

Real 퀀트

가치성장 상대 강도/코스피 예측부터
코인 가격 예측까지



Contents

Key Chart: 마켓 타이밍 모델 누적수익 추이	4
1. 정성적 예측이 어려운 주식 시장	6
지긋지긋한 “경기침체 vs. 금리 인하 기대” 구도	6
2. 시장(market)을 분석하는 것이 중요한 이유.....	8
“팩터 모델적 사고”를 해보자	8
3. 가치(Value) 팩터의 등장	10
파마프렌치의 3 요인 모형	10
3 요인 모형의 문제점: 양(+)의 프리미엄의 상실	11
4. 파마프렌치 사고방식에서 벗어나야.....	12
상상할 수 있는 모든 팩터를 감안해야 한다	12
5. 기술적 지표를 감안해야 하는 이유	14
차티스트(chartist)의 냉혹한 승부 세계: 코인 시장 예시	14
변동성이 높은 자산은 모멘텀으로 대해줘야 한다	15
6. 퀀트의 기초 덕목: 표본 외 검증.....	16
‘과거의 나’의 전략, 좋은 성과로 이어졌을까?	16
7. 실전! 마켓 타이밍 모델 설계하기	20
모델 방법론	22
Backtesting Result 1. 가치/성장 상대 강도	24
Backtesting Result 2. 코스피 지수	26
Backtesting Result 3. 비트코인	28
Backtesting Result 4. 이더리움	30
8. 현재는 성장주 우위 국면: 실적 스크리닝 종목.....	32
왜 실적을 고려해야 하는가?	32
성장 + 실적 모멘텀 스크리닝 상위 종목	33
Appendix. 잠재 요인을 추출하는 요인 모형과 그 한계.....	34

Quant Strategy

허석민 Quant Analyst

Real 퀀트는 어떻게 분석하는가?

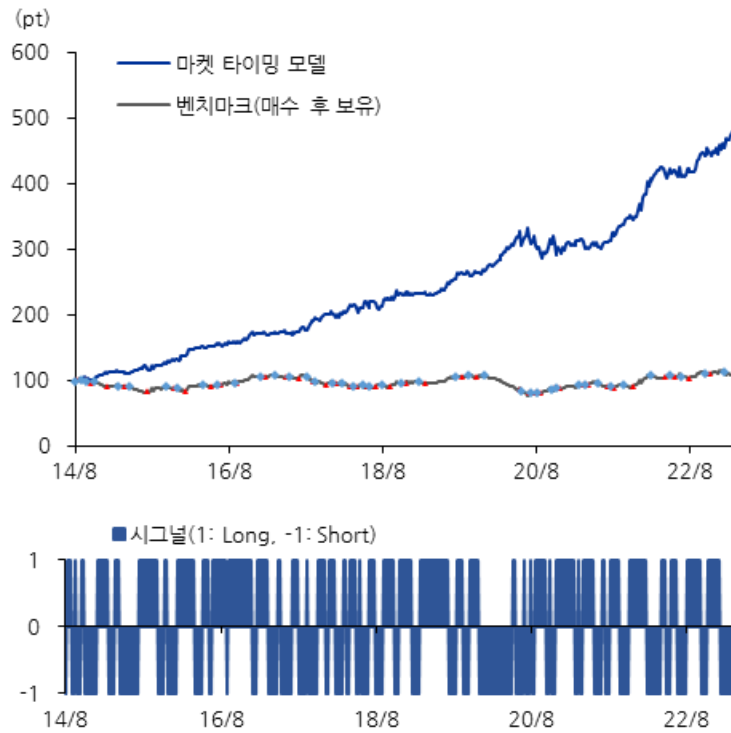
퀀트(Quant)는 ‘계량적인’이란 뜻의 영어단어 ‘quantitative’에서 파생됐다. 따라서 일반적으로 퀀트는 상품 개발, 리스크 관리, 알고리즘 트레이딩 등 수치를 통한 정량적인 분석을 수행하는 금융분석가를 의미한다. 하지만 필자는 퀀트의 핵심이 ‘계량적인 분석’ 그 자체는 결코 아니라고 생각한다. 퀀트의 핵심은 바로 과거 데이터로 최대한 보수적인 검증을 하고자 하는 시도이며, 수치 분석은 그러한 검증 과정에 있어서의 수단에 불과하다.

모든 투자 전략의 성패는 지속성에 달려있다. 대부분의 경우 투자 원칙을 정하더라도 그 원칙을 고수하지 못한다. 가장 큰 이유는 바로 심리다. 사람은 태생적으로 위험회피적인 성향을 가지고 있기 때문이다. 퀀트가 미래를 예측할 수는 없다. 하지만 면밀하고, 보수적인 접근 방식은 심리적 지지선으로 작용할 수 있다.

본 자료에서는 비지도학습 모델을 활용하여 1) 가치/성장 스타일 우위 국면, 2) 코스피 지수, 3) 주요 코인(비트코인, 이더리움) 예측을 시도하였으며, 표본 외 샘플 백테스팅을 통해 마켓 타이밍 전략의 유효성을 확인하였다. **4/10 일 기준으로 현 국면은 코스피 롱, 성장주 우위, 비트코인 및 이더리움 숏 우위로 판단된다.** 코스피 지수/코인 가격 예측은 일간, 가치/성장 상대 국면 예측은 주간(매주 마지막 영업일 종가 기준)으로 이뤄지며, 추후 시그널 변경시마다 업데이트할 예정이다.

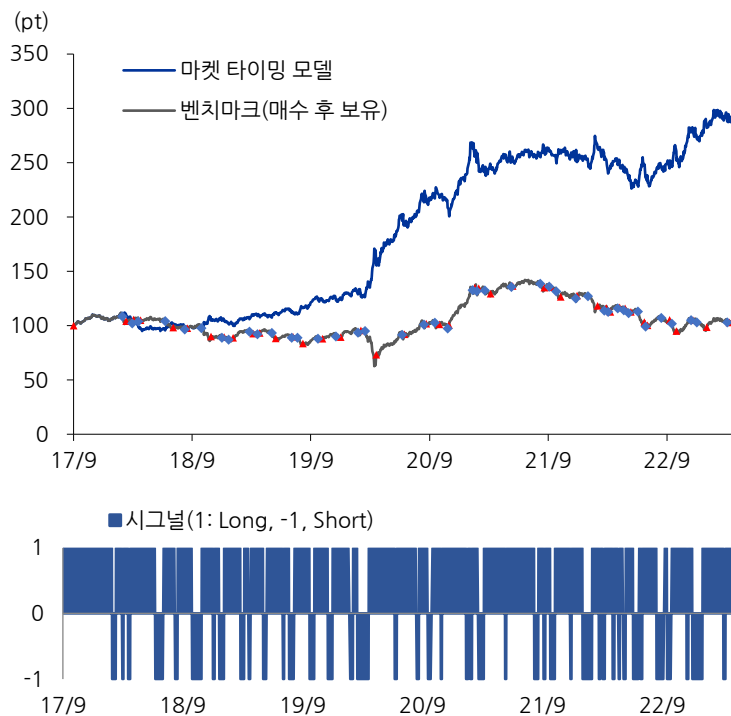
Key Chart: 마켓 타이밍 모델 누적수익 추이

Backtesting Result 1. 가치/성장 상대 강도



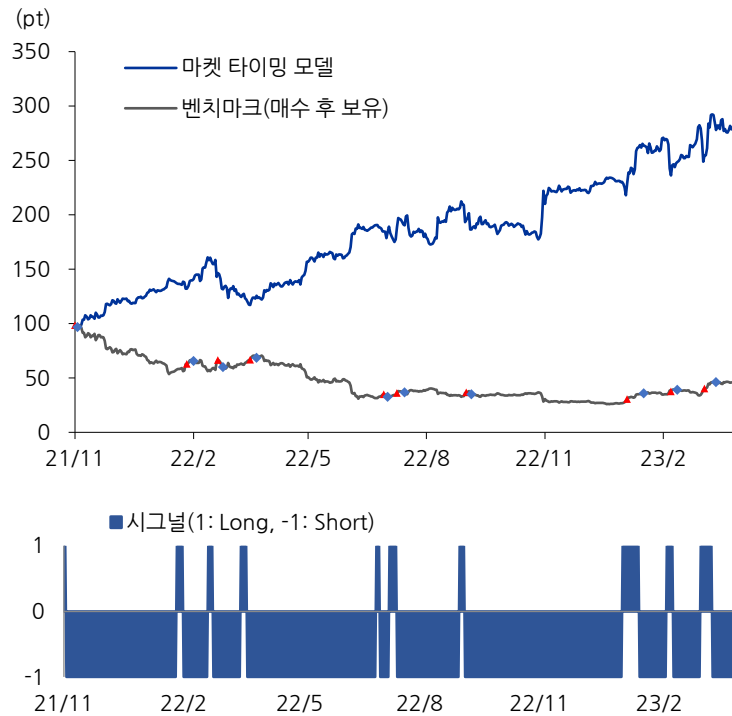
자료: 유진투자증권

Backtesting Result 2. 코스피 지수



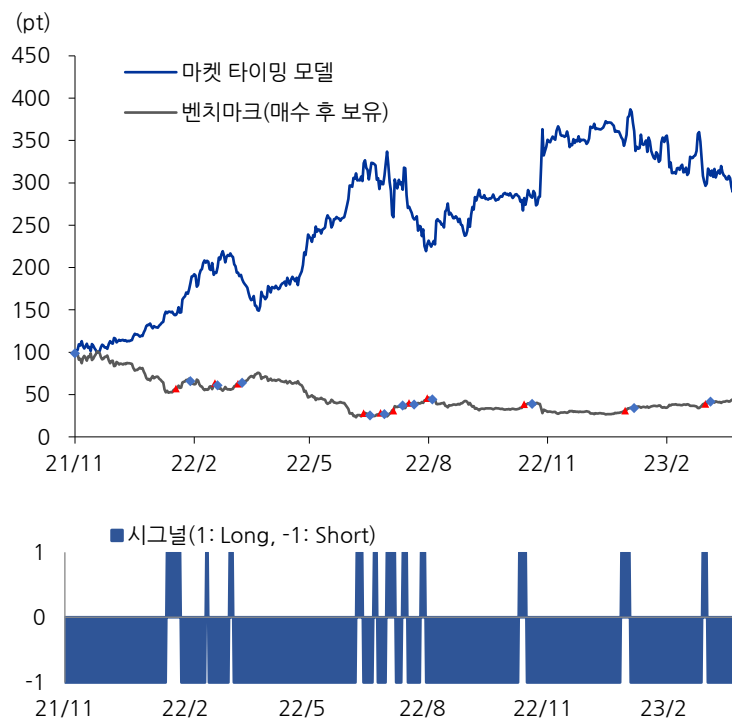
자료: 유진투자증권

Backtesting Result 3. 비트코인



자료: 유진투자증권

Backtesting Result 4. 이더리움



자료: 유진투자증권

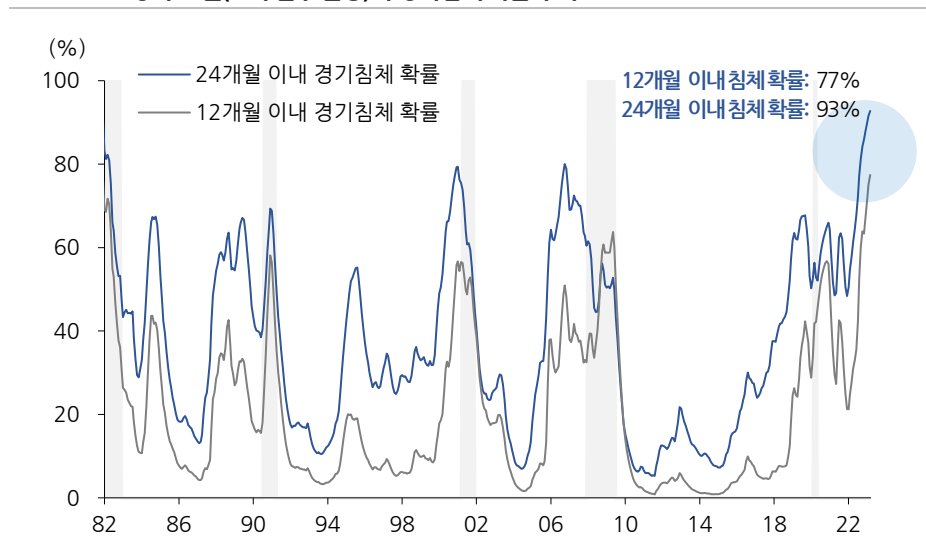
1. 정성적 예측이 어려운 주식 시장

지긋지긋한 “경기침체 vs. 금리 인하 기대” 구도

사실상 미국 경제 지표는 침체가 임박했거나, 이미 진입했음을 가리키고 있다. 당사 경기침체 예측 모델이 제시한 침체 확률은 22년초 이후부터 상승 추세를 유지하며 22년 3월 기준 모델 샘플 구간 내 최대치를 기록 중이다(도표 1 번). 타 기관 대비 당사 모델 예측 성능의 우위를 감안하면 결코 무시할 수 없는 수치다(도표 2 번). 하지만 작년 중순부터 나타난 ‘경기 침체 우려’, 그리고 이로 인한 ‘긴축 완화 가능성’의 상충 구도가 아직까지도 지속되고 있다. 경제 지표와 같은 전통적인 요인만으로는 단기적 흐름을 예측하는 것이 점점 어려워지고 있다.

