 형태로 구현하며, 이를 이용한 투자 전략의 성과 또한 평가한다. 2장에서는 모형의 구조와 이러한 모형이 어떤 통계학적 근거를 가지고 있는지를 탐색하며, 3장에서는 실제 데이터에 적용하여 성과를 분석한다. 4장에서는 보고서를 결론짓는다.

2. 모형

해당 연구에서는 두 가지 모형이 사용된다. 첫째는 펀더멘털 괴리도를 추출하는 모형, 둘째는 펀더멘털 괴리도를 이용한 거래 시그널 생성 모형이다. 펀더멘털 괴리도 추출 모형은 t시점에서 특정 시간 가격 후인 사이의 과거 총자산 변화율, 총부채 변화율, 매출액, 영업이익, 그리고 이전 주가 수익률을 input으로, 현재 주가 수익률을 output으로 내놓는 Decision Tree를 최적화하는 과정과, 시점의 out of sample 데이터를 이용해 실제 현재 주가 수익률에서 머신러닝 알고리즘이 추정하는 현재 주가 수익률을 빼는 작업을 진행한다.

이렇게 계산된 값이 unbiased 추정치를 이용한 점추정에 비해 더 효과적으로 본연 가치로부터의 괴리도를 계산할 수 있다는 수리통계적 근거는 다음과 같다.

앞서 선행연구에서 보았듯, 초과변동성에서 기원한 본연 가치로부터의 괴리도는 과거 및 현재 시장 변동성, 과거 재무제표 데이터 등에 의존적인 투자자의 비이성적인 태도에 의한 것으로 볼 수 있다. 그렇다면, 어떤 주식의 특정 시점의 수익률은 다음과 같이 분해할 수 있다.

여기서 는 i 시점의 주식 수익률, 는 i 시점에서의 주식 기대수익률, 는 i 시점에서의 본연 가치로부터의 괴리 수준, 는 i 시점에서의 정보 충격이다. , 라는 각각의 분포 가정 또한 필요할 것이다.

여기서, 만약 본연 가치로부터의 괴리 수준이 과거 재무제표 및 과거 주가 수익률에 대한 정보를 제공한다면, 세 변수에 대해 다음과 같은 연관관계를 가정할 수 있다.

여기서, 는 i-1 시점에서의 재무제표 데이터이고, f는 알려지지 않은 어떤 결정적(deterministic) 함수이다.

만약 이렇게 되면, 모형은 에 대한 설명력을 지녀 추정치의 분산은 만큼 하락하지만, 이므로 현재 수익 추정량 은 이고 그 분산은 이다. 그러므로 한 시점의 현재 수익(단일값)에 모형의 추정치 간의 차 는 이고, 그 분산은 이다. 분산이 의 분산을 포함하지 않으므로, 통계적으로 유의한 의 검정이 가능하다.

이렇게 생성된 본연 가치로부터의 괴리도 의 크기에 비례하게 반대매매를 진행하면, 초과수익이 있을 것으로 예상할 수 있다. 다만, 괴리도가 양수일 때에는 빠르지 않은 차익거래로 인해 수익이 다음 거래일 안에 발생하지 않을 수 있으므로, 괴리도가 양수일 때의 공매도는 허용하지 않는 것으로 하였다. 거래 시그널, 혹은 매수 비중은 생성된 괴리도의 첫 1/4 데이터의 평균과 표준편차를 이용해 나머지 3/4를 표준화한 값을 이용하였다. 여기서, 나머지 3/4 중 1/3은 의사결정나무의 max depth를 최적화하기 위한 validation set으로 사용하였으며, 전략 수익률에 대한 보고 사항은 나머지 2/3을 이용해 이루어졌다.

3. 실증분석 및 전략분석

본 보고서엔 위의 전략의 유의성을 검증하기 위해 코스피 상위 100개, 코스닥 상위 100개 종목을 선택하여 이들의 월별 수익률과 연간 재무제표 정보(총자산 변화율, 총부채 변화율, 매출액, 영업이익)를 이용해 전략을 구현하였다. 여기서, 데이터 구조상 해당 연도의 재무제표 정보는 당시 접근 가능한 정보가 아닌 미래에 발표된 해당 연도의 재무정보에 관한 것으로, 투자 전략 구현 중 연간 재무제표 정보의 data leakage를 막기 위해, 18개월만큼 재무제표 정보의 시간대를 움직일 필요가 있었다. 또한, 거래 시그널을 이용한 전략에서, 투자 수익률은 시그널 생성 시점에서 1달 뒤의 수익률을 사용하여 data leakage를 방지하였다. 거래 시그널은 모든 개별 주식에 대해서 생성되었으며, 매매 시그널에 따라 월마다 종목들의 비중을 바꾸었다. 이렇게 생성된 매수 종목들을 동일가중으로 매수한 포트폴리오를 분석 대상으로 삼았다.

* 1. 전략의 젠센 알파

가장 기본적으로 확인할 수 있는 전략 평가법은 벤치마크 지수의 변동이 설명하지 못하는 전략의 평균 수익을 계산한 젠센 알파이다. 그 계산을 위한 회귀식은 아래와 같다.

는 포트폴리오 수익률, 은 벤치마크 수익률, 는 젠센 알파이다.

투자 대상 자산의 동일가중포트폴리오 (EWP)를 벤치마크로 하여 젠센 알파를 도출한 결과, test set데이터 구간 내에서 월간 2.54%의 젠센 알파가 통계적으로 유의하게 존재하는 것으로 보았다. 통계량은 0.01 유의수준에서 유의하였다.

