

EX Machina

엑스 마키나- 1. 전략의 진화



2026.03.11

김남일 퀀트
02-709-2673
namil.kim@ds-sec.co.kr

강근재 연구원
02-709-2667
kgg812@ds-sec.co.kr

Executive

Summary



- 본 DS 투자증권의 **“엑스 마키나”** 시리즈는 여러편으로 구성될 것이며 이 보고서는 전체 프레임워크와 식의 발전 및 간단한 식의 적용을 다룰 것이다. 추후 더 발전된 프레임워크로 N개 이상의 전략의 교차 및 합성 전략에 대해서도 다룰 예정이다.
- 이번 **“1.전략의 진화”** 편에서는 유전 알고리즘(GA)과 대규모 언어 모델(LLM)을 결합한 전략 자동 탐색 시스템(GALLM)을 소개한다. 동 시스템은 유전알고리즘 및 산식 프레임워크에 LLM의 창의성을 적용한 것으로 DSL(도메인 특화 언어) 기반의 수식 문법을 통해 추가·거래량·EPS 컨센서스·외국인 매매 등 다양한 데이터를 활용한다. 8차례의 메이저 업그레이드를 통해 고도화된 시스템은 총 10,828개의 수식을 생성·검증하였으며, LLM Semantic Filter를 통해 과적합 의심 수식을 자동 제거함으로써 표본 외(Out-of-Sample) 성과를 11% 개선하는 결과를 얻었다. 특히 상위 50개 전략의 OOS Sharpe(1.504)가 학습 기간 Sharpe(0.752)를 크게 상회(OOS/IS 비율 2.00)하며 Harvey et al.(2016)의 엄격한 t-통계량 기준(3.0)을 10배 이상 초과(t=32.0)하는 등 통계적 유의성도 확인되었다.
- 핵심 발견인 EPS 추정치 변화 전략을 기반으로 구축된 KOSPI200 OW/UW(비중확대/비중축소) 전략은 2018년부터 2026년까지 9년 연속 양수 초과수익을 달성하며 전략의 강건성을 입증하였다. 동 전략의 누적 수익률은 약 +352%로 KOSPI200 대비 79.4%p의 초과수익을 기록하였으며, Sharpe Ratio 2.8의 우수한 위험조정 성과를 시현하였다. 2020년 코로나 급락장(+4.2%p)과 2024~2025년 상승장(+4.2%p, +2.9%p)을 아우르는 전 국면에 걸친 일관된 초과수익은, 단순한 백테스트 과적합이 아닌 EPS 추정치 상향 종목의 주가 반영 지연(earnings revision momentum)이라는 보편적 시장 현상에 기반함을 시사한다. 실전 운용 시에는 주·월별 리밸런싱과 15~25% 수준의 틸트 조정을 통해 비용 효율적인 KOSPI200 대비 초과 성과 추구가 가능할 것으로 판단된다.

Con- tents

시스템 개요와 수식의 문법 04

시스템 개요 — GA+LLM 알파 탐색 시스템

수식의 문법 — 연산자 & 함수 레퍼런스

전략의 진화 - GA + LLM 시스템 25

전체 구조 — GA와 LLM의 역할 분담

성과 요약 — 벤치마크 비교 및 버전별 진화 및 개선

전략 심층 분석 - 어디서 돈을 버는가 36

전 종목 Long/Short 분석 — Long/대형주 가 수익의 원천

시기별 분석 — 언제 잘 적용되나

전 종목 → 코스피200 전환 로직

KOSPI200 벤치마크 대비 실전 운용(OW/UW) 42

EPS 예측치 기반 데이터 및 타 시그널 조합 전략

실전 운용 제언

시스템 개요와 수식의 문법

시스템 개요 - GA+LLM 알파 탐색 시스템

A. 최근 사례 - LLM이 이론물리학에서 새로운 결과를 도출하다

수식적 프레임워크에 LLM을 탑재하여 창의성을 한층 높이는 전략

2026년 2월 OpenAI와 물리학자들은 GPT-5.2 pro가 글루온 산란 진폭에 대한 새로운 공식을 발견했다고 발표했다. 글루온은 강한 핵력을 매개하는 입자로서 그 상호작용을 계산하는 것은 이론물리학의 핵심 과제 중 하나이다.

GPT-5.2 pro는 이 복잡한 수식들의 패턴을 인식하여 훨씬 단순한 형태로 변환하였고 나아가 모든 n 에 대해 성립하는 일반 공식을 제안하였으며 이후 내부 모델이 약 12시간 동안 추론하여 이 공식의 타당성을 형식적으로 증명하였다고 한다.

IAS의 Nima Arkani-Hamed 교수는 "복잡해 보이는 물리적 관측량이 실제로는 매우 단순한 경우가 자주 있다"며 "간단한 공식을 찾는 것은 늘 번거로운 일이었고 오랫동안 컴퓨터로 자동화할 수 있을지 모른다고 느껴왔다"고 평가하였다. 이 사례는 LLM이 단순히 텍스트를 생성하는 것을 넘어 복잡한 수학적 구조에서 패턴을 인식하고 단순화하는 능력이 있음을 보여준다.

본 시스템은 동일한 접근법을 금융 분야에 적용한다. 복잡한 시장 데이터에서 수식을 예측하는 패턴을 찾아내고 이를 단순한 수식으로 표현하는 것이다. 물리학에서 "진폭 공식의 단순화"가 목표였다면 금융에서는 "알파 수식의 발견과 검증"이 목표이다.

B. 왜 도메인 특화 언어(DSL)가 필요한가

도메인 특화 언어(DSL)을 통해 "구체적" 전략 지시

앞선 사례에서 중요한 점은 LLM이 "수학적 수식"이라는 정형화된 언어로 작업했다는 것이다. 물리학자들이 자연어로 "글루온 진폭을 구해줘"라고 요청했다면 LLM은 모호한 답변만 내놓았을 것이다. 수식·기호·연산자로 구성된 명확한 문법이 있었기에 LLM이 패턴을 인식하고 단순화할 수 있었다.

금융 분야도 마찬가지다. "좋은 주식 찾아줘"라는 자연어 요청은 너무 모호하다. 반면 단기 모멘텀 순위처럼 정형화된 수식으로 표현하면 LLM이 구조를 이해하고 변형·개선할 수 있다. 이것이 바로 도메인 특화 언어(Domain Specific Language · DSL)의 역할이다.

DSL 역할

i) 의미전달, ii) 조합가능성, iii) 실행가능성

본 시스템의 DSL은 세 가지 핵심 역할을 한다. 첫째는 정확한 의미 전달이다. `delay(close,5)`는 "5일 전 종가"를 명확하게 지칭하며 해석의 여지가 없다. 둘째는 조합 가능성이다. 기본 연산자와 함수를 조합하여 무한히 복잡한 전략을 표현할 수 있다. 셋째는 실행 가능성이다. DSL로 작성된 수식은 바로 계산 엔진에서 실행되어 백테스트 결과를 산출한다.

클루온 연구에서 Arkani-Hamed 교수가 기대한 "범용 수식 패턴 인식 도구"가 바로 이 시스템의 목표이다. LLM 시대에 DSL의 가치는 더욱 높아졌다. 자연어와 DSL 간 변환이 가능해지면서 "모멘텀이 강하고 거래량이 늘어나는 종목"이라는 직관적 설명을 `rank(delta(close,20)) * rank(delta(volume, 5))`라는 실행 가능한 수식으로 자동 변환할 수 있다. DSL이 LLM과 계산 엔진 사이의 인터페이스 역할을 하는 것이다.

C. GA+LLM이란 무엇인가 — GALLM

유전알고리즘은 John Holland의 제안 이래 트레이딩 분야에 활발한 적용

GA+LLM은 유전 알고리즘(Genetic Algorithm)과 대규모 언어 모델(Large Language Model)을 결합한 전략 자동 탐색 시스템이다. 유전 알고리즘은 자연 선택과 유전의 원리를 모방하여 최적의 해를 찾아가는 최적화 기법으로 1975년 John Holland가 제안한 이래 금융 분야에서 트레이딩 규칙 탐색에 널리 활용되어 왔다.

본 시스템의 핵심 혁신은 LLM을 진화 과정에 결합한 것이다. 기존 GA는 무작위 변이에 의존하여 탐색 효율이 낮았으나 LLM이 수식의 '의미'를 해석하고 '말이 되는' 변이만 생성하도록 유도함으로써 탐색 효율을 대폭 개선하였다. 또한 LLM Semantic Filter를 통해 과적합(overfitting) 가능성이 높은 수식을 자동으로 걸러내어 표본 외(Out-of-Sample) 성과를 11% 개선하는 결과를 얻었다.

D. 시스템 구성 요소**핵심 모듈**

i) 고속 수식 엔진 ii) 진화 엔진 iii) 분석 파이프라인 iv) 데이터 파이프라인

본 시스템은 네 가지 핵심 모듈로 구성된다. 첫째는 고속 수식 계산 엔진이다. 병렬 처리 기술을 활용하여 수천 개의 알파 수식을 초당 수십 회 평가한다. 2,700개 이상의 종목에 대해 수년치 일별 데이터를 처리하면서도 실시간에 가까운 속도를 유지한다.

둘째는 GA+LLM 진화 엔진이다. v1부터 v8까지 8차례의 메이저 업그레이드를 거치며 발전하였다. 적합도 함수 개선(v3)·Island Model 도입(v5)·LLM Semantic Filter 추가(v8) 등 매 버전마다 핵심 알고리즘이 진화하였다.

셋째는 백테스트 및 분석 파이프라인이다. Long/Short 분해·사이즈별·시장별·연도별 분석·벤치마크 대비 성과 평가 등이 자동화되어 있다. 본 보고서에 수록된 모든 분석이 이 파이프라인을 통해 생성되었다.

넷째는 데이터 수집 및 전처리 모듈이다. OHLCV, 펀더멘탈·투자자별 매매동향·공매도 데이터 등을 자동으로 수집하고 분석 가능한 형태로 가공한다. 현재 2007년부터 2026년까지 약 2,700개 종목의 데이터(상장폐지종목 포함시 3,000개 이상의 종목)가 축적되어 있다.