역설적이게도 미국 증시는 이익 추정치 하향 폭만 감안하면 침체 가능성을 거의 반영하지 않고 있다. 증시 하락이 거의 밸류에이션 하락에 기인하고 있기 때문이다(도표 4 번). 국내 증시는 지수 레벨/이익 추정치 하향 폭 모두 미국 증시 대비 상대적으로 큰 폭으로 나타났으나, 이익 하향 폭 대비 지수 레벨 하향 폭이 충분치는 않다(도표 5 번). 특히 연초 이후 이익 추정치 하향은 지속된 반면 지수는 상승하면서 오히려 현재 밸류에이션 부담을 무시할 수 없는 상황이 됐다(도표 6 번).

도표 1. 당사 모델(5개 변수 활용)의 경기침체 확률 추이

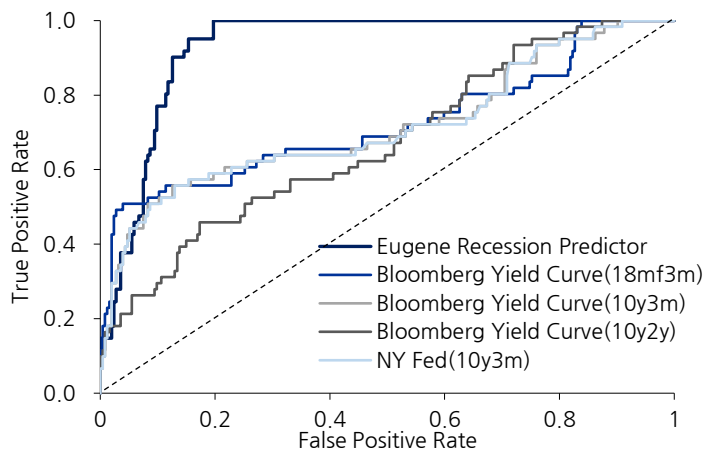


자료: 유진투자증권

주: 음영은 NBER 분류 기준 침체 구간

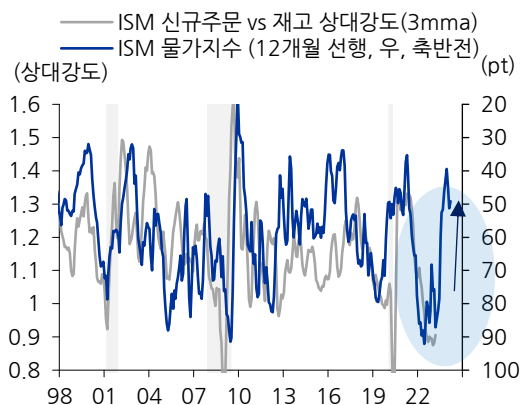
주: 사용 변수 등 구체적인 방법론은 <퀀트로 경기침체 확률 읽기 feat. AI 예측 모델, 22.08.01> 참고

도표 2. 기관별 침체 모델의 ROC 커브(좌상단에 위치할수록 모델의 성능이 좋음)



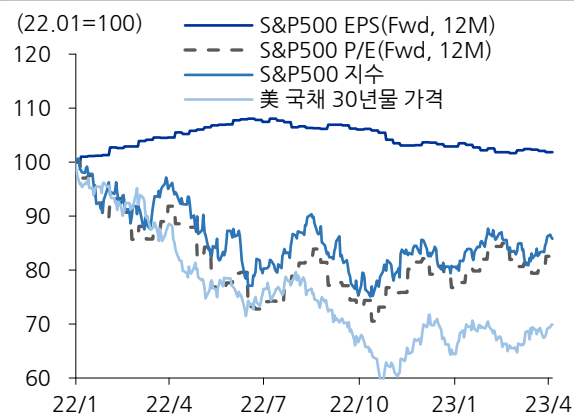
자료: Bloomberg, NY Fed, 유진투자증권

도표 3. 물가 둔화에 따른 턴어라운드 기대 작용하는 듯



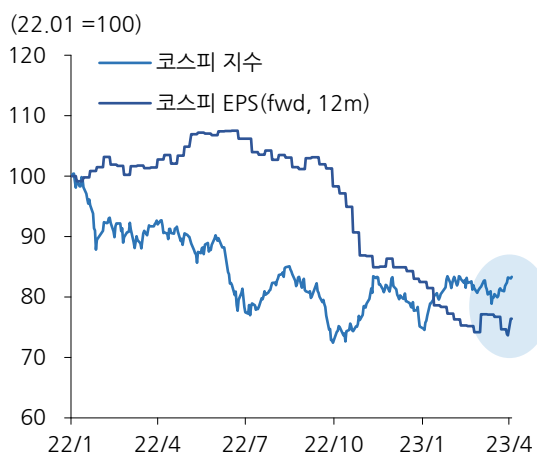
자료: Bloomberg, 유진투자증권

도표 4. S&P500 하락은 주로 밸류에이션 하락에 기인



자료: Bloomberg, 유진투자증권

도표 5. 코스피 지수/이익 하향 폭은 상대적으로 크지만...



자료: Quantwise, 유진투자증권

도표 6. ...밸류에이션 부담은 오히려 높아진 상황



자료: Quantwise, 유진투자증권

2. 시장(market)을 분석하는 것이 중요한 이유

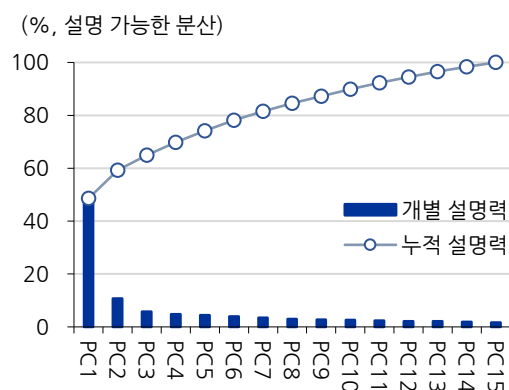
“팩터 모델적 사고”를 해보자

앞서 살펴봤던 것처럼 코스피와 같은 시장 지수의 흐름을 분석하는 것은 무슨 의미가 있을까? 너무도 당연하게 여겨질 수 있는 질문이지만, 팩터 모델의 관점에서 따져보도록 하자. 원래 진리는 단순한 법이다.

금융시장에 참여하고 있는 사람 모두는 이미 팩터 모델적 사고에 익숙하다. 금융에서의 팩터(factor)란 말 그대로 자산의 수익률에 영향을 끼치는 모든 요인을 의미한다. 그리고 시장 요인은 그 중 가장 강력한 팩터다. 소위 “장세”가 좋고 나쁨에 따라 개별 주가는 큰 영향을 받게 된다. 실무에서 자주 들을 수 있는 ‘시장 베타’가 바로 시장 요인(주가 지수 등)에 대한 주가 수익률의 민감도(베타)이며, 실제 개별 주가 수익률에 대해 가장 높은 설명력을 가진다.

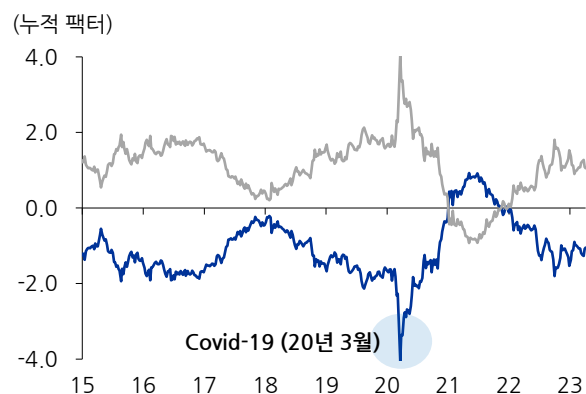
퀀트답게 조금 더 계량적으로 살펴보도록 하자. 도표 7 번은 코스피 시총 상위 50 개 종목의 수익률에 영향을 끼친 상위 요인을 통계적으로 추출한 후 각 요인의 설명력을 나타낸 것이다. 첫 번째 요인이 약 50%에 가까운 높은 설명력을 가지고 있으며, 상위 잠재 요인 6 개가 주가 수익률을 약 80% 수준 설명 가능한 것이 확인된다. 여기서 이미 짐작 가능하겠지만, 첫 번째 요인이 바로 시장 요인이다(도표 8 번). 즉, 시장이 주가 수익률에 대해 약 50%에 가까운 설명력을 갖고 있다는 것이다.

도표 7. 주가를 설명하는 잠재 요인의 설명력



자료: 유진투자증권

도표 8. 첫 번째 잠재 요인 누적 시계열: 시장 요인



자료: 유진투자증권

주: 요인 추출 시의 민감도(베타) 부호에 따라 요인의 부호가 달라질 수 있는 점 유의

결국 시장이 주가를 결정 짓는다

'시장의 설명력'과 같은 표현으로는 쉽게 와닿지 않을 수 있다. 그렇다면 추출한 민감도(베타)와 요인 시계열을 활용해 주가를 재구성해보자. 예시로 도표 9~12 번은 HMM과 현대차 주가를 각각 시장 요인 단일 또는 상위 6개 요인으로 재추정하여 실제 주가와 비교한 결과다.

놀랍지 않은가? 많은 분석자들이 시장의 방향성을 예측하고자 노력하는 본질적인 이유가 여기에 있다. 결국 시장의 방향성을 예측하는 것이 곧 개별 종목의 방향성을 예측하는 것과 다름없다고 봐도 과언이 아닌 것이다. 만약에 시장을 예측할 수 있다면 방향성에 따라 포트폴리오의 시장 베타를 올리거나 낮춤으로써 아웃퍼폼하는 것이 가장 유효한 방법이다. 너무도 당연한 말이지만, 팩터 모델적 관점에서 이 논리를 이끌어와야 새로운 논리로 확장시킬 수 있다.

도표 9. 시장 요인만으로 재구성한 현대차 주가



자료: 유진투자증권

도표 10. 상위 6개 요인을 활용해 재구성한 현대차 주가



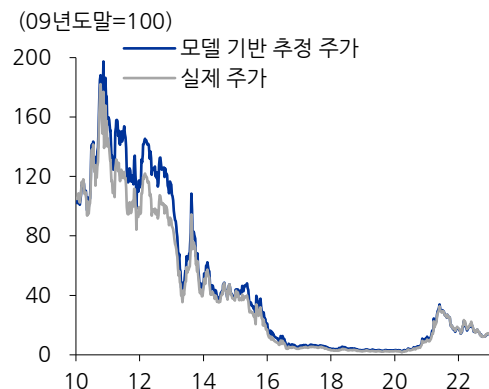
자료: 유진투자증권

도표 11. 시장 요인만으로 재구성한 HMM 주가



자료: 유진투자증권

도표 12. 상위 6개 요인으로 재구성한 HMM 주가



자료: 유진투자증권

3. 가치(Value) 팩터의 등장

파마-프렌치의 3 요인 모형

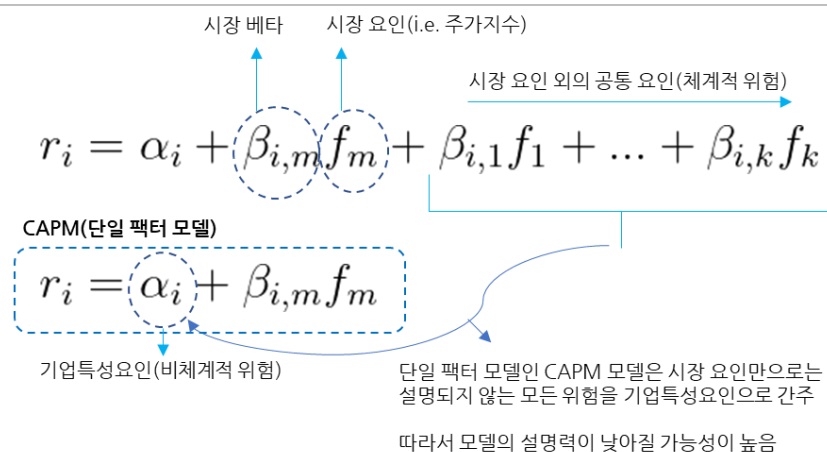
팩터 모델 관점에서의 첫 번째 시도는 1961 년 CAPM(자산가격결정모형)으로 볼 수 있다. CAPM 은 시장 요인만을 활용하여 자산의 수익을 설명하고자 하였다. 시장의 설명력을 보면 60 년대에 등장한 모델이 왜 아직까지 현대 재무학에서 활용되는 지 알 수 있다.