Table 1

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | coef | std err | t | P>|t| | 유의성 |
| const | 0.0254 | 0.007 | 3.749 | 0 | \*\*\* |
| 벤치마크 | -0.2206 | 0.111 | -1.993 | 0.051 | \* |

* 1. HM 시점선택 모형 기반 시점선택 능력 평가

Henriksson과 Merton (1981)의 방식을 원용한 HM 시점선택 모형 또한 전략의 유효성 검증에 사용되었다. (Henriksson & Merton, 1981) HM 시점선택 모형의 수학적 형태는 아래와 같다.

는 포트폴리오 수익률, 은 벤치마크 수익률, 는 이 양수일 땐 해당값, 음수 및 0인 경우엔 전부 0으로 처리한 가상 옵션 수익률, 는 상수이고 는 가우시안 노이즈이다. 통계적 분석 결과는 Table 2에 있다.

Table 2

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | coef | std err | t | P>|t| | 유의성 |
| const | 0.0128 | 0.01 | 1.288 | 0.203 |  |
|  | -0.5569 | 0.224 | -2.482 | 0.016 | \*\* |
|  | 0.5741 | 0.335 | 1.714 | 0.092 | \* |

여기서는 일반적인 0.05 유의수준에서 유의한 시점선택 능력이 측정되지 않았다. 하지만 이것은 과 이 가지는 높은 상관관계가 발생시킨 다중공선성으로 인한 결과일 확률이 높다. Table 3엔 독립변수 및 종속변수 간의 상관계수가 나와 있다.

Table 3

|  |  |  |  |
| --- | --- | --- | --- |
|  | port | RM | RM+ |
| port | 1 | -0.255195 | -0.118417 |
| RM | -0.255195 | 1 | *0.874408* |
| RM+ | -0.118417 | *0.874408* | 1 |

Table 3를 참조하면, RM과 RM+ 간엔 0.9에 가까운 피어슨 상관계수를 가지고 있음을 알 수 있다. 이는 EWP의 월별 수익률이 대부분 양수였기 때문을 추정되며, 그것은 회귀계수의 표준오차가 팽창하여 올바른 통계검정이 이루어지지 못하였을 가능성을 시사한다.

* 1. Jiang의 비모수 시점선택 모형

위의 HM 시점선택 모형과 같은 모수적 방법론이 가지는 문제점을 해결하기 위해, Jiang이 제시한 비모수 시점선택 검정 방법을 원용하여 시점선택 능력을 다시 테스트하였다. (Jiang, 2003) 구체적인 검정 방법은 매우 복잡하여 본 보고서에 기재하지 않는다. 다만, Jiang이 제시한 통계량은 벤치마크 기준 전략 수익률의 convexity를 측정하는 통계량으로, 표본이 커짐에 따라 정규분포로 수렴함이 증명되었기에 t-test 및 z-test가 가능함을 언급한다. 그 결과는 Table 4에 제시되어 있다.

Table 4

|  |  |
| --- | --- |
| 통계량 명 | 통계량 |
| U-statistic추정치 | 0.119 |
| 표준오차 | 0.035 |
| 통계량 | 3.345 |
| p값 | <0.01 |

검정 결과, z-statistic은 3.345로 유의수준 0.01에서 유의하였다. 그러므로, 본 투자 전략의 시점선택 능력이 존재한다고 보는 것이 적절하다.

4. 결론

이러한 분석 결과를 통해, 본 보고서는 과거 정보가 미래 주가 수익률에 영향을 미치지 않는 효율적 시장 가설에 반하지 않으면서도 초과수익을 누릴 수 있는 전략에 대한 가능성을 열었다는 점에서 의의를 가진다. 비록 상장폐지 종목 및 가치가중평균을 통한 비체계적 위험의 제거 등이 이루어지지 않은 상태에서의 평가이기에 여러모로 초과수익의 유의성의 적절함에 대한 논의가 필요하긴 하나, 행동경제학적 특성을 데이터 기반 투자전략으로 연결짓는 시도에 의의가 있으며, 이러한 한계점들은 후속 연구에 의해서 적용되어 검증되어야 할 필요가 있다.

References

Andrei, D., & Hasler, M. (2015). Investor attention and stock market volatility. *The Review of Financial Studies, 28*(1), 33-72.

Fama, E. F., & French, K. R. (1993). Common risk factors in the returns on stocks and bonds. *Journal of Financial Economics, 33*(1), 3-56.

Henriksson, R. D., & Merton, R. C. (1981). On market timing and investment performance. II. Statistical procedures for evaluating forecasting skills. *Journal of Business,* , 513-533.

Hollstein, F., Prokopczuk, M., & Wese Simen, C. (2020). The conditional Capital Asset Pricing Model revisited: Evidence from high-frequency betas. *Management Science, 66*(6), 2474-2494.

Hong, K., & Wu, E. (2016). The roles of past returns and firm fundamentals in driving US stock price movements. *International Review of Financial Analysis, 43*, 62-75.

Jiang, W. (2003). A nonparametric test of market timing. *Journal of Empirical Finance, 10*(4), 399-425.

Peng, L., & Xiong, W. (2006). Investor attention, overconfidence and category learning. *Journal of Financial Economics, 80*(3), 563-602.

Ross, S. A. (2013). The arbitrage theory of capital asset pricing. *Handbook of the fundamentals of financial decision making: Part I* (pp. 11-30). World Scientific.

Sharpe, W. F. (1964). Capital asset prices: A theory of market equilibrium under conditions of risk. *The Journal of Finance, 19*(3), 425-442.

Shiller, R. J. (1987). The volatility of stock market prices. *Science, 235*(4784), 33-37.

Shiller, R. J. (1990). Market volatility and investor behavior. *The American Economic Review, 80*(2), 58-62.

stylefix