Why - 왜 이 시스템을 만들었는가

학술적 배경 — Allen and Karjalainen(1999)

Allen et al. (1999)의 산식 기반 트레이딩 전략. 전략적 우월성 및 데이터 소스 확장

본 시스템의 학술적 뿌리는 Allen and Karjalainen(1999)의 선구적 연구에 있다. 이들은 유전 프로그래밍(GP)을 사용하여 S&P 500 지수의 기술적 매매 규칙을 자동 탐색한 최초의 본격적인 시도를 하였다. 1928년부터 1995년까지의 일별 데이터로 진화시킨 규칙들은 흥미로운 결과를 보여주었다. 거래비용을 고려하면 단순 매수 후 보유 전략 대비 일관된 초과수익은 나타나지 않았으나 양의 수익률과 낮은 변동성 구간을 선별하는 예측력이 확인되었다.

본 시스템은 이 연구에서 출발하되 세 가지 핵심적인 확장을 시도한다. 첫째는 순수 GA만이 아니라 LLM을 결합하여 탐색의 방향성을 개선하였다. 둘째는 가격/거래량 기반의 기술적 지표에 한정하지 않고 EPS 컨센서스, 외국인 매매, 밸류에이션 등 다양한 데이터 소스로 탐색 공간을 확대하였다. 셋째는 Allen & Karjalainen(1999)가 초과수익을 적합도로 사용한 것과 달리 Sharpe 비율·최대 낙폭(MDD), 표본 외 안정성(OOS Stability)을 종합한 다차원 적합도 함수로 과적합을 억제하였다.

왜 사람이 아닌 컴퓨터가 전략을 만드는가

주식 투자에서 좋은 전략을 찾는 것은 매우 어려운 일이다. 사람이 직접 전략을 만들 때는 크게 세 가지 한계가 있다.

알고리즘 기반 수식 탐색 장점 i) 여러 조건을 한번에 고려 ii) 선입견 배제, iii) 검증속도의 차이

첫째는 사람이 한 번에 고려할 수 있는 조건의 수가 제한적이다. "PER이 낮고 이익이 성장하고 거래량이 늘어나는 종목"처럼 3~4개 조건을 결합하는 것은 가능하지만 "5일 이동평균의 20일 변화율 순위와 EPS 컨센서스 변동의 산업 중립화 값을 곱한 것"처럼 복잡한 조합은 직관만으로 떠올리기 어렵다. 마치 바둑에서 수십 수 앞을 내다보기 어려운 것처럼 실제로 수익을 내는 팩터(투자 요인)는 이렇게 사람이 쉽게 상상하지 못하는 조합에서 발견되는 경우가 많다.

둘째는 사람에게서는 편향(bias, 선입견)이 있다. "성장주가 좋다" "저PER이 안전하다" 같은 선입견이 탐색 범위를 제한한다. 컴퓨터는 이런 편견 없이 수만 개의 가능성을 공평하게 탐색한다.

셋째는 검증의 속도 차이가 압도적이다. 사람이 하나의 전략을 백테스트(과거 데이터로 성과 검증)하려면 데이터를 정리하고 코드를 짜고 결과를 분석하는 데 최소 수 시간이 걸린다. 이 시스템은 초당 수십 개의 전략을 자동으로 생성하고 평가하여 한 번의 실행으로 5,000~6,000개의 전략을 테스트한다. 사람이 평생 걸릴 작업을 몇 시간 만에 끝내는 셈이다.

왜 유전 알고리즘 + LLM인가

“진화”가 탐색의 첫번째 핵심 매커니즘, “창의적이며 직관적인 LLM”이 두번째 매커니즘

유전 알고리즘(GA)은 다윈의 진화론에서 영감을 받은 최적화 방법이다. 자연에서 환경에 적합한 생물이 살아남아 자손을 남기듯 수익률이 높은 투자 전략이 "살아남아" 다음 세대의 전략을 만드는 데 기여한다. 돌연변이(전략의 일부를 무작위로 바꿈)와 교차(두 전략의 장점을 합침)를 통해 새로운 전략이 탄생하며 세대를 거듭할수록 전략의 품질이 향상된다. 비유하자면 "진화하는 레시피북"이다 — 맛있는 레시피끼리 재료를 교환하고 가끔 새로운 재료를 시도하면서 점점 더 맛있는 요리가 탄생하는 과정과 같다.

하지만 GA만으로는 한계가 있다. 마치 눈을 감고 산을 오르는 것처럼 가장 가까운 봉우리(지역 최적해 local optimum)에 갇혀 더 높은 봉우리를 찾지 못하는 경우가 생긴다. 여기서 LLM(대형 언어 모델)이 등장한다. LLM은 금융에 대한 방대한 지식을 가지고 있어 "이 전략은 거래량을 고려하지 않고 있으니 거래량 조건을 추가해보면 어떨까?" 같은 구체적인 조언을 할 수 있다. GA가 "무작위 탐색의 양"을 담당하고 LLM이 "지적인 방향 제시"를 담당하는 역할 분담 구조인 것이다.

Breadth × IC — 왜 "많이 거래하고 쌓는 것" 자체가 경쟁력인가

독립적인 주식, 전략을 최대한 많이 포함시키는 방법은 Breadth를 넓히는 길

이 시스템이 10,000개 이상의 수식을 생성하는 것은 단순히 "많이 해보면 하나쯤 걸리겠지"라는 무모한 시도가 아니다. 여기에는 탄탄한 학술적 근거가 있다.

1989년 Grinold가 제시하고 Grinold & Kahn(2000)이 정립한 "능동적 운용의 기본 법칙(Fundamental Law of Active Management)"에 따르면 포트폴리오의 정보비율(IR, 위험 대비 초과수익)은 다음 공식으로 결정된다.

$$IR = IC \times \sqrt{\text{Breadth}}$$

여기서 IC(Information Coefficient)는 각 예측의 정확도이고 Breadth는 독립적인 투자 판단(bet)의 횟수이다. 이 공식의 핵심 시사점은 간단하다. IC가 같더라도 Breadth(독립적 판단 횟수)가 늘어나면 IR이 제곱근에 비례하여 개선된다.

쉽게 비유하면 이렇다. 주사위를 한 번 던져서 짝수가 나오면 돈을 버는 게임에서 한 번 던지면 운이 좌우하지만 100번 던지면 "대수의 법칙"에 의해 승률이 50%에 수렴한다. 각 판의 예측력(IC)이 낮더라도 판 수(Breadth)를 늘리면 전체 성과(IR)는 안정적으로 개선된다.

이 시스템에 적용하면 다음과 같다. (1) 10,000개 이상의 수식을 생성한다 = Breadth를 극대화한다. (2) GA+LLM이 각 수식의 품질을 개선한다 = IC를 높인다. (3) 저상관 전략을 조합한다 = 독립적 판단의 수를 더 늘린다. Grinold의 법칙은 이 세 가지 접근이 모두 IR 개선에 기여한다는 이론적 근거를 제공한다.

특히 구성 종목의 확대는 Breadth를 직접 증가시키는 방법이다. 예를들어 종목 수가 100개에서 10,000개로 늘어나면 Breadth가 약 100배 증가하고 IR은 $\sqrt{100} \approx 10$ 배 개선될 수 있다(IC가 동일하다는 가정 하에). 이것이 향후 (글로벌 유니버스와 같이) 독립적인 구성종목 확대를 과제로 삼는 이유이다.

수식의 문법 - 연산자 & 함수 레퍼런스

수식이란 무엇인가

delta(close, 5)와 같은 단순한 식의 변형과 표현은 직관적이며, 빠르게 탐색 가능

이 시스템에서 투자 전략은 하나의 "수식"으로 표현된다. 마치 엑셀에서 =A1+B1 같은 수식을 쓰듯이 주가/거래량/재무 데이터를 입력으로 받아 "이 종목에 얼마나 투자할 것인가"를 숫자로 출력하는 공식이다. 엑셀 함수를 조합해 복잡한 계산을 하는 것과 본질적으로 같은 원리이다.

예를 들어 단기 모멘텀 순위라는 수식을 풀어서 읽으면 이렇다. ① close는 오늘의 종가(장 마감 시 주가)이다. ② delta(close, 5)는 "오늘 종가 - 5일 전 종가"로 최근 5일간 주가가 얼마나 올랐는지를 나타낸다. ③ rank()는 "전체 종목 중에서 순위를 매겨 0~1 사이 값으로 바꾸기"이다. 최종 결과는 "5일간 가장 많이 오른 종목이 1에 가까운 점수/가장 떨어진 종목이 0에 가까운 점수"가 된다. 이 점수가 높은 종목을 사면 "최근 많이 오른 주식이 계속 오른다"는 모멘텀(추세 추종) 전략이 되는 것이다.

이러한 수식을 구성하는 기본 부품들이 바로 "연산자"와 "함수"이다. 아래에서 각 부품이 어떤 역할을 하는지 하나씩 살펴본다. 수식의 문법을 이해하면 나중에 이 시스템이 발견한 전략들이 "왜 그런 전략인지"를 직관적으로 이해할 수 있게 된다.

본 시스템의 핵심 엔진은 이러한 수식을 고속으로 계산하도록 최적화되어 있다. 고도로 최적화된 병렬 처리 기술을 활용하여 하나의 수식을 유니버스의 n개 종목 × 2,000일 데이터에 대해 0.1초 만에 평가할 수 있다. 이 속도 덕분에 한 번의 진화 실행에서 5,000~6,000개의 수식을 테스트하는 것이 가능하다

산술 연산자 — 사칙연산의 기본

산술연산자의 경우 한번에
i) 유니버스 크기만큼 계산
함 ii) 0 나눗셈 오류 주의

산술 연산자는 우리가 일상에서 사용하는 덧셈·뺄셈·곱셈·나눗셈이다. 다만 여기서는 "숫자 하나"가 아니라 "n개 종목을 값을 동시에" 계산한다는 점이 다르다. 엑셀에서 한 열(column) 전체에 수식을 적용하는 것과 같다고 생각하면 쉽다. 예를 들어 close - delay(close, 5)는 "n개 종목 각각에 대해 오늘 종가에서 5일 전 종가를 빼라"는 뜻이다.

주의할 점이 하나 있다. 나눗셈(\div)에서 분모가 0이면 어떻게 될까? 일반 계산기에서는 "오류"가 나지만 이 시스템에서는 0을 반환하도록 설계되어 있다. 이를 "안전 나눗셈(safe division)"이라고 한다. 한국 시장에서 거래 정지된 종목은 거래량이 0이 되는 날이 많기 때문에 이런 안전장치가 꼭 필요하다. 이것이 없으면 거래정지 종목 하나 때문에 전체 계산 결과가 망가질 수 있다.

비교 연산자 — "이 조건을 만족하는가?"