만약 통계적으로 추출한 요인들이 주가를 거의 완벽히 설명한다면, 그 요인들을 사용하면 주가를 완벽하게 예측할 수 있는 걸까? 그렇지 않다. 일단, 통계적으로 추출한 요인은 시장 요인 외에는 명확히 정의할 수도 없다. 이에 관련한 자세한 내용은 <Appendix. 잠재 요인을 추출하는 요인 모형과 그 한계>를 참고하길 바란다.

CAPM 이후 시장 요인 외에 주가 수익률에 영향을 끼치는 다른 펀더멘털적 요인을 찾아보려 하는 시도로 등장한 것이 바로 그 유명한 파마-프렌치의 3 요인 모형(1992 년)이다. 3 요인 모형이 시장 요인 외에 더하고자 한 2 가지 팩터는 '가치주 효과'와 '소형주 효과'다. 각 효과는 주식이 장부가치 대비 시장가치가 낮을수록(저평가 되있을수록), 또는 시가총액이 작을수록, 주식 수익률이 유의미하게 높다는 것을 의미한다.

3 요인 모형은 이러한 효과를 계량화하여 시계열 팩터로 활용하기 위해 시장 요인처럼 지수화한다. 예를 들어 가치주 효과는 B/P(P/B 의 역수)가 중앙값보다 높은 종목을 롱, 작은 종목을 숏하는 포트폴리오의 누적 수익으로 팩터를 구성하여 시장 중립적인 지수화된 팩터를 구성한다. 만약 이 팩터에 대한 주가 수익률의 민감도(베타)가 양(+)의 부호에 통계적으로 유의미하다면, 가치주 효과는 존재하는 것이다.

도표 13. 시장 요인 외에 새로운 팩터를 찾고자 하는 시도, 3 요인 모형



자료: 유진투자증권

3 요인 모형의 문제점: 양(+)의 프리미엄의 상실

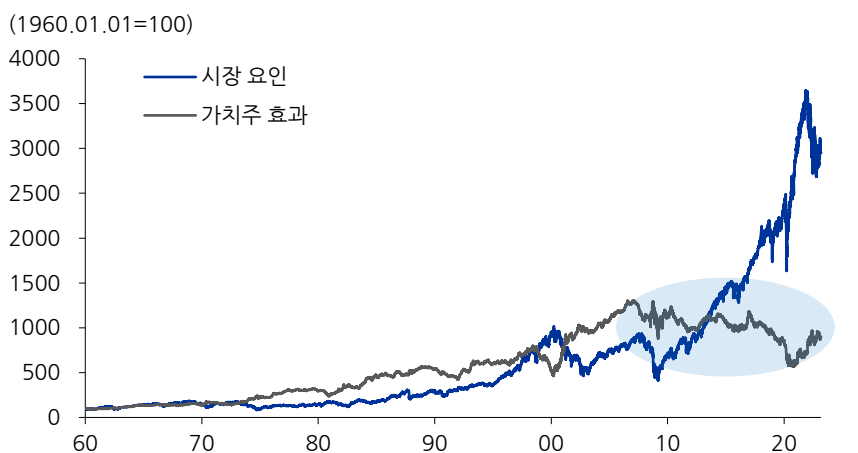
지속적인 알파가 없는 팩터는 미래에 대한 예측력이 없다

파마-프렌치 모형이 등장하고 저평가주에 투자하는 '가치주'라는 새로운 자산 수익의 동인이 등장하면서 가치주 투자가 각광을 받게 된다. 하지만 치명적인 문제가 발생한다. 파마-프렌치 두 교수의 논문이 발표되고 나서도 꾸준히 지속되던 가치주 효과가 어느 순간부터 전혀 작동하지 않게 된 것이다(도표 14 번). 가치주 효과 곡선은 시장 중립적인 통속 포트폴리오를 나타내기 때문에 우하향하는 곡선은 가치주에 투자할 시 시장(벤치마크) 대비 오히려 언더퍼폼하게 됨을 의미한다.

3 요인 모형이라는 멀티 팩터 모델이 가치가 있었던 이유는 가치/소형주 팩터가 장기적으로 우상향할 것, 즉 장기적으로 양(+)의 프리미엄, 즉 주가 수익률에 대해 양의 민감도를 보장할 것이라는 믿음이 있었기 때문이다. 만약 팩터가 지속적으로 양(+)의 프리미엄을 추구할 수 있다면 가치주에 대한 익스포저를 단순히 높임으로써 시장 대비 아웃퍼폼하는 것이 가능하다. 반대로 지속적으로 알파를 유지할 수 없다면, 팩터는 향후 주가의 수익에 대한 어떠한 정보도 내포하지 않는다.

그렇다면 시장 팩터는 안전할까? 투자의 대가 워렌 버핏은 비전문투자자에게 S&P500 지수 추종 ETF에 투자할 것을 권하는 것으로 유명하다. 이는 우리가 이전에 확인했듯이 시장 요인이 개별 종목의 수익률을 거의 결정짓고(우리는 흔히 개인의 능력으로 알파를 얻어냈다고 착각하지만), 시장 외의 추가적인 알파를 만들어내는 팩터를 찾아내기 어렵기 때문이다. 다만 지수 추종 역시 시장 요인이 우상향할 것이라는 믿음에 기반한 것이다. 과연 그럴까? 경험적으로 우리는 알고 있다. 장기적 우상향이 가능하더라도, 긴 하락장에서의 인내가 요구된다는 것을 말이다.

도표 14. 장기적 양(+)의 프리미엄을 상실한 가치주 효과



자료: Kenneth R. French Data Library, 유진투자증권

4. 파마-프렌치 사고방식에서 벗어나야

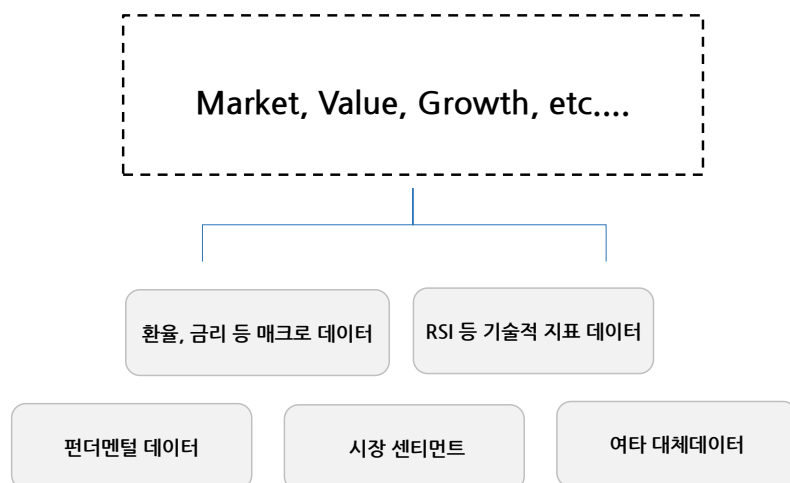
상상할 수 있는 모든 팩터를 감안해야 한다

시장과 기존 펀더멘털 팩터가 우상향할 것이라는 믿음이 깨진 이상, 단순히 기존 팩터에 대한 베타 익스포저를 높이는 전략은 의미가 없다. 우리가 해야 할 것은 시장과 여타 팩터의 움직임, 그 배후에 어떤 '진짜' 팩터들이 영향을 끼치고 있는지를 찾아서(도표 15 번) 시장/가치 등의 팩터 움직임을 예측해야 한다. 그리고 그 과정에서는 상상할 수 있는 모든 팩터를 감안해야 하고, 경제적 함의가 아예 없는 팩터가 아닌 이상 상상에 제한을 둘 필요는 없다.

기술적 지표도 유효한 팩터다

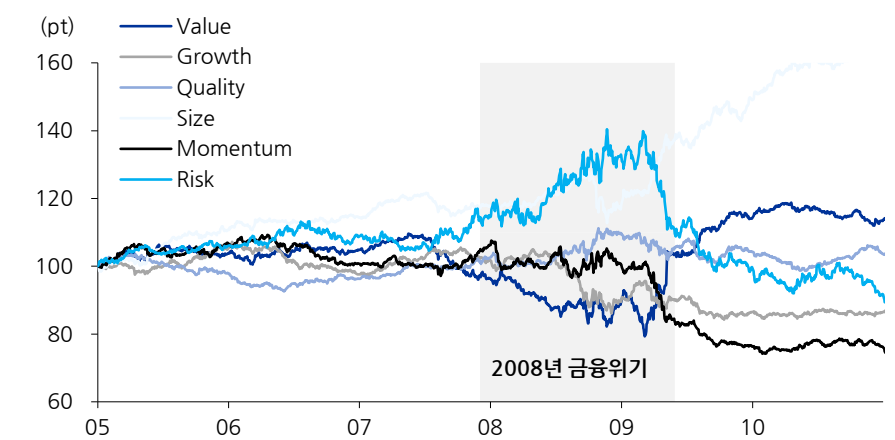
예를 들어 많은 사람들이 실적과 같은 펀더멘털 팩터, 또는 금리 환율 등과 같은 매크로 팩터는 용인하지만 이동평균선, RSI(상대강도지수) 등의 기술적 지표는 무시한다. 경제적 함의가 없다는 것이다. 물론 판단은 개인의 자유이지만, 금융공학 학계에서도 기술적 요인 중 하나인 모멘텀 효과는 주요 팩터로 인정받고 있다. 칼하트의 4 요인 모형(1997)은 3 요인 모형에 모멘텀 팩터를 더함으로써 팩터가 가지는 경제적 함의를 군집현상 등의 행동경제학적 영역까지 끌어온다. 즉, 이미 자산의 가격에 정보가 반영된 이후에도 시장의 과잉반응으로 인한 알파 추구가 가능하다는 것이다.

도표 15. 기존에 활용하던 팩터에 영향을 끼치는 '진짜' 팩터를 찾아라



자료: 유진투자증권

도표 16. 매크로 팩터가 급변하는 경우, 스타일 팩터의 차별화는 두드러진다



자료: Bloomberg, 유진투자증권

주: 미국증시 스타일 지수(롱숏), 2005.01.01=100을 기준으로 지수화

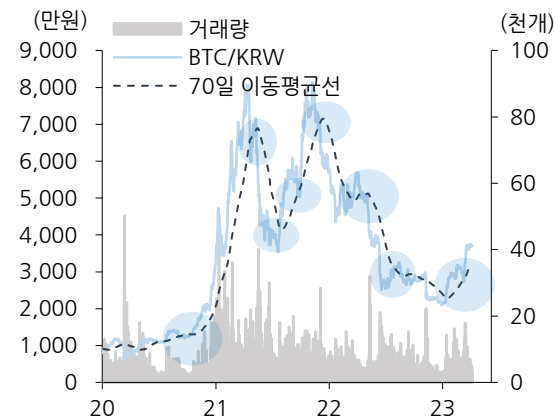
5. 기술적 지표를 감안해야 하는 이유

차티스트(chartist)의 냉혹한 승부 세계: 코인 시장 예시

기술적 분석을 간과하는 것은 성공한 트레이더들의 지속적인 성과를 단순히 운으로 치부하는 셈이다. 조금 자극적일 수 있지만 코인 시장 예시를 한번 들어보자. 코인 시장은 상대적으로 주식 시장 대비 본질적인 내재가치를 측정하기 어려운 시장이기 때문이다. 따라서 코인 시장에서는 인간의 선천적인 편향(cognitive bias)을 읽는 능력, 즉 일관적인 행동 패턴 파악을 위한 기술적 지표를 잘 선정하는 능력이 요구된다.

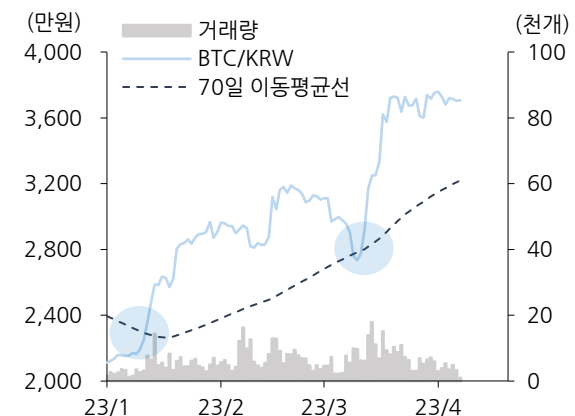
가장 간단하지만, 강력한 예시를 들어보겠다. 비트코인과 이더리움의 경우 70 일선을 기준으로 상/하향 돌파시 매수/매도하는 전략이 항상 유효했다(도표 17~20 번). 매우 간단하지만 장기적으로 작동했던 전략으로, 최근 3 월중에도 하향 돌파한 직후 바로 상향 돌파하자 재차 큰 폭 상승하는 흐름을 나타냈다.

도표 17. 비트코인 가격 추이: 장기



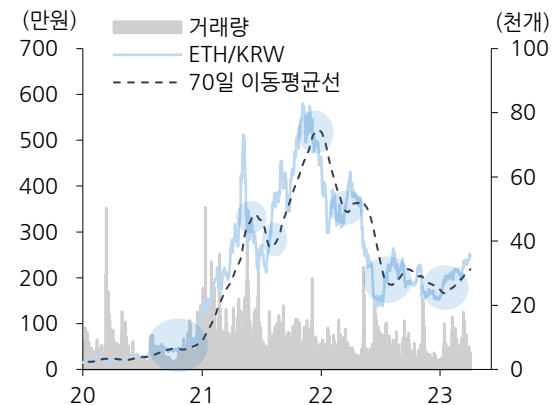
자료: 업비트, 유진투자증권

도표 18. 비트코인 가격 추이: 연초 이후



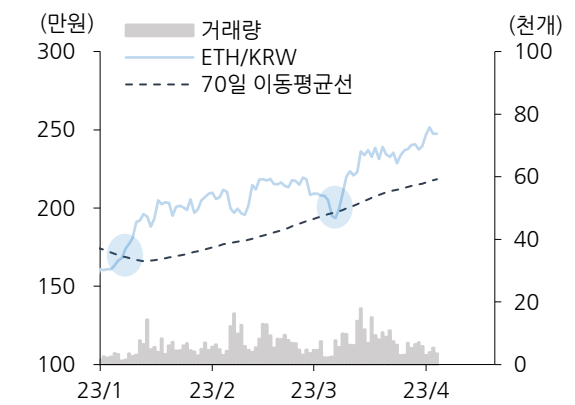
자료: 업비트, 유진투자증권

도표 19. 이더리움 가격 추이: 장기



자료: 업비트, 유진투자증권

도표 20. 이더리움 가격 추이: 연초 이후



자료: 업비트, 유진투자증권

변동성이 높은 자산은 모멘텀으로 대해줘야 한다

70 일선 돌파 전략은 추세를 하향 돌파하면 매도하기 때문에 트레일링 손절 방식(trailing stop loss)에 가깝다. 트레일링 손절은 진입가 대비 손절 라인을 설정하는 것이 아닌, 매수 이후의 고점 가격 대비로 손절 가격을 설정하는 것을 의미한다. 어깨에서 사서 어깨에서 파는 추세 전략이라고 볼 수 있다.

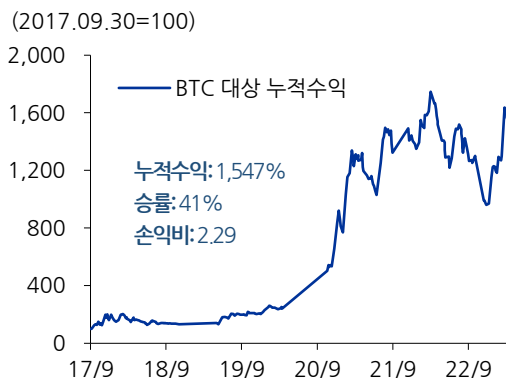
어깨에서 사서 무릎에서 팔다?

반대로 추세(모멘텀) 전략을 어깨에서 사서 무릎에서 파는 방식으로도 시도할 수 있다. 언뜻 듣기에는 전혀 말도 되지 않는 전략일 수 있다. 보통의 경우 트레이딩의 격언은 무릎에서 사서 어깨에서 파는 것이 아닌가? 과매수(어깨)에서 매수하여 과매도(무릎)에서 수익 실현하는 방식은 트레일링 손절을 하지 않고 추세적 상승 구간의 모든 이익을 향유하고자 하는 전략이다. 물론 매수가 대비 손절 라인은 설정해야 한다. 이럴 경우, 틀린 시그널이 더 많이 발생하기 때문에 손절로 인한 잦은 손실을 보게 되지만, 시그널이 맞게 되면 막대한 이익을 실현하게 된다(높은 손익비). 실제로 어깨에서 사서 무릎에서 파는 방식의 추세 전략을 이용하면 매수 후 보유(buy and hold)와는 '확연히' 다른 자산 흐름의 양상을 목격할 수 있다.

실제 추세 전략 예시

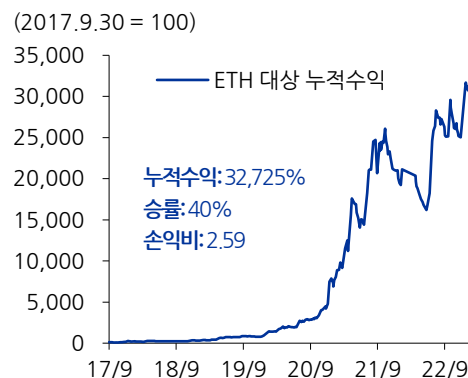
예를 들어 도표 21, 22 번은 필자가 사용하는 추세 전략(과매수 매수/과매도 매도/손절 라인 진입가 대비 설정)을 사용했을 때의 누적수익 추이를 나타낸다. 비트코인, 이더리움 각각 누적수익 1,547%, 32,725%를 기록한다. 당연하게도 승률은 40%대로 낮다. 절반이 안 되는 승률이지만 손익비가 1:2 이상이다. 즉, 손실 대비 이익이 두 배 이상이라는 것이다. 또한 기본적으로 손절을 통한 위기 관리가 병행되기 때문에 단순 매수 후 보유 전략 대비 오히려 하락 폭이 크게 나타나지 않는다.

도표 21. 비트코인 대상 추세 전략 누적수익 추이



자료: 업비트, 유진투자증권
주: 매매 1 회시 회전율 100%, 거래비용 1% 이상 가정

도표 22. 이더리움 대상 추세 전략 누적수익 추이



자료: 업비트, 유진투자증권
주: 매매 1 회시 회전율 100%, 거래비용 1% 이상 가정

6. 퀀트의 기초 덕목: 표본 외 검증

'과거의 나'의 전략, 좋은 성과로 이어졌을까?

퀀트(Quant)는 '계량적인'이란 뜻의 영어단어 'quantitative'에서 파생됐다. 이에 따라 퀀트는 일반적으로는 수치를 통한 정량적인 분석을 수행하는 금융분석가를 의미한다. 하지만 필자는 퀀트의 핵심이 '계량적인 분석' 그 자체는 결코 아니라고 생각한다. 퀀트의 핵심은 바로 과거 데이터로 최대한 보수적인 검증을 하고자 하는 시도이며, 수치 분석은 그러한 검증 과정에 있어서의 수단에 불과하다.

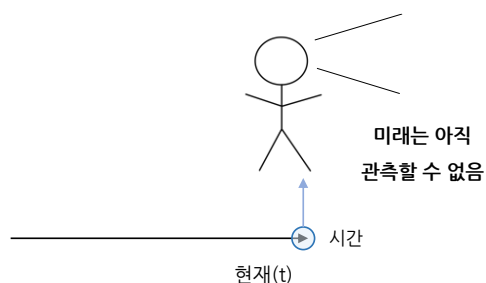
만약 활용 가능한 모든 데이터에 최적화된 팩터 모델을 만든다면, 모델을 검증할 수 있는 방법은 실전밖에 없다(도표 23 번). 하지만 분석자는 관측되지 않은 미래 데이터를 미리 알 수 없기 때문에 사전적으로 검증할 수 있는 방법론이 필요하다.

모델의 강건함을 검증하는 최소한의 방법론: 표본 외 구간 검증하기

따라서 특정 팩터 모델을 활용해 전략을 만들 때는 반드시 '과거의' 나는 어떤 결정을 내렸을지를 상상해봐야 한다(도표 24 번). 즉, 과거의 내가 가용 가능한 데이터만을 활용했을 때의 최선의 결정이 이후 좋은 성과로 이어졌는지를 보수적으로 검증해보는 것이다. 가장 기본적인 최소한의 방법론이 과거 구간을 나눠 모델 성능을 검증하는 표본 외 구간 검증이다(도표 25 번).

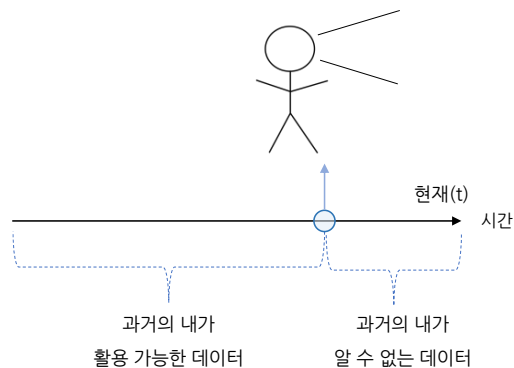
표본 외 검증 결과에서 모델의 강건함이 어느 정도 입증되더라도 모델이 미래에 준수한 성능을 보일지는 장담할 수 없다. 현실 세계에서는 투자자들의 행동 패턴이 바뀌면서 새로운 팩터가 나타나기도 하고, 혹은 시장 참여자들이 기존의 패턴을 인지하고 원래 존재했던 팩터가 수명을 다하며 사라지기도 한다(가치주 효과의 부진을 상기시켜보자). 따라서 팩터 모델의 유효성에 대한 검증은 끊임없이 이뤄져야 한다.

도표 23. 현재 시점에서 미래의 데이터는 관측할 수 없다



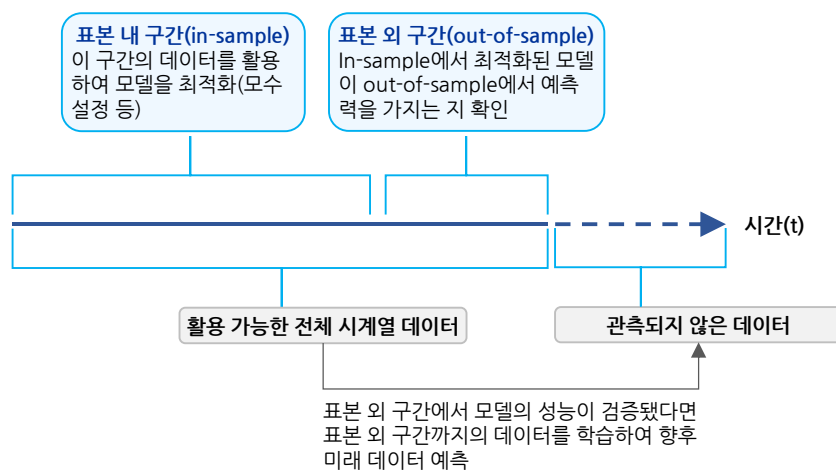
자료: 유진투자증권

도표 24. '과거의 나의 결정은 좋은 성과로 이어졌을까?'



자료: 유진투자증권

도표 25. 사전적 모델 성능 확인을 위한 표본 외 검증(out-of-sample testing)



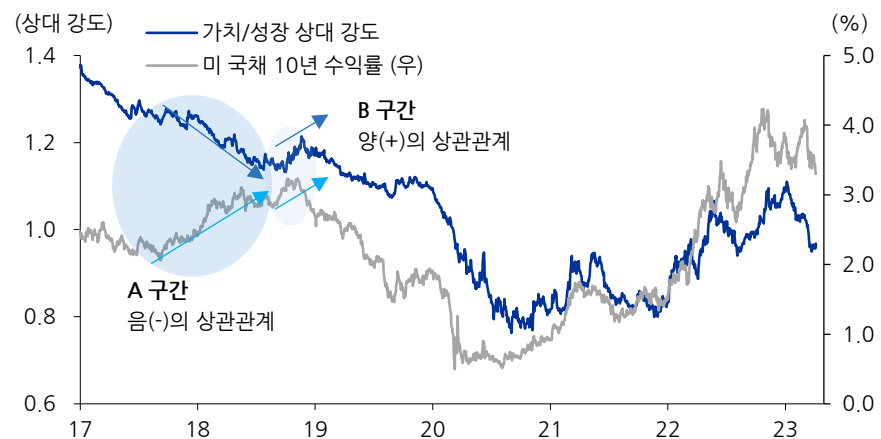
자료: 유진투자증권

끊임없는 검증이 필요한 이유: 가치/성장 상대 강도 예시

가치/성장 스타일 상대 강도와 국채 수익률 추이는 매크로 팩터 추정을 통한 수익률 예측 가능성을 보여준다(도표 26 번). 이론적으로 높은 할인율은 밸류에이션이 높은(미래가치를 더 크게 반영한) 종목의 주가에 더 큰 부담으로 작용하게 된다. 따라서 금리가 상승하는 구간에서는 가치주(저평가) 효과를 추구하는 가치 스타일이 성장주(높은 실적 성장성 반영, 대체로 높은 밸류에이션) 효과를 추구하는 성장 스타일 대비 아웃퍼폼 할 가능성이 높다. 즉, 이론적으로 금리라는 팩터에 대해 가치/성장 스타일 팩터 상대 강도는 양(+)의 민감도를 가지게 된다. 경제적 함의를 내포하게 되는 것이다.