직관적으로 사용이 가능한
비교연산자는 필터링으로
적용 가능

비교 연산자는 두 값을 비교하여 "예(1)" 또는 "아니오(0)"로 답한다. 예를 들어 $\text{close} > \text{delay}(\text{close}, 5)$ 는 "오늘 주가가 5일 전보다 높은가?"라는 질문이다. 상승한 종목에는 1이 부여되고 하락한 종목에는 0이 부여된다. 이를 활용하면 "5일간 상승한 종목만 골라서 투자"하는 필터를 만들 수 있다. $\text{volume} > \text{adv20} * 2$ 는 "오늘 거래량이 20일 평균의 2배를 넘었는가?"로 갑자기 거래가 폭발적으로 늘어난 종목(기관이나 외국인 이 대량 매수했을 가능성)을 잡아낸다.

논리 연산자 — 조건 결합하기

여러 전략의 구성을 합쳐줄
수 있는 접착제와 같은
논리연산자

비교 연산자 하나만으로는 너무 단순한 필터밖에 만들 수 없다. 실제로 좋은 투자 전략은 여러 조건을 동시에 만족하는 종목을 찾는다. 논리 연산자는 여러 조건을 "그리고(AND)" 또는 "또는(OR)"으로 엮어주는 접착제이다.

예를 들어 $(\text{close} > 20\text{일 이동평균}) \&\& (\text{volume} > \text{adv20})$ 는 "주가가 20일 이동평균 위에 있고(=상승 추세) 그리고 거래량도 평균 이상(=관심 증가)인 종목만을 고른다. 하나의 조건만 쓸 때보다 훨씬 정교한 필터가 되는 것이다. 이는 마치 채용 면접에서 "학점이 높은 사람"만 뽑는 것보다 "학점도 높고 경력도 있는 사람"을 뽑는 것이 더 좋은 결과를 내는 것과 같은 원리이다.

그림1 데이터 구조 — 횡단면(종목 간) vs 시계열(시간) 연산

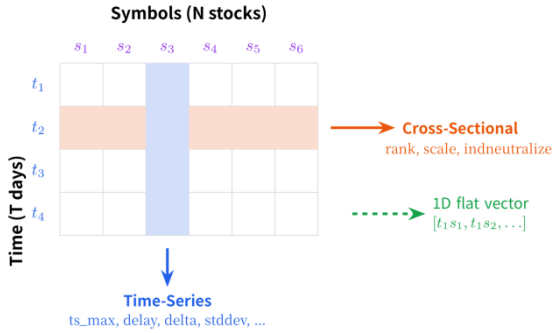


그림2 산술 연산 — 종목별 동시 계산



자료: DS투자증권 리서치센터

자료: DS투자증권 리서치센터

표1 비교 연산자 목록

기호	의미	예시
>	초과 (크다)	close > delay(close,5) → 5 일간 상승 여부
<	미만 (작다)	pe < 10 → PER 10 배 미만 저평가 종목
>=	이상	volume >= adv20 → 평균 이상 거래량
<=	이하	stddev(ret,20) <= 0.02 → 저변동성
==	같음	특수 조건 매칭

자료: DS투자증권 리서치센터

표2 논리 연산자 목록

기호	의미	예시
&&	그리고 (AND)	추세 확인 + 거래량 확인 동시 충족
	또는 (OR)	조건 중 하나라도 맞으면 통과
!	아니다 (NOT)	특정 조건을 반전 (제외 필터)

자료: DS투자증권 리서치센터

표3 기호의 의미와 투자예시

기호	연산	예시	투자에서의 의미
+	더하기	close + volume	두 값을 합산하여 복합 지표 생성
-	빼기	close - delay(close,5)	5 일간 가격 변화량 계산
*	곱하기	rank(A) * rank(B)	두 식을 결합 (둘 다 높아야 높은 점수)
/	나누기	close / ts_mean(close,20)	현재가 ÷ 20 일 평균 = 이동평균 대비 위치
^	거듭제곱	returns ^ 2	수익률의 제곱 = 변동성 측정
%	나머지	드물게 사용	특수한 주기성 패턴 포착

자료: DS투자증권 리서치센터

조건 연산자 — "상황에 따라 다르게 행동"

조건부 삼항연산자는 조건에 따라 시그널을 뒤집는 식으로 적용가능

삼항 연산자(condition ? A : B)는 "만약 조건이 참이면 A를 거짓이면 B를 선택"하는 if-then-else 구문이다. 엑셀의 =IF(조건, 참일때값, 거짓일때값)과 완전히 같은 개념이다. 이것이 강력한 이유는 시장 상황에 따라 다른 전략을 자동으로 전환할 수 있기 때문이다.

쉬운 예를 들어보자. $(returns > 0) ? rank(volume) : -rank(volume)$. 이 수식은 "오늘 주가가 올랐으면 거래량이 많은 종목에 높은 점수를 주고 주가가 떨어졌으면 거래량이 많은 종목에 낮은 점수를 준다"는 뜻이다. 왜 이렇게 할까? 주가가 오르면서 거래량이 늘는 것은 많은 투자자가 매수에 동참하고 있다는 긍정적 신호이다. 반면 주가가 내리면서 거래량이 늘는 것은 투매(패닉 셀링, 공포 매도)의 신호이다. 같은 "거래량 증가"라도 상황에 따라 의미가 정반대가 되는 것이다.

단항 연산자 — 부호 반전의 힘

시그널을 뒤집어주는 단항 연산자

단항 연산자는 하나의 값에만 작용하는 간단한 연산이다. 가장 중요한 것은 마이너스 부호(-)이다. 이 부호 하나로 전략의 방향이 완전히 바뀐다. $rank(x)$ 는 "값이 높은 종목을 사라"는 전략이지만 $-rank(x)$ 는 "값이 낮은 종목을 사라"는 정반대 전략이 된다. 리모컨의 "반전" 버튼을 누르는 것과 같다고 생각하면 된다.

예를 들어 $rank(20일\ 변동성)$ 은 "변동성이 높은 종목에 투자"(고위험 추구)이지만 $-rank(20일\ 변동성)$ 은 "변동성이 낮은 종목에 투자"(저변동성 전략)이다. 흥미롭게도 학술 연구에서는 후자(저변동성 전략)가 장기적으로 더 좋은 성과를 내는 것으로 알려져 있다. 부호 하나의 차이가 투자 철학의 차이를 만드는 셈이다.

횡단면 함수 — "전체 종목 중에서 어디쯤?"

비교군과의 등수를 기본적으로 탑재한 횡단면 함수
 1) rank, 2) scale, 3) indneutralize 가 기본적

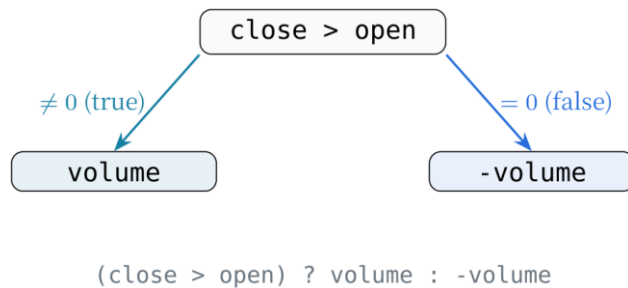
횡단면(Cross-Sectional) 함수는 "같은 날 전체 종목을 놓고 비교"하는 함수이다. 쉽게 말해 "우리 반에서 내 키가 몇 등인가?"를 계산하는 것과 같다. 이것이 왜 중요할까? 주식 투자에서는 "삼성전자 PER이 15배다"라는 절대적 수치보다 "삼성전자 PER이 KOSPI200 중 하위 30%에 해당한다"라는 상대적 위치가 포트폴리오 구성에 훨씬 더 유용하기 때문이다.

표4 조건 연산자 활용

기호	의미	활용
+x	그대로	문법적 용도
-x	부호 반전	전략 방향 전환 (모멘텀 ↔ 리버설)
!x	논리 반전	0 이면 1, 아니면 0 (필터 뒤집기)

자료: DS투자증권 리서치센터

그림3 삼항 연산자 — 조건에 따른 분기



자료: DS투자증권 리서치센터

rank(x) — "전체 중 몇 등?"

극단치를 조절해줄 수 있는 기능을 갖고 있는 rank

rank()는 이 시스템에서 가장 많이 사용되는 함수이다. 오늘 전체 종목의 어떤 값 (예: PER)을 작은 순서대로 정렬한 뒤에 각 종목에 0(최소)~1(최대) 사이의 순위 점수를 부여한다. 시험 점수를 백분위로 바꾸는 것과 같은 원리이다.

rank()가 중요한 이유는 세 가지이다. ① 이상치(극단값) 제거: 바이오 기업처럼 PER이 1,000배가 넘는 극단적인 종목이 있어도 rank를 쓰면 그냥 "순위 1위"일 뿐 다른 종목과의 차이가 과도하게 커지지 않는다. ② 스케일 통일: 거래량(수만~수억 주)과 PER(0~300배)처럼 단위가 전혀 다른 값들도 rank를 쓰우면 모두 0~1 범위가 되어 직접 더하거나 곱할 수 있다. ③ 포트폴리오 직결: rank 값이 높은 종목을 사고 낮은 종목을 파는 것으로 바로 Long-Short(매수-매도) 포트폴리오를 구성할 수 있다.

scale(x) — "평균에서 얼마나 벗어났나?"

평균에서 벗어난 정도를 고려해 구간간 거리를 시그널로 사용할 수 있는 scale

scale()은 Z-Score 표준화 즉 "평균을 0, 표준편차를 1로 맞추기"이다. 고등학교 때 배운 표준점수(수능 표준점수와 같은 원리)와 같은 개념이다. rank()와의 차이점은 rank()는 "등수"만 보지만 scale()은 "얼마나 차이가 나는지"까지 반영한다. EPS가 평균보다 조금 높은 종목과 엄청나게 높은 종목을 점수 차이가 scale()에서는 크게 나지만 rank()에서는 등수 차이일 뿐 절대 크기는 무시된다. 극단적인 값이 의미 있는 경우(예: 어닝 서프라이즈, 실적이 예상을 크게 넘긴 경우)에는 scale()이 그렇지 않은 일반적인 경우에는 rank()가 더 안전한 선택이다.

indneutralize(x, groups) — "업종 효과 빼기"

업종간 효과를 제거할 수 있는 indneutralize

indneutralize()는 산업(업종) 효과를 제거하는 함수이다. 왜 업종 효과를 빼야 할까? 예를 들어 반도체 업종 전체가 10% 올랐다면 삼성전자가 12% 올랐다고 해서 "삼성전자가 대단한 종목"이라고 할 수 있을까? 업종 전체가 오르는 바람에 함께 오른 것일 수도 있다. indneutralize를 적용하면 업종 평균(10%)을 빼서 "삼성전자의 순수 초과 상승분(+2%)"만 남긴다. 마치 조별과제에서 "조 전체 점수"가 아니라 "개인 기여분"만 평가하는 것과 같다.