실제 팩터 모델을 만든다면, 이러한 양(+)의 민감도가 통계적으로 유의미한 지에 따라 변수 선정 여부가 결정된다. 그리고 그 민감도의 지속성을 확인해야 한다. 예를 들어 도표의 A 구간에서 상대 강도는 금리에 대해 음(-)의 민감도를 가지고 있다. 분석자가 음(-)의 민감도를 기반으로 B 구간에서 금리가 상승할 것으로 추정했다면 모델은 B 구간에서 상대 강도가 하락할 것이라고 추정했을 것이다. 실제 B 구간에서 상대 강도는 금리에 대해 양(+)의 민감도를 보였기 때문에 금리와 동행하여 상승했다. 즉, 구간에 따라 민감도(상수)가 변함으로써 팩터(금리) 방향성에 대한 추정이 맞았음에도 불구하고 모델의 예측이 어긋나게 된 것이다.

도표 26. 스타일 지수 상대 강도(가치 vs 성장)와 국채 수익률 추이



자료: Refinitiv, 유진투자증권

분석자는 통계적 상관관계에만 의존하면 안 되며, 경제적 함의와 통계적 함의가 일치하더라도 민감도의 지속성에 대해서 끊임없이 검증을 해야 한다. 민감도와 같은 모델의 변수는 특정 구간에 적합하여 나오는 값으로 파라미터(parameter)라고 표현한다. 파라미터는 특정 구간에 적합하여 나오기 때문에 실제 미래 시점의 예측 전, 위의 경우처럼 과거 구간을 나눠 민감도가 지속성을 가지는 지 여부를 확인하여 모델의 강건함을 검증해야 한다.

7. 실전! 마켓 타이밍 모델 설계하기

이제 우리는 팩터 모델적 사고를 갖췄고, 꾸준히 지속되는 알파가 보장된 팩터가 없음을 이해하였다. 이는 결국 우리가 시장/가치 등 주요 팩터의 움직임을 예측해야함을 의미한다. 그 과정에서 당연히 매크로/펀더멘털/기술적 팩터 등 모든 유의미한 팩터를 동원해야한다.

기본적으로 모든 모델링 방식은 목표 함수에 대한 최적화라는 관점에서 기계학습(머신러닝)이라고 볼 수 있다. 머신러닝이라고 하면 일반적으로 딥러닝을 떠올리는 경우가 많은데, 딥러닝은 머신러닝 알고리즘 중에서도 인공신경망 구조를 채택하는 경우를 의미할 뿐이다.

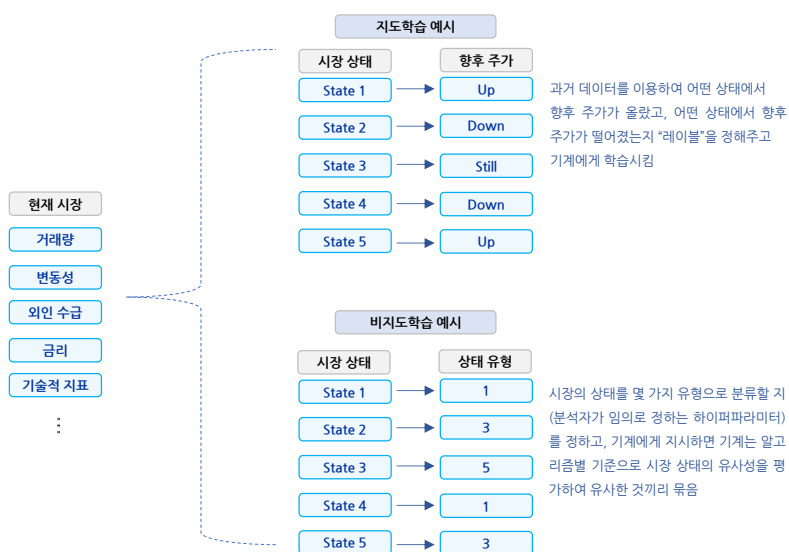
물론 최근 대형언어모델인 Chat GPT 를 필두로, 딥러닝의 높은 성능이 부각되고 있는 것은 사실이다. 하지만 필자는 금융 시계열 분석에 있어서는 딥러닝이 압도적 우위에 있다고는 생각하지 않는다. 딥러닝을 활용한 금융 시계열 분석 방법 예시로는 <딥러닝을 통한 주식 스타일 배분: 11 월 포트폴리오 제시, 22.10.31> 자료를 참고하길 바란다. 필자는 비지도학습 계열 머신러닝 알고리즘을 채택하여 모델을 설계하였다.

비지도학습이란?

기계학습은 크게 지도학습과 비지도학습의 형태로 나뉘어진다. 주가 예측 모델 설계 관점에서 설명하자면, 지도학습은 과거 데이터를 이용하여 어떤 상태에서 향후 주가가 올랐고, 어떤 상태에서는 향후 주가가 떨어졌는지를 기계에게 학습시키는 방식이다. 학습이 완료된 후, 기계에게 최근까지의 상태를 기준으로 향후 주가가 어떻게 될지를 물어보면, 기계는 그동안 학습된 데이터와 그에 기반하여 산출된 파라미터를 기반으로 향후 주가의 방향을 제시하게 된다.

반대로 비지도학습은 상황에 따른 정답을 알려주지는 않고 기계에게 유형 분류를 지시하는 방식이다. 시장의 상태를 몇 가지 유형으로 분류할 것을 기계에게 지시하면, 기계는 알고리즘별로 나름의 기준을 통해 상태의 유사성을 평가하여 유사한 것끼리 묶게된다(clustering). 학습이 완료된 후 기계에게 최근의 상태는 어느 상태와 유사한지 물어보면 기계는 학습 결과를 바탕으로 어느 유형에 속하는지를 알려준다.

도표 27. 기계학습의 두 가지 유형 비교



자료: 유진투자증권

모델 방법론

예측 대상

마켓 타이밍 모델의 예측 대상은 다음과 같다.

1) 가치/성장 상대 강도, 2) 코스피 지수, 3) 주요 코인(비트코인 및 이더리움)

가치/성장 상대 강도와 코스피 지수(일봉 캔들)는 FnGuide 데이터를 활용했다. 가치/성장 상대 강도 산출을 위한 지수로는 WMI500 스타일(가치, 성장) 지수를 채택했다(도표 28 번). WMI500은 코스피+코스닥 상장사 중 시총 상위 80% 내 기업을 커버하기 때문에 범용성이 높고, 유니버스 관리에 용이하다. 다만 FnGuide에서 스타일 지수의 OHLC 데이터는 제공하지 않기 때문에 종가 지수만 활용 가능하다.

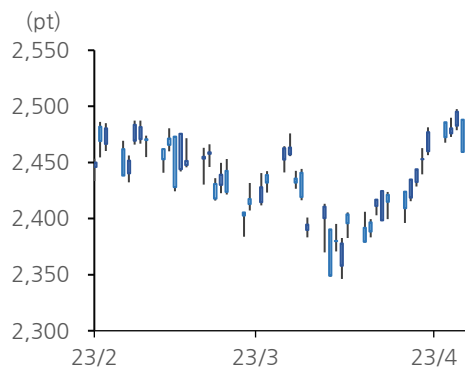
주요 코인(비트코인/이더리움) 가격(일봉 캔들)은 업비트 데이터(원화 기준)를 활용했다. 코인은 24시간 거래가 가능하다. 따라서 업비트 기준 일봉 캔들은 매일 UTC 기준 0시(한국 시간 오전 9시)에 업데이트되며, 오전 9시부터 24시간 동안의 시가/고가/저가/종가를 표시한다.

도표 28. 국내 증시 가치/성장 상대 강도 추이 (WMI500 스타일 지수 기준)



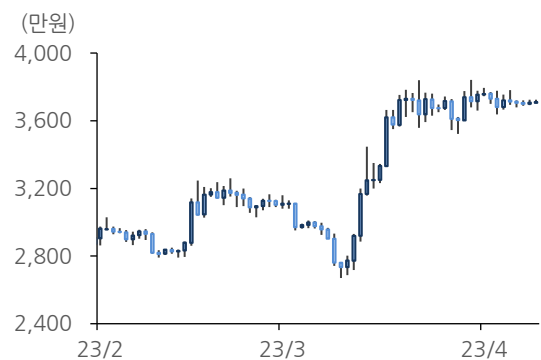
자료: Quantwise, 유진투자증권

도표 29. 코스피 지수 추이(일봉, OHLC)



자료: Quantise, 유진투자증권

도표 30. 비트코인 가격 추이(일봉, OHLC)



자료: 업비트, 유진투자증권

모델 학습 과정

향후 진행되는 모델 학습은 다음과 같은 과정을 거친다.

1. 독립 변수 선정: 예측 대상(종속 변수)별로 유의미한 독립 변수 선정
2. 표본 내/외 구간 구분: 충분한 표본 외 구간이 확보되도록 함
3. 데이터 표준화: 표본 외 구간 정규화시 표본 내 구간의 통계량 활용
4. 클러스터링: 표본 내 구간에서 비지도학습을 통해 시장 유형 분류
5. 우호 국면 선정: 표본 내 구간의 유형별 수익률 분포 확인 후 Long/Short 국면 선정
6. 백테스팅 진행: 국면에 따른 표본 내/외 구간 내 투자 백테스팅 진행 후 성과 확인

모델 예측 주기

코스피 지수/주요 코인의 예측 주기는 일간(daily)으로 설정한다. 시스템 리스크에 따른 급락 가능성을 대비하기 위해서는 일간이 가장 적합하다고 판단한다. 일간 이하의 주기를 활용할 시 시계열 데이터의 노이즈 자체도 커질 수밖에 없고, 매매 시그널이 필연적으로 자주 발생하게 된다. 전략을 설계할 때는 반드시 매매 빈도 및 거래비용(슬리피지 포함)을 고려하여 전략의 경제성을 평가해야 한다.

가치/성장 스타일 상대 강도 예측 주기는 주간(weekly)으로 설정한다. 가치/성장 스타일 상대 강도는 "시장 중립적(market neutral)"인 정보다. 즉, 시장 요인 외의 추가적인 알파를 추구하기 위한 팩터라고 볼 수 있으며 단기 등락보다는 상대적으로 더 긴 흐름 변화에 주목할 필요가 있다.

거래 비용

매매 1 회당 0.5%로 설정한다. 다만 시그널에 따라 기존 포지션 정리와 새로운 포지션 진입이 동시에 이뤄지는 구조로, 사실상 시그널 변경시마다 거래비용 1%를 가정하는 셈이다.

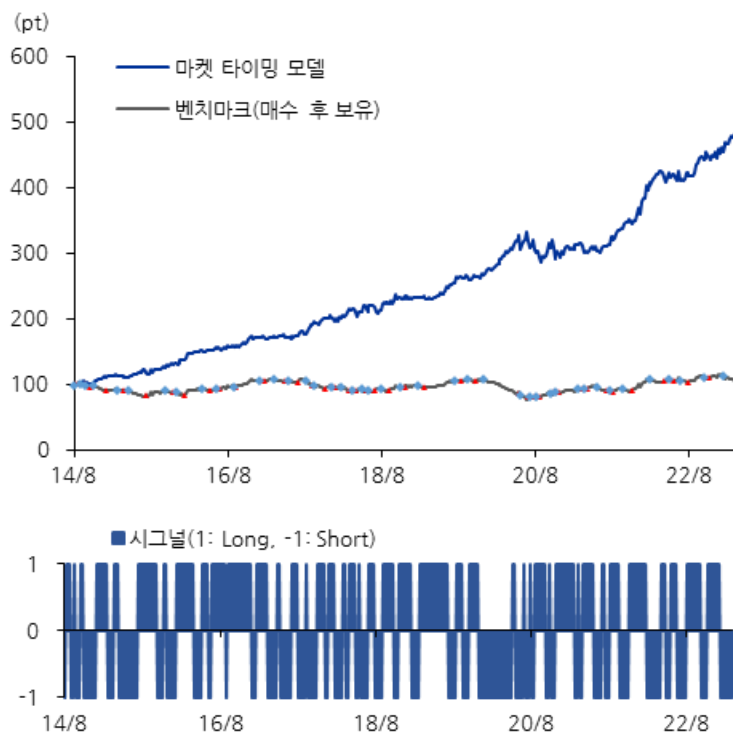
Backtesting Result 1. 가치/성장 상대 강도

예측 주기: 주간(weekly)

매매 시점: 주간 마지막 영업일 종가 기준 매매 시그널 산출, 차주 첫 영업일 종가에 포지션 진입/변경

현재 시그널: 2023 년 1 월 30 일부터 Short 포지션 유지 중(가치 Short, 성장 Long)

도표 31. 마켓 타이밍 모델 누적수익 추이



자료: 유진투자증권

도표 32. 벤치마크 및 전략 성과 지표

성과지표	Market Timing Model	Benchmark
누적 수익	380.6%	5.8%
CAGR	19.9%	0.7%
Volatility	12.3%	12.6%
Sharpe Ratio	1.62	0.05

자료: 유진투자증권

도표 33. 표본 외 구간 매매 내역 (가치/성장 상대 강도, 최근 30 회)

회차	진입일	정리일	포지션	지속 기간 (주)	수익률 (%, 거래비용 차감)
1	2019-12-29	2020-06-01	short	23	24.9
2	2020-06-01	2020-06-15	long	2	-2.9
3	2020-06-15	2020-07-20	short	5	2.2
4	2020-07-20	2020-07-27	long	1	-1.3
5	2020-07-27	2020-08-17	short	3	-0.9
6	2020-08-17	2020-08-24	long	1	-2.7
7	2020-08-24	2020-09-14	short	3	-2.9
8	2020-09-14	2020-11-02	long	7	7.3
9	2020-11-02	2020-11-16	short	2	-2.6
10	2020-11-16	2020-11-23	long	1	-1.8
11	2020-11-23	2020-12-14	short	3	1.0
12	2020-12-14	2021-03-15	long	13	5.7
13	2021-03-15	2021-04-05	short	3	-2.0
14	2021-04-05	2021-04-12	long	1	-1.8
15	2021-04-12	2021-04-26	short	2	-0.9
16	2021-04-26	2021-06-14	long	7	0.6
17	2021-06-14	2021-07-26	short	6	3.8
18	2021-07-26	2021-08-09	long	2	1.2
19	2021-08-09	2021-08-30	short	3	2.9
20	2021-08-30	2021-10-11	long	6	5.1
21	2021-10-11	2021-11-29	short	7	1.8
22	2021-11-29	2022-02-21	long	12	17.9
23	2022-02-21	2022-04-25	short	9	2.3
24	2022-04-25	2022-05-23	long	4	1.2
25	2022-05-23	2022-06-13	short	3	0.2
26	2022-06-13	2022-07-11	long	4	-1.2
27	2022-07-11	2022-08-22	short	6	1.3
28	2022-08-22	2022-10-31	long	10	6.5
29	2022-10-31	2022-11-28	short	4	0.7
30	2022-11-28	2023-01-30	long	9	2.4

자료: 유진투자증권

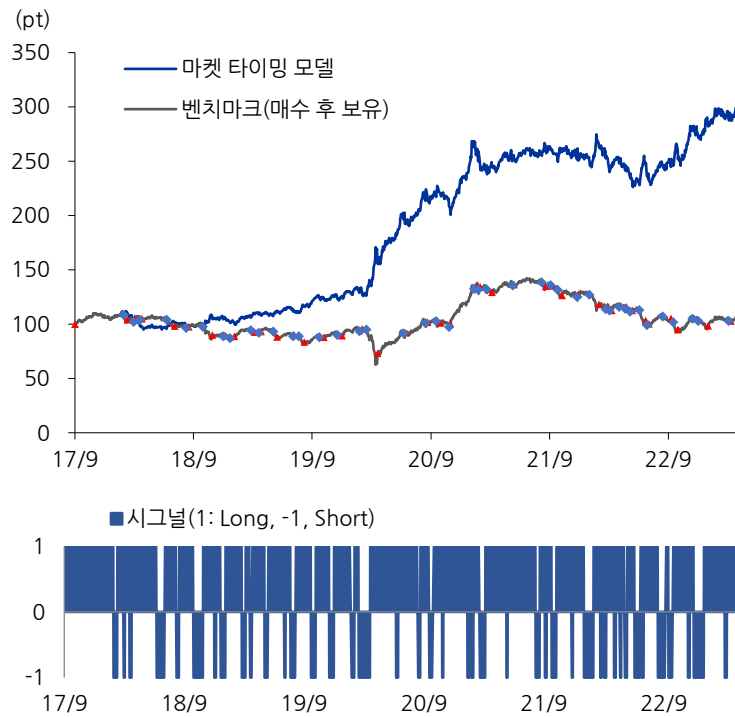
Backtesting Result 2. 코스피 지수

예측 주기: 일간(daily)

매매 시점: 일간 장 마감 이후 매매 시그널 산출, 다음 영업일 시가에 포지션 진입/변경

현재 시그널: 2023 년 3 월 20 일부터 Long 포지션 유지 중

도표 34. 마켓 타이밍 모델 누적수익 추이



자료: 유진투자증권

도표 35. 벤치마크 및 전략 성과 지표

성과지표	Market Timing Model	Benchmark
누적 수익	203.6%	8.1%
CAGR	20.3%	1.3%
Volatility	16.9%	18.4%
Sharpe Ratio	1.20	0.07

자료: 유진투자증권

도표 36. 표본 외 구간 매매 내역 (코스피 지수, 최근 30 회)

회차	진입일	정리일	포지션	지속 기간 (일)	수익률 (%, 거래비용 차감)
1	2021-09-15	2021-09-30	long	8	-2.9
2	2021-09-30	2021-10-14	short	8	2.9
3	2021-10-14	2021-11-30	long	33	-1.2
4	2021-11-30	2021-12-06	short	4	-0.7
5	2021-12-06	2022-01-06	long	21	-1.0
6	2022-01-06	2022-02-07	short	19	6.4
7	2022-02-07	2022-02-25	long	14	-2.6
8	2022-02-25	2022-03-03	short	3	-1.9
9	2022-03-03	2022-03-10	long	4	-2.5
10	2022-03-10	2022-03-16	short	4	0.4
11	2022-03-16	2022-04-08	long	17	2.1
12	2022-04-08	2022-04-14	short	4	-0.5
13	2022-04-14	2022-04-26	long	8	-1.7
14	2022-04-26	2022-05-02	short	4	0.2
15	2022-05-02	2022-05-11	long	5	-3.1
16	2022-05-11	2022-05-18	short	5	-1.8
17	2022-05-18	2022-06-09	long	14	-0.6
18	2022-06-09	2022-06-28	short	13	8.9
19	2022-06-28	2022-07-04	long	4	-3.9
20	2022-07-04	2022-07-08	short	4	-1.8
21	2022-07-08	2022-08-22	long	30	4.9
22	2022-08-22	2022-09-14	short	15	3.2
23	2022-09-14	2022-09-21	long	5	-1.6
24	2022-09-21	2022-10-05	short	9	4.6
25	2022-10-05	2022-11-21	long	32	8.8
26	2022-11-21	2022-11-25	short	4	0.2
27	2022-11-25	2022-12-07	long	8	-2.3
28	2022-12-07	2023-01-09	short	22	3.0
29	2023-01-09	2023-03-13	long	42	3.7
30	2023-03-13	2023-03-20	short	5	0.4

자료: 유진투자증권

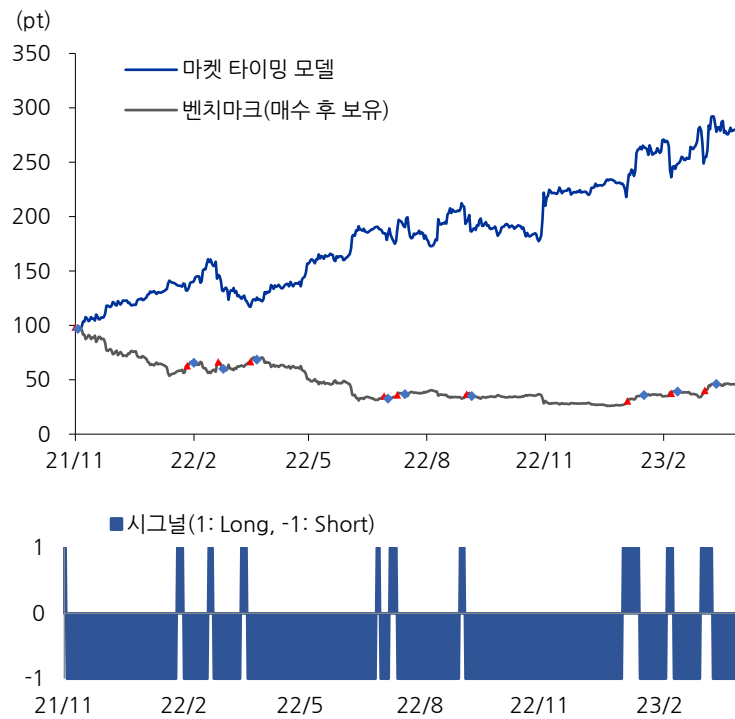
Backtesting Result 3. 비트코인

예측 주기: 일간(daily)

매매 시점: 오전 9 시 일봉 생성 시, 시그널 산출과 동시에 포지션 진입/변경

현재 시그널: 2023 년 3 월 23 일부터 Short 포지션 유지 중

도표 37. 마켓 타이밍 모델 누적수익 추이



자료: 유진투자증권

도표 38. 벤치마크 및 전략 성과 지표

성과지표	Market Timing Model	Benchmark
누적 수익	181.0%	-54.3%
CAGR	65.6%	-31.8%
Volatility	34.4%	43.3%
Sharpe Ratio	1.91	-0.73

자료: 유진투자증권

도표 39. 표본 외 구간 매매 내역 (비트코인, 최근 30 회)

회차	진입일	정리일	포지션	지속 기간 (일)	수익률 (%, 거래비용 차감)
1	2021-11-10	2021-11-13	long	3	-3.3
2	2021-11-13	2022-02-05	short	84	54.4
3	2022-02-05	2022-02-11	long	6	4.8
4	2022-02-11	2022-03-01	short	18	1.4
5	2022-03-01	2022-03-06	long	5	-7.1
6	2022-03-06	2022-03-26	short	20	-9.7
7	2022-03-26	2022-03-31	long	5	5.1
8	2022-03-31	2022-07-07	short	98	110.0
9	2022-07-07	2022-07-11	long	4	1.4
10	2022-07-11	2022-07-17	short	6	-1.9
11	2022-07-17	2022-07-24	long	7	6.6
12	2022-07-24	2022-07-30	short	6	-5.4
13	2022-07-30	2022-08-02	long	3	-1.8
14	2022-08-02	2022-09-10	short	39	5.1
15	2022-09-10	2022-09-15	long	5	-3.4
16	2022-09-15	2023-01-13	short	120	20.4
17	2023-01-13	2023-01-26	long	13	22.0
18	2023-01-26	2023-02-16	short	21	-7.8
19	2023-02-16	2023-02-21	long	5	2.2
20	2023-02-21	2023-03-14	short	21	0.7
21	2023-03-14	2023-03-23	long	9	14.8

자료: 유진투자증권

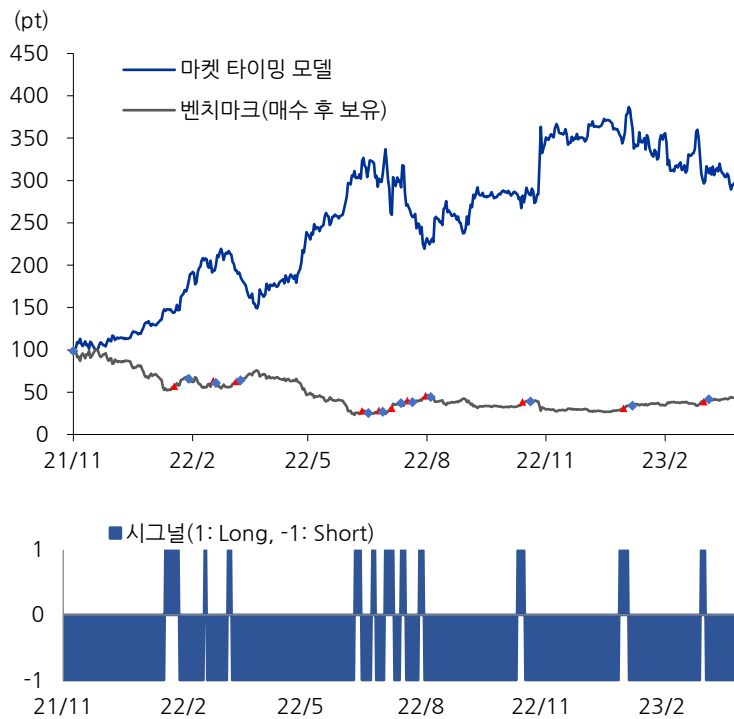
Backtesting Result 4. 이더리움

예측 주기: 일간(daily)