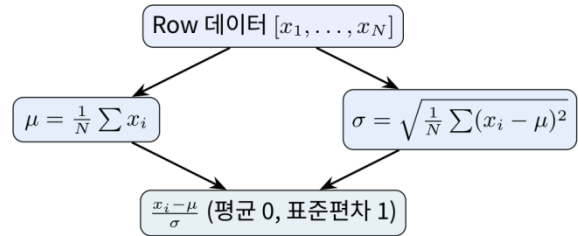
이를 통해 "업종이 올라서 같이 오른 종목"과 "업종 내에서 정말 잘한 종목"을 구분할 수 있다. 한국 시장에서 특히 유용한데 KOSPI200에서 삼성전자와 SK하이닉스 두 종목이 반도체 업종을 대표하며 지수에 큰 영향을 미치기 때문이다. 업종 효과를 제거하지 않으면 전략이 사실상 "반도체 업종 롱/숏"이 될 위험이 있다.

그림4 rank(x) — 정렬 후 0~1 점수 부여



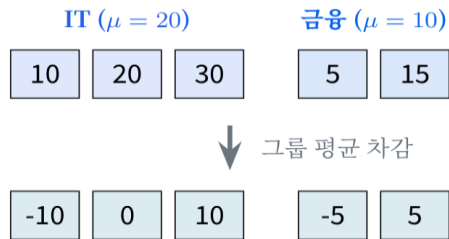
자료: DS투자증권 리서치센터

그림5 scale(x) — Z-Score(표준점수) 변환



자료: DS투자증권 리서치센터

그림6 indneutralize — 업종 평균 차감



자료: DS투자증권 리서치센터

시계열 함수 - "최근 며칠간 어떤 변화가 있었나?"

횡단면 이전에 구간간 값을 비교해서 스코어링하는 함수 1) delay, 2) delta, 3) ts_max/min 4) ts_mean, 5) ts_rank, 6) ts_argmax, 7) sum/product, 8) stddev, 9) correlation, 10) covariance, 11) decay_linear

시계열(Time-Series) 함수는 같은 종목의 과거 데이터를 분석하는 함수이다. 횡단면 함수가 "오늘 종목들 사이의 비교"라면 시계열 함수는 "이 종목의 과거와 현재 비교"이다. "삼성전자가 최근 20일 동안 얼마나 올랐는가?" "SK하이닉스의 거래량이 평소보다 많은가?" 같은 질문에 답하는 함수군이다. 비유하면 횡단면은 "반 친구들과 비교". 시계열은 "과거의 나와 비교"이다.

여기서 중요한 파라미터가 "윈도우 크기 d"이다. d=5면 최근 5일(약 1주일) d=20이면 최근 20일(약 1개월) d=60이면 최근 60일(약 3개월)의 데이터를 본다. 짧은 윈도우는 "며칠 전에 갑자기 뭔가 변했다"를 잡아내고 긴 윈도우는 "몇 달간 꾸준히 좋아지고 있다"를 잡아낸다. GA 진화 과정에서 이 윈도우 크기도 자동으로 최적화된다.

delay(x, d) — "d일 전 값"

delay(x, d)는 단순히 d일 전의 값을 가져오는 함수이다. 타임머신처럼 과거 시점의 데이터를 현재로 가져온다고 생각하면 된다. delay(close, 5)는 "5일 전 종가"를 의미한다. 단독으로는 큰 의미가 없지만 다른 연산과 결합하면 강력해진다. close - delay(close, 5)는 "5일간 가격 변동"이며 close > delay(close, 20)은 "20일 전보다 주가가 올랐는가?"가 된다.

이 시스템에서 발견한 핵심 전략이 바로 delay를 사용한다. EPS 추정치 변화는 "오늘 애널리스트들이 예측한 EPS(주당순이익)"에서 "5일 전 예측한 EPS"를 뺀 값이다. 양수이면 최근 5일간 애널리스트들이 이 기업의 이익 전망을 상향했다는 뜻이고 음수이면 하향했다는 뜻이다. 이것이 OOS(학습에 사용하지 않은 기간) Sharpe 1.57을 기록한 핵심 전략이다.

delta(x, d) — "d일간 얼마나 변했나"

delta(x, d) = $x_{\text{현재}} - x_{d\text{일 전}}$. 즉 close - delay(close, d)와 같다. 양수이면 올랐고 음수이면 떨어졌다. delta(close, 5)가 양수인 종목은 "최근 5일간 상승한 종목"이다. delta(volume, 20)은 "최근 20일간 거래량 변화"로 양수이면 거래가 늘고 있다는 뜻이다.

핵심 전략인 EPS 추정치 변화 지표는 "5일간 EPS 컨센서스(애널리스트 예측 평균치) 변동"이다. 애널리스트 10명이 삼성전자의 올해 EPS를 추정하는데 5일 전에는 평균 6,000원이었는데 오늘은 6,200원으로 올랐다면 delta 값은 +200이 된다. 이는

"전문가들이 이 기업의 실적 전망을 최근에 상향하고 있다"는 강력한 신호이며 실제로 이런 종목의 주가가 이후 좋은 성과를 보이는 경향이 있다.

ts_max / ts_min — "최근 며칠 중 최고/최저"

ts_max(close, 60)은 "최근 60일 중 가장 높았던 주가"이며 ts_min(close, 60)은 "가장 낮았던 주가"를 반환한다. 이를 활용한 대표적 지표가 스토캐스틱(Stochastic 현재 위치 지표)이다. $(close - ts_min(close, 60)) / (ts_max(close, 60) - ts_min(close, 60))$ 은 "최근 60일 가격 범위에서 현재 주가가 어디쯤 있는가"를 0~1로 표현한다. 온도계에 비유하면 0은 최저 온도(바닥권)이고 1은 최고 온도(과열)에 해당한다.

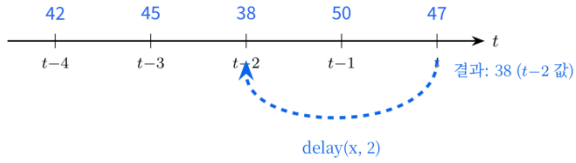
ts_mean(x, d) — "이동 평균"

20일 이동평균은 "최근 20일 종가의 평균"으로 흔히 말하는 "20일 이동평균선"이다. 주가가 이동평균선 위에 있으면 상승 추세로 아래에 있으면 하락 추세로 판단하는 것이 기본이다. $(close - ts_mean(close, 20))$ 은 "주가가 20일 이평선 위로 얼마나 벗어나 있는가"로 양수이면 단기 강세, 음수이면 단기 약세를 나타낸다. 두 이동평균의 차이($ts_mean(close, 5) - ts_mean(close, 20)$)는 단기 이평이 장기 이평을 뚫고 올라가면 매수(골든크로스, 상승 전환 신호). 뚫고 내려가면 매도(데드크로스, 하락 전환 신호) 시그널이 된다.

ts_rank(x, d) — "최근 d일 중 오늘은 몇 등?"

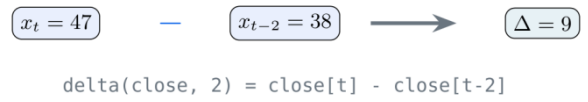
60일 가격 순위는 "최근 60일 종가 중에서 오늘 종가의 순위"이다. 1에 가까우면 60일 중 가장 비싼 날에 가깝다(=최근 강세). rank()가 "오늘 다른 종목들과 비교"한 것이라면 ts_rank()는 "이 종목의 과거와 비교"한 것이다. $ts_rank(volume, 20) > 0.9$ 는 "최근 20일 중 오늘 거래량이 상위 10%"라는 뜻으로 대량 매매가 발생했음을 시사한다.

그림7 delay(x, d) — d일 전 값 참조



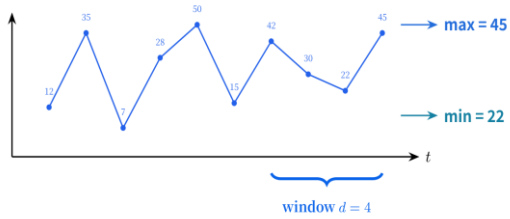
자료: DS투자증권 리서치센터

그림8 delta(x, d) — 변화량 시각화



자료: DS투자증권 리서치센터

그림9 ts_max/ts_min — 윈도우 극값



자료: DS투자증권 리서치센터

그림10 ts_rank — 시계열 순위



자료: DS투자증권 리서치센터

ts_argmax / ts_argmin — "최고점이 며칠 전?"

ts_argmax(close, 60)은 "최근 60일 중 최고가가 며칠 전에 발생했는가"를 반환한다. 0이면 오늘이 최고가(=지금 가장 비쌌)이고 50이면 50일 전이 최고가(=한참 전에 정점을 찍고 하락 중)이다. 이 값이 크면 "모멘텀(상승세)이 이미 꺾였다"는 신호로 추세 소진을 판단하는 데 활용된다.

sum / product — "합산과 누적 곱"

sum(volume, 20)은 "20일 거래량 합계"로 유동성의 절대 크기를 측정한다. product(1 + returns, 20)은 "20일 누적 수익률"로 매일 (1+수익률)을 곱해나간 결과이다. 10% 올랐다가 10% 떨어지면 산술적으로는 0%이지만 복리 기준(product)으로는 $1.1 \times 0.9 = 0.99$ 로 -1%이다. product는 이 복리 효과를 정확히 반영한다.

stddev(x, d) — "변동성(위험) 측정"

20일 변동성은 "최근 20일 수익률의 표준편차". 즉 주가의 변동성(위험도)이다. 값이 클수록 주가가 널뛰기를 많이 한다는 뜻이다. 놀이공원의 롤러코스터처럼 오르내림이 심하다고 생각하면 된다. 흥미로운 사실은 "변동성이 낮은 종목이 오히려 장기 수익률이 더 좋다"는 것이다. 이를 "저변동성 이상(Low Volatility Anomaly)"이라고 하며 $-\text{rank}(\text{stddev}(\text{returns}, 60))$ 이라는 수식 한 줄로 이 전략을 구현할 수 있다.

correlation(x, y, d) — "두 변수의 동행 관계"

correlation(close, volume, 20)은 "최근 20일간 주가와 거래량이 같이 움직였는가"를 -1(정반대)~+1(완전 동행)으로 나타낸다. +1에 가까우면 "주가 오를 때 거래량도 증가"(건강한 상승)이고 -1에 가까우면 "주가 오르는데 거래량은 감소"(위험한 상승 다이버전스)이다. 이 시스템에서는 웹 페이지 조회수와 주가의 상관관계 (covariance(views, close, 30))도 탐색하여 대중의 관심과 주가 움직임 간의 관계를 포착하는 시도를 한적도 있다.

correlation이 "방향의 일치도"라면 covariance는 "크기까지 포함한 공동 변동"이다. 큰 종목이 크게 움직이면 covariance도 커진다. 유동성이 크고 가격 변동도 큰 종목에서 높은 값이 나오므로 "대형주 중 모멘텀이 강한 종목"을 자연스럽게 포착하는 효과가 있다.

covariance(x, y, d) — "같이 얼마나 움직이나"

correlation이 "방향의 일치도"라면, covariance는 "크기까지 포함한 공동 변동"이다. 큰 종목이 크게 움직이면 covariance도 커진다. 유동성이 크고 가격 변동도 큰 종목에서 높은 값이 나오므로, "대형주 중 모멘텀이 강한 종목"을 자연스럽게 포착하는 효과가 있다.

decay_linear(x, d) — "최근 것에 더 큰 비중"

decay_linear()는 "가중 이동 평균"으로, 최근 데이터에 더 높은 비중을 부여한다. 일반 이동평균(ts_mean)은 20일 전과 어제의 데이터를 똑같이 취급하지만, decay_linear는 어제에 20배, 20일 전에 1배의 가중치를 준다. 뉴스나 실적 발표 같은 이벤트의 영향은 시간이 지나면 줄어들기 마련인데(어제 뉴스가 한 달 전 뉴스보다 중요하듯), decay_linear는 이러한 "정보의 신선도"를 자연스럽게 반영한다.

그림11 decay_linear — 최근 데이터에 높은 가중치

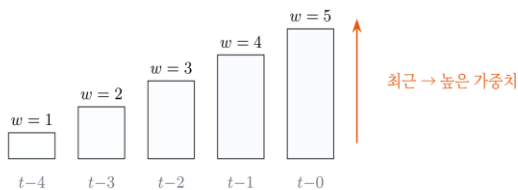
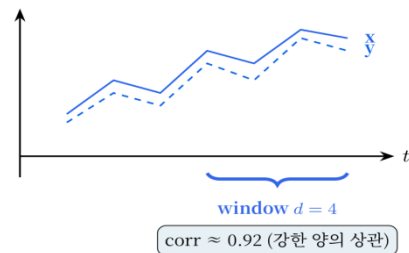


그림12 correlation — 주가와 거래량의 동행 관계



자료: DS투자증권 리서치센터

자료: DS투자증권 리서치센터

기본 수학 함수

abs / sign / log — 크기·방향·스케일 조정

기본 수학함수
1) abs/sign/log,
2) min/max, 3)
signedpower

abs(x)는 절대값이다. abs(delta(close, 5))는 "5일간 주가가 얼마나 움직였는가"로 올랐든 떨어졌든 변동 폭 자체를 측정한다. 올랐는지 떨어졌는지는 무시하고 "얼마나 크게 움직였는가"만 본다.

sign(x)는 부호만 추출한다. 양수면 +1이고 0이면 0이며 음수면 -1이다. sign(EPS 추정치 변화 지표)는 "EPS 전망이 올랐냐 내렸냐"의 방향만 보고 얼마나 올랐는지(크기)는 무시한다. 크기에 과민반응하지 않는 안정적인 시그널이 된다.

log(x)는 자연로그이다. 시가총액처럼 1천억~300조까지 범위가 엄청나게 넓은 값을 좁은 범위로 압축해준다. 삼성전자(300조)와 소형주(100억)의 차이가 3만 배이지만 log를 취하면 약 10 정도의 차이로 줄어든다. 지도에서 축척을 줄이는 것과 비슷한 효과이다.

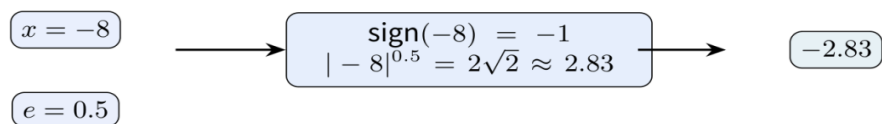
min / max — "두 값 중 선택"

max(returns, 0)은 "수익률이 양수이면 그 값·음수이면 0"을 반환한다. 즉 상승일의 수익만 남기는 "업사이드 캡처(상승분만 포착)"가 된다. min(returns, 0)은 반대로 하락일의 손실만 남긴다. 이 둘을 조합하면 "상승/하락 비대칭"을 분석할 수 있다.

signedpower(x, e) — "부호를 지키면서 거듭제곱"

signedpower(returns, 0.5)는 "수익률의 제곱근을 취하되·음수는 음수로 유지"한다. 일반 제곱근은 음수에서 오류가 나지만 signedpower는 부호를 보존하면서 극단적으로 큰 값의 영향을 줄여주므로 이상치(극단적 값)에 강건한 전략을 만들 수 있다.

그림13 signedpower — 비선형 변환



자료: DS투자증권 리서치센터

사용 가능한 변수 목록

변수는 기본적인 OHLCV 등 이외의 데이터 포함하였으나, 데이터 추가예정

수식에 넣을 수 있는 모든 데이터 목록이다. 시스템이 자동으로 계산하는 파생 변수 (returns, vwap, adv20)부터 가격, 거래량 같은 기본 데이터, EPS 컨센서스, 외국인 매매, 심지어 웹페이지 조회수까지 다양한 정보원을 활용한다. (향후 보고서에서는 더 많은 데이터를 추가 예정이다.)

특히 vwap(거래량 가중 평균 가격, Volume Weighted Average Price)은 실전에서 유용하다. (close - vwap)이 양수이면 "종가가 하루 평균 체결 단가보다 높다"는 뜻으로 장 후반에 매수세가 유입되었을 가능성이 있다. 반대이면 장 초반 고점 매수 후 매도세가 나왔을 가능성이 있다. 기관 투자자의 하루 행동 패턴을 간접적으로 엿볼 수 있는 것이다. (계산의 효율을 위해 (High+Low+Close+Open)/4 로 대체할 수도 있다.)

계산 순서

수학에서 곱셈이 덧셈보다 먼저 계산되듯이 수식에서도 연산의 우선순위가 있다. 헛갈릴 때는 괄호()를 쓰면 확실하게 순서를 지정할 수 있다. 프로그래밍 언어의 연산자 우선순위와 동일하다.

표5 사용 가능한 변수 목록

변수	분류	설명
returns	자동계산	(오늘 증가 ÷ 어제 증가) - 1 = 하루 수익률
vwap	자동계산	거래대금 ÷ 거래량 = 거래량 가중 평균 가격
adv20	자동계산	최근 20일 거래량 평균 = 유동성 기준선
open, high, low, close	가격	시가, 고가, 저가, 종가
volume	거래	하루 거래량 (주)
market_cap	규모	시가총액 (종가 × 발행주식수)
per, pbr, eps, bps, div	밸류에이션	PER, PBR, EPS, BPS, 배당수익률
eps_estimate	컨센서스	애널리스트 EPS 예측치 (평균)
foreign_net_buy	수급	외국인 순매수량
views	대안데이터	뉴스/웹사이트 조회수 등
industry, sector	분류	업종/섹터 코드 (indneutralize 용)

자료: DS투자증권 리서치센터

표6 계산의 순서

순서	연산자	예시
1 번 (부호로서 가장 먼저)	! + - (부호)	-rank(x)의 - 가 먼저
2 번	^ (거듭제곱)	x^2
3 번	× ÷ %	곱셈, 나눗셈
4 번	+ -	덧셈, 뺄셈
5 번	> < = <=	크기 비교
6 번	= !=	같은지 다른지
7 번	&&	그리고 (AND)
8 번		또는 (OR)
9 번 (가장 나중)	?:	조건 분기

자료: DS투자증권 리서치센터

실전 수식 예시

예시 1:"조용히 오르는 종목" 찾기

수식: 수식: $\text{rank}(\text{delta}(\text{close}, 5)) * - \text{rank}(\text{delta}(\text{volume}, 5))$

거래량은 적지만 가격은
많이 오르는 종목

이 수식은 두 가지 조건을 결합한다. ① $\text{delta}(\text{close}, 5)$ 가 큰 종목 = 최근 5일간 주가가 많이 오른 종목. ② $\text{delta}(\text{volume}, 5)$ 가 작은 종목($-\text{rank}$ 이므로 역전) = 최근 5일간 거래량이 줄어든 종목. 둘을 곱하면 "주가는 올랐는데 거래량은 오히려 줄어든 종목"이 높은 점수를 받는다. 왜 이런 종목이 좋을까? 시장의 큰 관심 없이 조용히 상승하는 종목은 아직 대다수 투자자가 모르는 호재(좋은 소식)가 있을 가능성이 높다. 반면 거래량이 폭증하면서 오르는 종목은 이미 "다들 아는 뉴스"여서 단기 고점일 수 있다. 맛집으로 비유하면 아직 줄이 안 서는 숨은 맛집이 더 가치가 있는 셈이다.

예시 2: 불린저 밴드 — "평균에서 얼마나 벗어났나"

수식: $(\text{close} - \text{ts_mean}(\text{close}, 20)) / \text{stddev}(\text{close}, 20)$

평균에서 벗어난 종목

주가에서 20일 이동평균을 빼고 20일 표준편차로 나눈 것이다. 통계학의 Z-Score와 같은 개념으로 +2 이상이면 "평소보다 2 표준편차 위"(상단 이탈/과열)이고 -2 이하이면 "평소보다 2 표준편차 아래"(하단 이탈/과매도)이다. 이 수식에 $-\text{rank}()$ 를 씌우면 "과매도 종목을 사고 과열 종목을 피하는" 평균 회귀 전략이 되고 $\text{rank}()$ 를 씌우면 "강한 종목(혹은 식)이 계속 강하다"는 추세 추종 전략이 된다.