매매 시점: 오전 9 시 일봉 생성 시, 시그널 산출과 동시에 포지션 진입/변경

현재 시그널: 2023 년 3 월 19 일부터 Short 포지션 유지 중

도표 40. 마켓 타이밍 모델 누적수익 추이



자료: 유진투자증권

도표 41. 벤치마크 및 전략 성과 지표

성과지표	Market Timing Model	Benchmark
누적 수익	201.3%	-57.3%
CAGR	71.9%	-34.2%
Volatility	47.9%	58.2%
Sharpe Ratio	1.50	-0.59

자료: 유진투자증권

도표 42. 표본 외 구간 매매 내역 (이더리움, 최근 30 회)

회차	진입일	정리일	포지션	지속 기간 (일)	수익률 (%, 거래비용 차감)
1	2021-11-13	2022-01-30	short	78	76.7
2	2022-01-30	2022-02-10	long	11	21.7
3	2022-02-10	2022-03-01	short	19	10.8
4	2022-03-01	2022-03-03	long	2	1.5
5	2022-03-03	2022-03-19	short	16	0.3
6	2022-03-19	2022-03-22	long	3	-1.2
7	2022-03-22	2022-06-24	short	94	138.4
8	2022-06-24	2022-06-29	long	5	1.1
9	2022-06-29	2022-07-07	short	8	-3.6
10	2022-07-07	2022-07-10	long	3	2.2
11	2022-07-10	2022-07-17	short	7	-10.8
12	2022-07-17	2022-07-24	long	7	15.1
13	2022-07-24	2022-07-29	short	5	-9.2
14	2022-07-29	2022-08-02	long	4	-4.0
15	2022-08-02	2022-08-12	short	10	-13.2
16	2022-08-12	2022-08-16	long	4	2.6
17	2022-08-16	2022-10-26	short	71	23.6
18	2022-10-26	2022-11-01	long	6	6.5
19	2022-11-01	2023-01-12	short	72	27.0
20	2023-01-12	2023-01-19	long	7	8.2
21	2023-01-19	2023-03-15	short	55	-15.9
22	2023-03-15	2023-03-19	long	4	4.5

자료: 유진투자증권

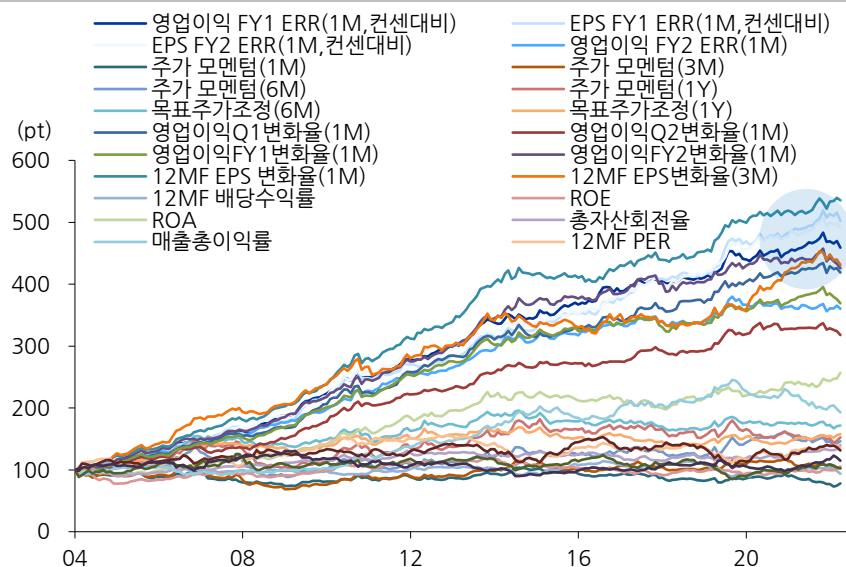
8. 현재는 성장주 우위 국면 : 실적 스크리닝 종목

왜 실적을 고려해야 하는가?

시장이나 주요 팩터에 대한 개별적 예측을 하더라도, 기초 펀더멘털에 대한 확신은 필요하다. 하지만 펀더멘털 스타일 팩터를 적극적으로 활용하려면 앞서 언급했듯이 해당 팩터가 우상향할 것이라는 강한 믿음이 기반돼야 한다. 국내 증시에서는 이런 관점에서 실적 모멘텀 팩터(이익조정비율, 실적 컨센서스 변화율 등)가 가장 유효한 것으로 판단된다. 실제로 켈리티, 가치, 주가 모멘텀 등 다른 스타일 팩터 대비 가장 높은 누적수익을 기록하고 있을뿐만 아니라, 다양한 경제 국면에서도 상대적으로 변동성이 적은 모습이 나타난다(도표 43 번).

실적 기준은 연도별, 분기별, 또는 현재 시점으로부터 고정된 구간(향후 12·24 개월, etc)으로 롤링하는 등 다양하게 설정하여 컨센서스 흐름을 관찰할 수 있는데, 가장 직관적인 기준은 시점에 따라 고정된 구간 내의 흐름을 보는 것이다. 연도·분기별 흐름은 시간이 흐름에 따라 주목도가 달라지기 때문이다. 예를 들어, 올해(FY1) 실적의 경우 연말쯤이 되면 거의 고점에 근접하게 되고, 주가에 큰 영향을 주지 않을 수 있다. 필자는 12 개월 선행 EPS의 단기(1 개월) 변화율을 실적 스크리닝 최우선 기준으로 선정한다.

도표 43. 국내 증시 누적수익 상위 팩터는 모두 실적 관련 변수 (이익조정, 컨센 변화율 위주)



자료: 유진투자증권
주: 월간리밸런싱, 개별팩터 상위 20% 롱, 하위 20% 숏 포트폴리오 누적수익

성장 + 실적 모멘텀 스크리닝 상위 종목

도표 44. 4/10 일 기준 스크리닝 (WMI500 편입 종목 대상)

티커	종목	업종 (W26 기준)	시가총액 (십억원)	상장된 시장	수익률(%)		P/E(배) (Fwd, 12M)	EPS 변화율(%) (Fwd 12M, 1 개월)
					1M	3M		
A251270	넷마블	소프트웨어	6,197	KOSPI	20.2	39.5	223.2	310.5
A196170	알테오젠	건강관리	2,612	KOSDAQ	35.6	42.6	42.0	178.0
A326030	SK 바이오팜	건강관리	4,902	KOSPI	-4.7	-11.7	-334.0	34.9
A010140	삼성중공업	조선	4,611	KOSPI	-1.0	4.8	21.8	18.6
A402340	SK 스퀘어	반도체	5,546	KOSPI	-6.0	15.1	-11.7	9.6
A263750	풀어비스	소프트웨어	2,958	KOSDAQ	8.6	8.2	27.4	9.0
A015760	한국전력	유틸리티	11,607	KOSPI	1.0	-7.3	-2.0	8.9
A086520	에코프로	화학	14,983	KOSDAQ	98.3	428.3	20.4	6.4
A066570	LG 전자	IT 가전	18,705	KOSPI	-0.1	26.0	10.2	4.6
A066970	엘앤에프	IT 가전	11,400	KOSDAQ	29.2	76.8	30.1	4.4
A373220	LG 에너지솔루션	IT 가전	135,720	KOSPI	2.5	30.6	76.0	4.3
A012450	한화에어로스페이스	상사,자본재	5,615	KOSPI	14.0	65.5	16.1	3.9
A003670	포스코퓨처엠	화학	22,929	KOSPI	13.6	57.0	68.7	3.8
A068270	셀트리온	건강관리	22,461	KOSPI	2.6	-2.5	34.5	3.2
A064350	현대로템	기계	2,898	KOSPI	4.9	1.7	16.5	3.2
A008770	호텔신라	호텔,레저서비스	3,222	KOSPI	6.1	-3.1	23.0	2.4
A128940	한미약품	건강관리	3,882	KOSPI	16.2	16.8	32.7	2.0
A006400	삼성 SDI	IT 가전	50,748	KOSPI	-6.9	25.9	22.9	1.7
A000100	유한양행	건강관리	4,123	KOSPI	0.9	-1.7	33.6	1.5
A035420	NAVER	소프트웨어	31,727	KOSPI	-8.3	4.5	26.2	1.4
A352820	하이브	미디어,교육	8,974	KOSPI	15.1	29.2	42.8	1.4
A271560	오리온	필수소비재	5,385	KOSPI	8.9	13.0	14.2	1.3
A035720	카카오	소프트웨어	26,115	KOSPI	-4.7	2.5	46.6	1.3
A383220	F&F	화장품,의류,완구	5,252	KOSPI	-2.1	8.8	9.9	1.3
A207940	삼성바이오로직스	건강관리	56,655	KOSPI	0.3	-1.5	67.4	1.1
A047810	한국항공우주	상사,자본재	4,913	KOSPI	12.0	14.2	20.2	1.1
A028050	삼성엔지니어링	건설,건축관련	5,821	KOSPI	9.6	25.9	9.9	1.0
A247540	에코프로비엠	IT 가전	25,184	KOSDAQ	25.3	171.9	55.6	0.6
A034020	두산에너지빌리티	기계	10,743	KOSPI	-0.1	8.6	23.3	0.1
A091990	셀트리온헬스케어	건강관리	10,002	KOSDAQ	0.6	10.1	43.9	-0.3

자료: Quantwise, 유진투자증권

주: WMI500 Large 순수성장 편입 종목 대상

Appendix. 잠재 요인을 추출하는 요인 모형과 그 한계

최근 퀀트 애널리스트를 포함해 많은 분석자들이 PCA(Principal Component Analysis, 주성분 분석) 또는 Autoencoder 등 딥러닝 방식을 활용한 변수 추출을 통해 차원의 저주와 다중공선성 문제를 해결하고 있다. PCA는 고차원의 데이터를 저차원의 데이터로 환원시키는 통계적 기법으로, 이로부터 추출된 주성분(principal components)은 독립변수의 분산을 가장 잘 설명하며 서로 직교한다(독립적이다).

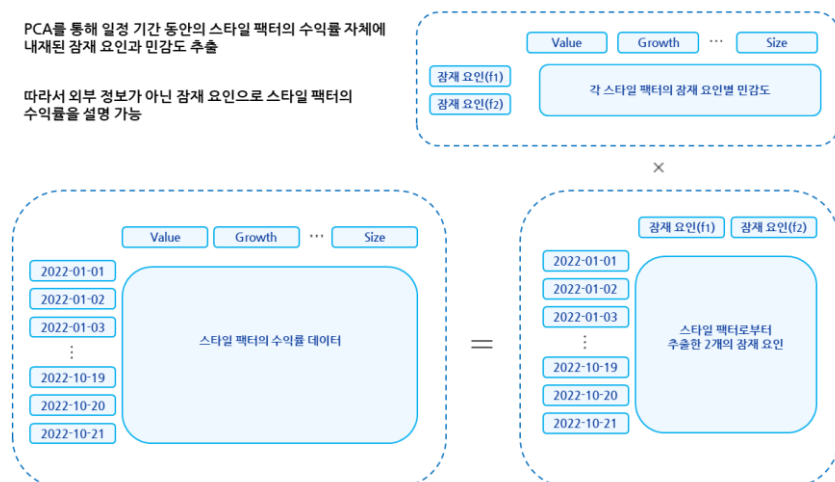
차원 축소라는 개념이 어려워보일 수 있지만, 차원 축소와 팩터 모델 관점의 요인 모형은 동일한 개념이다. "4개의 차원(변수)을 2개로 축소한다"와 "4개의 변수를 2개의 공통 요인(팩터)로 설명한다"는 동일한 개념이다.

PCA로 축소된 공통 요인은 여러 팩터에 내재된 잠재 요인

특히 PCA를 통해서 추출한 공통 요인은 서로 독립적이기 때문에 차원의 저주뿐만 아니라 팩터 간의 높은 상관관계(다중공선성) 또한 해결할 수 있다. 여러 개의 변수에서 설명력이 높은 변수를 추출해내기 때문에 PCA를 통해 추출한 공통 요인을 잠재 요인(latent factor)이라고 표현하며, 전체 데이터에 대해 매우 높은 설명력을 보인다.

여기까지 보면 PCA는 마치 마법과도 같다. 만약에 분석자가 팩터 모델을 구성할 때 PCA를 활용하면 스타일 팩터에 영향을 끼치는 매크로 팩터를 직접 찾을 필요가 없다. 단순히 여러 스타일 팩터 데이터에 내포된 정보를 추출하여 잠재 요인과 요인에 대한 개별 스타일 팩터의 민감도를 추정할 수 있다. 추출된 잠재 요인은 서로 독립적이고, 스타일 팩터를 설명하는 가장 효과적인 공통 요인(팩터)이다(도표 45번).