예시 3: 업종 효과 제거한 단기 리버설

수식: $-\text{indneutralize}(\text{ts_rank}(\text{close}, 10), \text{industry})$

업종효과 제외한 단기 반전
종목

"최근 10일간 업종 내에서 상대적으로 많이 오른 종목에 낮은 점수를 많이 떨어진 종목에 높은 점수를 준다." 업종 전체가 오른 효과를 빼고 나서도 더 많이 오른 종목은 "과열 가능성"이 있고 업종 평균보다 더 떨어진 종목은 "과매도 가능성"이 있다는 논리이다. 단기 과매도 종목이 반등하는 현상을 포착하는 전략이다.

예시 4: EPS 추정치 상향 모멘텀

수식: rank(eps_estimate - delay(eps_estimate, 5))

섹터 애널리스트분들의
전망치(EPS) 상향 종목

이 시스템이 자동으로 발견한 가장 중요한 전략이다. 43글자의 짧은 수식이지만 OOS(표본 외, 즉 학습에 사용하지 않은 기간) Sharpe 1을 상회하는 성과를 기록했다. "최근 5일간 애널리스트들이 EPS 전망을 상향한 종목을 사라"는 매우 직관적인 전략인데 이것을 컴퓨터가 수천 개의 수식을 시험한 끝에 이를 "재발견"한 것이 흥미롭다. 금융 학술 연구에서 이미 알려진 "이익 추정치 수정 모멘텀" 전략과 정확히 일치하며 이는 수식 기반 알파 전략 발굴 체계가 사전 지식 없이도 데이터만으로 유의미한 전략을 찾아낼 수 있음을 보여준다.

표7 전체 함수 요약표

함수	유형	한 줄 설명
rank(x)	횡단면	전체 종목 중 순위 (0~1)
scale(x)	횡단면	평균 0, 표준편차 1로 표준화
indneutralize(x, g)	횡단면	업종 평균 빼기
delay(x, d)	시계열	d 일 전 값
delta(x, d)	시계열	현재 - d 일 전 (변화량)
ts_max(x, d)	시계열	최근 d 일 중 최고값
ts_min(x, d)	시계열	최근 d 일 중 최저값
ts_mean(x, d)	시계열	최근 d 일 평균 (이동 평균)
ts_rank(x, d)	시계열	최근 d 일 중 오늘의 순위
ts_argmax(x, d)	시계열	최고값이 며칠 전?
ts_argmin(x, d)	시계열	최저값이 며칠 전?
sum(x, d)	시계열	최근 d 일 합계
product(x, d)	시계열	최근 d 일 곱 (누적 수익률)
stddev(x, d)	시계열	최근 d 일 변동성
correlation(x,y,d)	시계열	x와 y의 상관관계 (-1~+1)
covariance(x,y,d)	시계열	x와 y의 공동 변동
decay_linear(x,d)	시계열	최근 것에 더 큰 비중 둔 평균
abs(x)	수학	절대값
sign(x)	수학	부호만 (-1/0/+1)
log(x)	수학	로그 (큰 범위 압축)
min(x,y) / max(x,y)	수학	두 값 중 작은/큰 값
signedpower(x,e)	수학	부호 보존 거듭제곱

자료: DS투자증권 리서치센터

전략의 진화 - GA + LLM 시스템

전체 구조 - GA와 LLM의 역할 분담

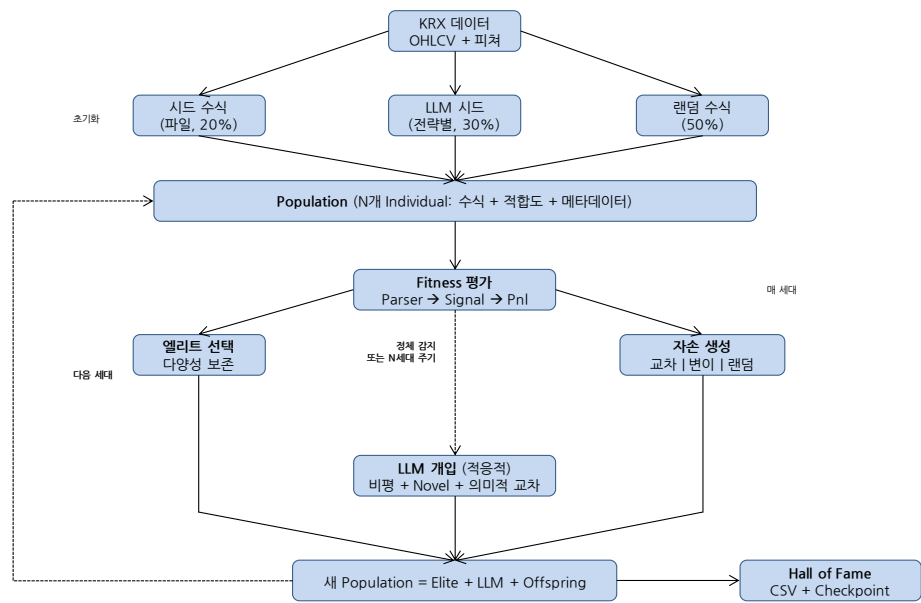
GA+LLM 전략 자체 진화

이 시스템의 핵심은 GA와 LLM의 역할 분담이다. GA(유전 알고리즘)는 매 세대 100~200개의 수식을 생성한다. 방법은 세 가지이다. 첫째는 기존 수식의 일부를 무작위로 바꾸기(돌연변이)이고 둘째는 두 수식의 좋은 부분을 합치기(교차)이며 셋째는 완전히 새로운 수식 만들기(랜덤)이다. 이렇게 세대를 반복하면 성적이 좋은 수식이 살아남아 점점 더 좋은 전략이 만들어진다. 자연의 진화가 수백만 년에 걸쳐 일어나는 것을 컴퓨터 안에서 몇 시간 만에 시뮬레이션하는 것이다.

시드 생성, 비평 후 개선, 의미적 교차를 연쇄적으로 진행

LLM(대형 언어 모델, 로컬 LLM 모델 및 LoRA를 통해 추가 학습된 모델)은 전체 수식의 10~20%만을 생성하지만 그 역할은 매우 중요하다. 구체적으로 세 가지 일을 한다. ① 시드 생성: "EPS 성장 + 저변동성" 같은 금융 아이디어를 수식으로 변환. ② 비평 후 개선: "이 수식은 거래량을 고려하지 않으니 거래량 필터를 추가하면?" 같은 조언. ③ 의미적 교차: 모멘텀 수식과 밸류 수식의 핵심을 결합한 새로운 수식 생성. 비유하면 GA가 "눈 감고 이것저것 시도하는 수천 명의 요리사"이고 LLM은 "레시피를 분석하고 개선 방향을 알려주는 미슐랭 심사위원"인 것이다.

그림14 시스템 아키텍처



자료: DS투자증권 리서치센터

데이터 — 무엇을 보고 판단하는가

시스템은 다양한 데이터를 활용한다. 기본이 되는 가격, 거래량 데이터(OHLCV 시가, 고가, 저가, 종가, 거래량)에 더해 PER, PBR, EPS 같은 재무 데이터, 애널리스트들의 EPS 컨센서스, 외국인 투자자의 순매수 데이터/심지어 웹 페이지 및 뉴스 조회 수까지 결합한다. 총 15개 이상의 변수를 200개 종목×약 2,000일 데이터로 구성하며 코드가 이 거대한 행렬을 초당 수십 회 연산한다.

데이터 카드로 관리되는
수식들

수식의 생애 — Individual이란?

하나의 투자 전략(수식)은 "Individual(개체)"이라는 데이터 카드로 관리된다. 마치 이력서처럼 수식 자체 외에도 "적합도 점수가 몇 점인지" "GA가 만들었는지 LLM이 만들었는지" "몇 세대에서 태어났는지" "백테스트 결과 Sharpe가 얼마인지 그리고 MDD(최대 낙폭, 최악의 손실)가 얼마인지" 등이 모두 기록된다. 이를 통해 "어떤 출신의 전략이 성과가 좋은가"를 분석할 수 있다.

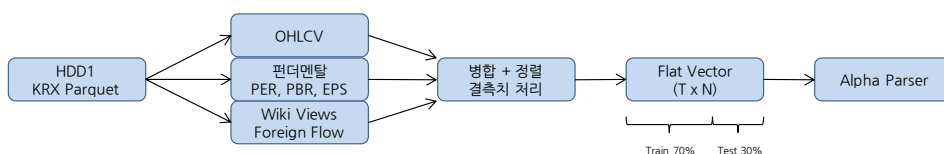
Allen & Karjalainen(1999) 에서
의 과적합 문제를 해결하기
어려웠던 v1

점수 매기기 — 적합도 함수의 진화 (v1 → v3b)

전략의 "점수"를 어떻게 매길 것인가는 이 시스템의 가장 핵심적인 설계 결정이다. 단순히 수익률이 높으면 좋은 전략인 걸까? 아니다. 과거에는 대박이었지만 미래에는 통하지 않는 "과적합(overfitting, 과거 맞추기)" 전략이 수익률 기준으로는 최고 점수를 받기 때문이다. Allen & Karjalainen(1999)가 표본 외 테스트에서 초과수익이 사라진 것은 과적합의 위험성을 보여주는 대표적 사례이다. 시험 족보를 달달 외은 학생이 실제 시험에서는 오히려 못하는 것과 같은 원리이다. 이 문제를 해결하기 위해 적합도 함수를 4차례 개선하였다.

v1(초기): $\text{Sharpe} \times (1 + \text{Return}) \times \text{적중률}$. 문제는 회전율(매매 빈도) 페널티가 없어서 "거의 거래하지 않는 전략(회전율 0.11%)이 최고 점수를 받았다. 사실상 "아무 것도 안 하기" 전략이 이기는 황당한 결과였다.