도표 45. PCA를 활용하면 여러 팩터를 독립적이고 설명력 있는 팩터로 축소할 수 있음



자료: 유진투자증권

마법과도 같은 PCA, 금융 시계열 데이터 적용에는 부적합하다

설명력이 높은 변수를 선정하고 노이즈를 제거하여 쓸만한 데이터로 만드는 전처리 과정은 분석자에게 있어서 가장 어려운 과제다. 그래서인지 많은 분석자들이 PCA를 금융 시계열 변수 축소 방법으로 활용하고 있지만, 많은 경우 PCA는 잘못 사용되고 있는 실정이다.

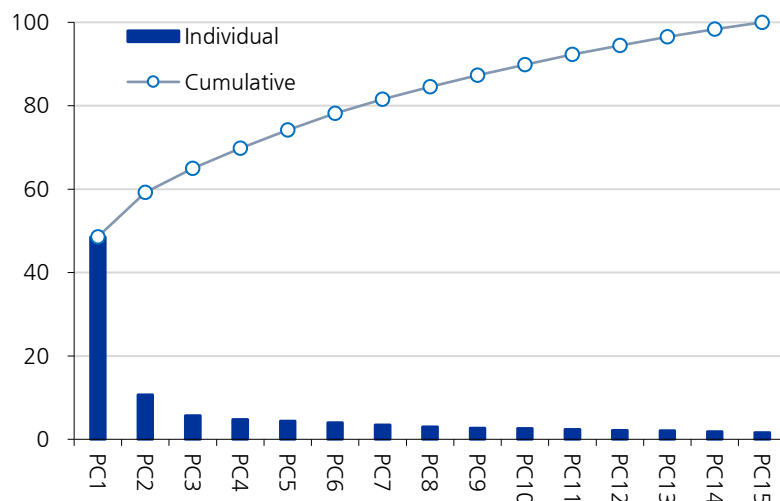
PCA의 가장 큰 문제점은 추출된 잠재 요인이 경제적 함의를 전혀 내포하지 않는다는 사실이다. 팩터 모델을 통해 예측을 하고자 하는 분석자는 자산가격에 대한 설명력이 높을 뿐만 아니라 '경제적 함의'를 가진 팩터를 선택해야 한다 언급한 바 있다. 경제적 함의가 없는 팩터는 통계적 방법을 활용한 추정에만 의존해야 하기 때문이다.

예시를 통한 이해가 가장 빠를 것이다. 도표 46번은 미 증시에 상장된 약 100여개 종목의 수익률에 대한 주성분(잠재 요인)의 누적 설명력을 나타낸다. 첫 번째 잠재 요인이 전체 전체 주성분 내 50%에 가까운 높은 설명력을 가지고 있으며, 상위 잠재 요인 6개가 주가를 약 80% 수준 설명 가능한 것이 확인된다.

그렇다면 적절한 개수의 잠재 요인을 선택하여 요인 모형을 만든다면 주가의 움직임을 설명할 수 있을까?

도표 46. 주가를 설명하는 잠재 요인(latent factor)의 누적 설명력 추이

(%, 설명 가능한 분산)



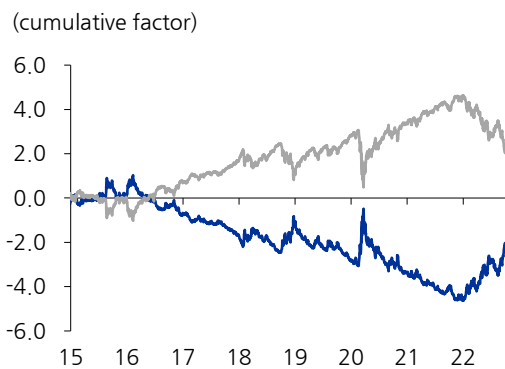
자료: 유진투자증권

잠재 요인을 누적 시계열로 나타내면 해당 팩터의 추세 정보를 시각화할 수 있다. PCA로 추출된 팩터는 민감도의 부호에 따라 반대가 될 수 있기 때문에 대칭 시계열도 관찰해야 한다. 첫 번째 잠재 요인의 누적 추이는 시장 요인과 매우 유사하다. 첫 번째 잠재 요인이 전체 주 성분 중 50%에 가까운 설명력을 갖고 있었다는 점을 감안하면 합리적인 결과다.

하지만 문제는 두 번째 잠재 요인부터 발생한다. 시장 요인 이후로는 어떠한 정보를 내포하고 있는 지 파악할 수 없다. PCA는 데이터로부터 추출한 팩터이기 때문에 금리나 환율처럼 "특정한 변수"가 아닐 가능성이 높다. 단순히 센터먼트를 의미할 수도 있으며, 여러 변수들의 종합적인 흐름을 나타낼 수도 있다.

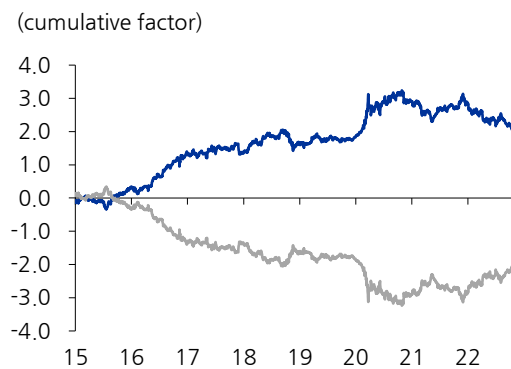
그렇다면 A라는 구간과 B구간에서 추출한 잠재 요인은 같은 팩터일까? 시장 요인을 제외하면 그렇지 않을 가능성이 클 것이다. 주가에 영향을 미치는 팩터는 구간에 따라서 달라질 것이며, 만약에 같더라도 민감도와 설명력, 심지어는 민감도에 따라 부호도 달라질 수 있다. 다른 구간에서 동일한 팩터로 작용할 지조차도 알 수 없기 때문에 시계열 분석을 통한 추정은 할 수 없다. 따라서 PCA와 같은 차원 축소법은 시간과 관계없는 비시계열 데이터의 적용에 적합하다.

도표 47. 잠재 요인(PC1) 누적 시계열: 시장 요인



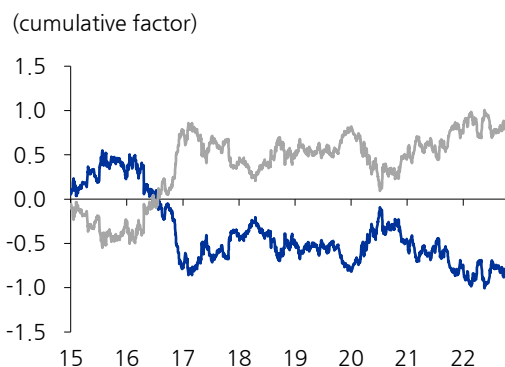
자료: 유진투자증권

도표 48. 잠재 요인(PC2) 누적 시계열: 정의 불가능



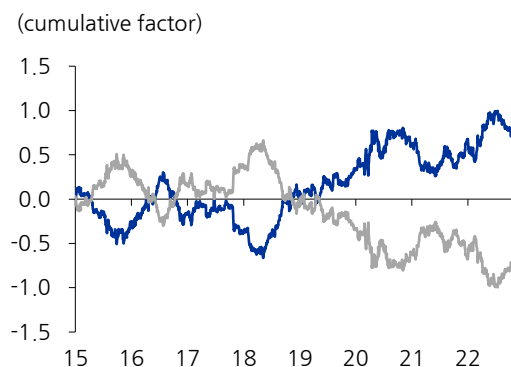
자료: 유진투자증권

도표 49. 잠재 요인(PC3) 누적 시계열: 정의 불가능



자료: 유진투자증권

도표 50. 잠재 요인(PC4) 누적 시계열: 정의 불가능



자료: 유진투자증권

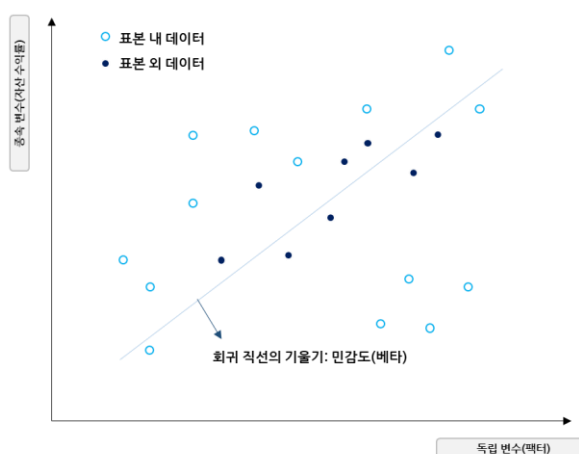
PCA의 잘못된 사용은 지나치게 이상적인 결과로 이어진다

만약 분석자가 PCA의 문제점에 대해 인지하지 못한다면 어떻게 될까? 실제로 많은 경우 PCA를 금융 시계열에 잘못 적용하여 사용하고 있다. 예를 들어 전체 데이터를 대상으로 PCA를 통해 공통 요인을 추출한 이후 그 공통 요인을 독립 변수로 활용했다고 가정해보자. 만약 전체 데이터로부터 추출한 공통 요인을 활용하여 표본 내/외 구간으로 나누고 표본 외 검증을 진행한다면, 모델의 성능은 매우 이상적으로 나타날 가능성이 높다.

PCA로 추출된 팩터와 민감도는 데이터를 가장 이상적으로 설명하는 형태(MECE)로 구성되게 된다. 이미 전체 데이터를 가장 잘 설명하는 팩터와 민감도를 추출해놓고 표본 외 검증을 진행하면 모델의 예측력은 높게 나올 수밖에 없다. 마치 데이터가 관측되기 전 미리 미래의 관측 데이터까지 포함된 회귀직선(민감도)을 그리고 나서, 실제 관측에서 예측력이 좋게 나타난다고 말하는 것과 동일하다(미래 참조의 오류, 도표 51 번).

종속변수를 잘 설명할 수 있는 적절한 독립변수를 선택하고, 가공하는 과정은 어려울 수 있다. 물론 팩터 모델은 단순해야 한다. 적은 변수로 현상을 해석할 수 있어야 모델이 샘플에 과잉적합되지 않기 때문이다. 하지만 충분히 고민하지 않은, 강건하지 않은 모델에 기반한 판단은 잘못된 투자 선택으로 이어질 수 있음을 명심해야 한다. PCA 처럼 변수의 정보를 잃지 않으면서도 규제를 통해 변수 개수를 축소할 수 있는 LASSO, Ridge 등의 회귀 역시 변수 선정의 좋은 방법으로 고려할 수 있다. 최근에는 Decision Tree의 중요도 분석 또한 유효한 변수 선정 방법으로 활용되고 있다.

도표 51. 표본 외 데이터를 미리 활용하는 미래 참조의 오류



자료: 유진투자증권

Compliance Notice

당사는 자료 작성일 기준으로 지난 3개월 간 해당종목에 대해서 유가증권 발행에 참여한 적이 없습니다

당사는 본 자료 발간일을 기준으로 해당종목의 주식을 1% 이상 보유하고 있지 않습니다

당사는 동 자료를 기관투자자 또는 제 3자에게 사전 제공한 사실이 없습니다

조사분석담당자는 자료작성일 현재 동 종목과 관련하여 재산적 이해관계가 없습니다

동 자료에 게재된 내용들은 조사분석담당자 본인의 의견을 정확하게 반영하고 있으며, 외부의 부당한 압력이나 간섭 없이 작성되었음을 확인합니다

동 자료는 당사의 제작물로서 모든 저작권은 당사에 있습니다

동 자료는 당사의 동의 없이 어떠한 경우에도 어떠한 형태로든 복제, 배포, 전송, 변형, 대여할 수 없습니다

동 자료에 수록된 내용은 당사 리서치센터가 신뢰할 만한 자료 및 정보로부터 얻어진 것이나, 당사는 그 정확성이나 완전성을 보장할 수 없습니다. 따라서 어떠한 경우에도 자료는 고객의 주식투자의 결과에 대한 법적 책임소재에 대한 증빙자료로 사용될 수 없습니다

투자기간 및 투자등급/투자의견 비율

종목추천 및 업종추천 투자기간: 12개월 (추천기준일 종가대비 추천종목의 예상 목표수익률을 의미함)

당사 투자의견 비율(%)

· STRONG BUY(매수)	추천기준일 종가대비 +50%이상	0%
· BUY(매수)	추천기준일 종가대비 +15%이상 ~ +50%미만	94%
· HOLD(중립)	추천기준일 종가대비 -10%이상 ~ +15%미만	5%
· REDUCE(매도)	추천기준일 종가대비 -10%미만	1%

(2023.3.31 기준)