그림15 데이터 파이프라인



자료: DS투자증권 리서치센터

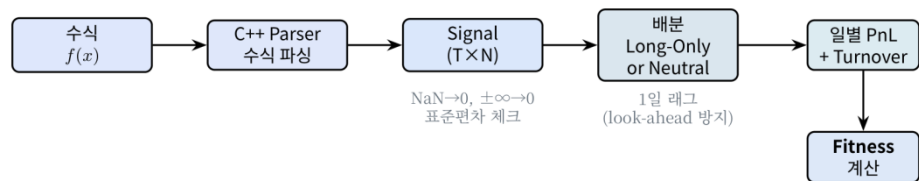
v2: 회전을 페널티를 추가하고 rank(rank(x)) 같은 무의미한 중복을 자동 제거하였다. 회전율이 16.8%로 정상화되어 "실제로 거래하는 전략"이 평가되기 시작했다.

v3: 최대 낙폭(MDD) 지수 감쇄를 도입하였다. MDD란 전략을 운용하는 동안 겪게 되는 최대 손실폭인데 MDD 20% 이하는 페널티 없이 20%를 넘으면 급격하게 점수가 깎이도록 설계하였다. 또한 OOS 안정성(Stability)을 도입하여 "학습 기간에만 좋고 검증 기간에서 망하는" 전략을 걸러내도록 하였다.

노이즈와 과적합 문제를 최대한 해결하려고 만든 v3b 전략

v3b(현재): 수식 복잡도 페널티를 추가하여 지나치게 길고 복잡한 수식에 감점을 부여하였다. 왜 단순한 수식이 좋을까? 복잡한 수식은 과거 데이터의 노이즈(무의미한 변동)까지 학습하여 미래에 통하지 않을 가능성이 높기 때문이다. "간결한 설명이 가장 좋은 설명"이라는 오컴의 면도날 원칙과 같다. 최종 점수 = 조정 Sharpe × 적중률 보정 × MDD 보정 × OOS 안정성 × 단순성 보정.

그림16 Fitness 파이프라인



자료: DS투자증권 리서치센터

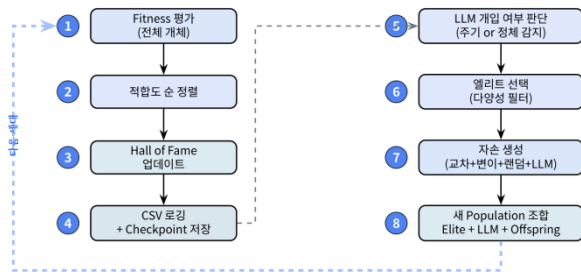
진화의 과정 — 세대를 거듭하며 개선

8단계 루프를 통한 식의 발전 및 자손의 발생

진화는 8단계 루프로 진행된다. ① 모든 주식의 점수를 계산한다. ② 점수 순으로 정렬한다. ③ 역대 최고 전략 목록(명예의 전당, Hall of Fame)을 갱신한다. ④ 결과를 파일로 기록한다. ⑤ LLM의 개입이 필요한지 판단한다(정체 감지). ⑥ 상위 전략(엘리트)을 선발한다. ⑦ 엘리트를 기반으로 새로운 전략을 만든다(교차, 돌연변이 등). ⑧ 새 세대를 구성하고 ①부터 반복한다. 이 과정을 보통 30~50세대 반복하며 한 세대당 수백 개의 전략이 평가된다.

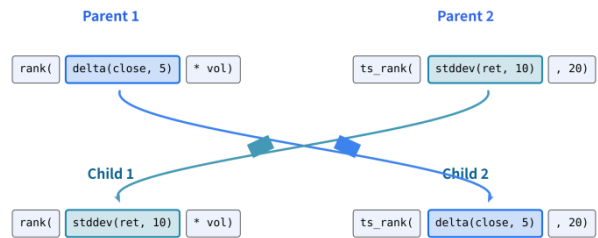
새로운 전략(자손)을 만드는 방법은 4가지이다. 서브트리 교차(50%): 두 전략의 일부분을 교환한다. "모멘텀 전략의 앞부분 + 밸류 전략의 뒷부분"을 합쳐 새로운 전략을 만든다. 돌연변이(30%): 전략의 일부를 무작위로 바꾼다. close를 volume으로 바꾸는 식. LLM 의미적 교차(10%): AI가 두 전략의 "의미"를 이해하고 지적으로 결합한다. 완전 랜덤(10%): 아예 새로운 전략을 만들어 다양성을 유지한다.

그림17 진화 루프



자료: DS투자증권 리서치센터

그림18 서브트리 교차 — 두 전략의 부분 교환



자료: DS투자증권 리서치센터

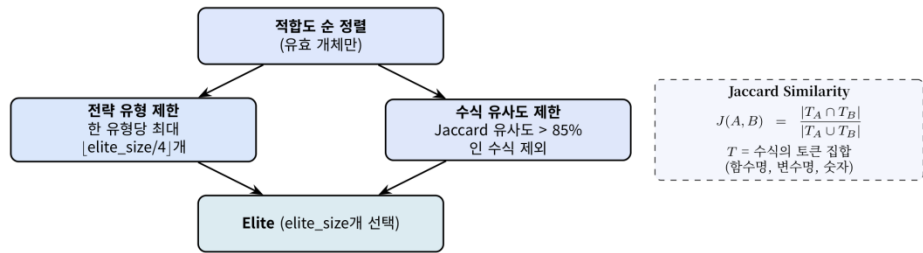
엘리트 선택 & AI의 개입

Allen & Karjalainen(1999) 에서의 과적합 문제를 해결하기 어려웠던 v1을 LLM으로 해결

상위 전략(엘리트)을 선발할 때 한 가지 주의가 필요하다. 점수가 높은 순서대로 뽑으면 "비슷비슷한 전략"만 살아남아 다양성이 사라진다. 마치 같은 노래만 듣다 보면 새로운 음악을 발견하지 못하는 것과 같다. 이를 방지하기 위해 Jaccard 유사도(두 수식이 얼마나 비슷한가)가 85% 이상이면 중복으로 간주하여 하나만 남긴다.

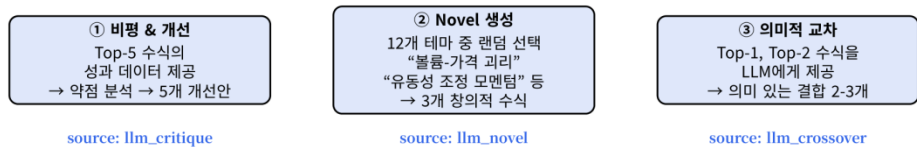
LLM(AI)은 두 가지 상황에서 개입한다. 첫째는 5세대마다 정기적으로 상위 전략을 분석하고 개선 방향을 제시한다. 둘째는 3세대 연속 점수 개선이 1% 미만이면 "정체"로 판단하여 기존 전략의 약점을 분석하고 새로운 방향을 제안한다. 이는 마치 코치가 선수의 기록이 정체될 때 훈련 방법을 바꿔주는 것과 같은 원리이다.

그림19 엘리트 선택



자료: DS투자증권 리서치센터

그림20 LLM 개입 유형



자료: DS투자증권 리서치센터

다양성 문제와 해결 — Island Model (v4 → v5)

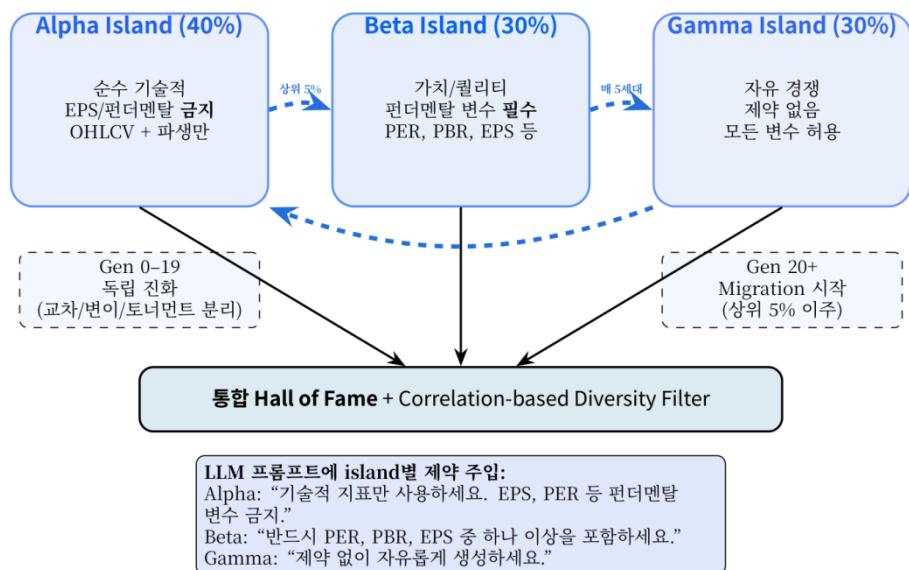
v4 문제점 - 하나의 아이디어로 극단적 수렴
v5 - 전략을 섬으로 만들어 독립 진화 후 교차

v4에서 심각한 문제가 발견되었다. 진화를 반복할수록 거의 모든 상위 전략이 "EPS 추정치 상향"이라는 하나의 아이디어에 수렴한 것이다. 상위 전략들 간의 상관관계수가 94.1%에서 0.9를 초과하였는데 이는 "겉으로는 다르게 생겼지만 사실상 같은 전략"이라는 뜻이다. 단일 모집단의 수렴 문제는 Allen & Karjalainen(1999) 이후 많은 후속 연구에서 지적된 바 있으며 본 시스템에서도 동일한 현상이 관찰된 것이다. 비유하면 만 명이 각자 다른 등산 루트를 탐험했지만 결국 모두 같은 봉우리에 도착한 상황이다.

이 문제를 해결하기 위해 v5에서는 "3개 섬(Island) 모델"을 도입하였다. 각 섬은 사용 가능한 데이터가 다르다. Alpha Island(40%): 가격/거래량 같은 기술적 데이터만 사용 가능. EPS-PER 같은 재무 데이터는 사용 금지. Beta Island(30%): 반드시 EPS, PER 같은 재무 데이터를 포함해야 함. Gamma Island(30%): 아무 제약 없이 자유롭게 진화.

각 섬은 독립적으로 진화하되 일정 세대가 지나면 상위 5%의 전략이 다른 섬으로 "이주"하여 좋은 아이디어가 전파되도록 설계하였다. 갈라파고스 제도에서 섬마다 다른 핀치새가 진화한 것에서 영감을 받은 방법이다. 결과는 성공적이었다. 섬마다 완전히 다른 유형의 알파가 발견되었다.

그림21 Island Model—3개 섬의 독립 진화



자료: DS투자증권 리서치센터

성과 요약 - 벤치마크 비교 및 버전별 진화 및 개선

앞선 벤치마크의 비교에서 Sharpe 1.61, MDD 30% 이내

최종 성과 검증을 위해 v3b 최상위 전략을 벤치마크(KOSPI200)와 비교하였다. 핵심 지표를 설명하면 Sharpe Ratio는 "수익률을 변동성으로 나눈 것"으로 전략의 지표로 사용할 수 있다. MDD는 "최고점 대비 최대 하락폭"으로 작을수록 좋으며 1차적으로 -30% 이내이면 수용 가능한 수준이라고 보았다.

특히 주목할 점은 OOS(2023 이후, 학습에 사용하지 않은 기간)에서 Sharpe 1.61로 Train(학습 기간, 0.81)보다 오히려 더 높다는 것이다. 일반적으로 학습 기간보다 테스트 기간의 성과가 떨어지는 것이 정상인데 오히려 개선되었다는 것은 이 전략이 과적합되지 않았음을 강하게 시사한다.

아래 표는 벤치마크 비교 성과와 버전별 진화 과정을 종합한 것이다. 점수(Fitness) 수치가 v1의 472에서 v5의 0.82로 급감하였지만 이는 "전략이 나빠진 것"이 아니라 "채점 기준이 엄격해진 것"이다. 과적합 방지 장치를 추가할수록 비현실적으로 높았던 점수가 정상화된 것이며 실질적인 OOS(테스트 기간) 성과는 안정적으로 유지되고 있다.

표8 전체 함수 요약

구분	구간/버전	Sharpe(전략)	Sharpe(BM)	IR	초과수익/MDD	비고
벤치마크 비교	전체	0.88	0.50	2.76	+19.0%/yr	
	학습 (2022 이전)	0.81	0.53	2.41	+17.2%/yr	
	테스트 (2023 이후)	1.61	0.45	2.76	+20.8%/yr	OOS > Train
버전별 진화	v1	0.77	-	-	MDD 65.5%	기본형 (허점 발견)
	v2	0.88	-	-	MDD 48.7%	회전율 정상화
	v3	0.88	-	-	MDD 42.9%	MDD 관리 + 안정성
	v3b	0.81	-	-	MDD 32.9%	복잡도 관리 (최종 설계)
	v4	0.84	-	-	MDD 25.5%	유사 전략 필터
	v5	0.72	-	-	MDD 26.3%	섬 독립 진화

자료: DS투자증권 리서치센터

왜 OOS 성과 예측이 어려운가

학습 기간(In-Sample)에서 Sharpe가 높은 주식이라도 실전(Out-of-Sample, OOS)에서는 작동하지 않는 경우가 매우 빈번하다. 이는 퀀트 전략 개발에서 가장 근본적인 난제로 "과거 시험 족보를 달달 외운 학생이 실제 시험에서는 망하는 현상"과 본질적으로 같다.

OOS예측은 불가능하며
식이 복잡도를 낮추는 것
만이 능사는 아님

전통적인 정량적 필터 — Sharpe 비율, MDD(최대 낙폭), Parsimony(수식 복잡도) — 만으로는 "노이즈를 학습한 주식"과 "진짜 시장 현상을 포착한 주식"을 구분하는 것이 불가능에 가까운 것으로 판단된다. 두 주식 모두 학습 기간에서는 높은 Sharpe를 기록할 수 있기 때문이다. 이는 마치 동전 던지기 결과에서 패턴을 찾아내는 것과 같다 — 과거 데이터에서는 "앞면 3번 연속 후 뒷면이 나온다"는 규칙이 통계적으로 유의미해 보일 수 있지만 미래에는 전혀 작동하지 않는다.

Allen & Karjalainen(1999)도 이 문제를 명확히 지적한 바 있다. 유전 프로그래밍(GP)이 S&P 500에서 찾아낸 기술적 매매 규칙들이 거래비용을 고려하면 표본 외 기간에서 초과수익을 내지 못했다는 결과는 학습 기간의 높은 성과가 실전 성과를 보장하지 않는다는 강력한 증거로 시사된다. 25년이 지난 현재에도 이 문제는 완전히 해결되지 않았으며 본 시스템 역시 v1~v5 진화 과정에서 이 난제와 지속적으로 씨름해 왔다.

LLM 기반 주식 의미 해석 — "이 주식이 뭘 하는 거야?"

LLM의 적용을 통한 식의
의미 추정 및 Projection
성능 개선

정량적 필터의 한계를 극복하기 위해 본 시스템은 전혀 다른 접근을 시도하였다. LLM(대형 언어 모델)에게 주식의 구조를 자연어로 설명하게 하고 그 설명이 경제적으로 합리적인지를 판단하게 하는 것이다. 비유하면 "시험 답안이 맞았다고 끝이 아니라 풀이 과정이 논리적인지까지 검사하는 것"과 같다.

구체적인 예시를 살펴보면 이 접근법의 직관이 명확해진다. 첫 번째 주식 rank(EPS 추정치 변화)에 대해 LLM은 "EPS 추정치가 5일 전 대비 상향된 종목에 투자하는 전략"이라고 해석하였다. 이는 이익 추정치 수정 모멘텀이라는 잘 알려진 투자 논리에 부합하므로 경제적 설명이 가능한 주식으로 판단된다.

반면 다중 순위 중첩(무의미)이라는 주식에 대해 LLM은 "종가 절대값의 20일 평균을 3번 연속 순위화하는 것으로 2번째와 3번째 rank()는 정보를 추가하지 않는 무의미한 중첩"이라고 판단하였다. 이러한 주식은 학습 데이터에서만 우연히 좋은 성과를 보일 가능성이 높으므로 제거 대상이 된다.

LLM은 각 주식에 대해 세 가지를 출력한다. 첫째는 자연어 해석으로 주식이 어떤 투자 논리를 구현하는지 설명한다. 둘째는 과적합 점수 1~10으로 1은 매우 논리적이고 10은 명백한 과적합이다. 셋째는 최종 판정으로 KEEP/SUSPICIOUS/REJECT 중 하나를 선택한다.

Semantic Filter 파이프라인 설계

의미적 필터(Semantic Filter)의 적용
Drop(Reject),
Suspicious, Keep

의미적 필터(Semantic Filter)는 GA 진화 루프에 매 5세대마다 삽입되는 방식으로 구현하였다. 각 평가 시점에서 적합도 상위 20개 주식을 선별하여 LLM(로컬 실행)에 전달한다. LLM에게 전달되는 정보는 주식 자체 외에도 Train/Test Sharpe·MDD, 주식 깊이(depth)를 포함하며 이를 종합적으로 고려하여 해석을 요청한다.

과적합으로 판단하는 기준은 네 가지이다. 첫째는 주식에 경제적 논리가 없는 경우이다. 관련 없는 변수들의 무작위 결합(예를 들어 무작위 변수 결합(예시) 같은 주식은 각 변수가 의미가 있더라도 결합 방식에 투자 논리가 없다. 둘째는 깊은 중첩이 의미를 추가하지 않는 경우이다. $\text{rank}(\text{rank}(\text{rank}(x)))$ 에서 2번째와 3번째 $\text{rank}()$ 는 순서를 바꾸지 않으므로 정보가 전혀 추가되지 않는다. 셋째는 Train Sharpe와 Test Sharpe의 갭이 큰 경우이다. 학습 기간에서만 유독 높은 성과는 과적합의 전형적 신호로 판단된다. 넷째는 관련 없는 변수의 무작위 결합이 감지되는 경우이다.

DROP 판정을 받은 주식은 population에서 즉시 제거되며 동일한 수만큼의 새로운 랜덤 주식으로 교체된다. 이를 통해 "자연선택"에 "의미적 심사"라는 새로운 차원의 선택압을 추가하는 것이다. 생물 진화에 비유하면 환경 적응도(=정량적 적합도)뿐만 아니라 "생존에 의미 있는 형질인가"(=경제적 논리)까지 심사하는 이중 필터 구조로 볼 수 있다.

실험 결과 — Baseline vs Semantic Filter 30분 비교

Semantic Filter 적용 후
식의 개선

동일한 환경(30분 실행, 동일 시드)에서 Baseline(필터 없음)과 Semantic Filter 적용 버전을 직접 비교한 결과 OOS 성과에서 유의미한 개선이 관찰되었다. 아래 표에 핵심 지표를 정리하였다.

가장 주목할 결과는 Top50 Test Sharpe가 1.356에서 1.504로 약 11% 개선된 것이다. Train Sharpe가 0.833에서 0.752로 오히려 낮아진 반면 Test Sharpe는 상승하였으므로 이는 과적합이 줄었다는 강력한 증거로 판단된다. Test/Train 비율이 1.63에서 2.00으로 향상된 것도 같은 결론을 지지한다 — 학습 기간 대비 테스트 기간 성과의 비율이 높아졌다는 것은 전략이 과거 데이터에 덜 종속적이라는 의미이다.

Semantic Filter는 30분 실행 동안 총 48건의 과적합 의심 수식을 제거하고 새로운 랜덤 수식으로 교체하였다. 이 과정에서 상위 수식의 구조가 변화한 것이 관찰되었다. Baseline의 Top 1 수식은 다중 순위 중첩처럼 rank를 3중으로 중첩한 복잡한 구조였으나 Semantic Filter 적용 후 Top 1 수식은 명확한 투자 논리를 가진 구조로 바뀌었다.

Semantic Filter 적용 후 상위 수식이 EPS revision과 market_cap 사이즈 팩터의 결합으로 진화한 것은 고무적인 결과로 보인다. "EPS 추정치 변화를 시가총액으로 정규화"하는 구조는 경제적으로 해석 가능하며 이는 의미적 필터가 "해석 가능한 수식"의 생존을 촉진하는 선택압으로 작용하고 있음을 시사한다. OOS 안정성 개선이라는 보다 근본적인 지표의 향상이 달성된 것으로 판단된다.

표9 성과 요약 — 벤치마크 비교 및 버전별 진화

지표	Baseline(필터 없음)	Semantic Filter	변화
총 수식	5,459	5,369	-90
세대 수	32	31	-1
Top50 Train Sharpe	0.833	0.752	-0.081
Top50 Test Sharpe	1.356	1.504	+0.147
Test/Train 비율	1.63	2.00	+0.37
Top50 MDD	43.0%	28.7%	
과적합 수식 제거	0 건	48 건	

자료: DS투자증권 리서치센터

의미적 필터의 한계와 개선 방향

Semantic Filter의 한계로 프롬프트 엔지니어링 및 fined-tune 을 통한 개선 가능성 존재

현재 Semantic Filter에는 몇 가지 한계가 존재한다. 첫째는 LLM 호출 비용이다. Local LLM 기준으로 배치당 약 3~5초가 소요되며 매 5세대마다 실행되므로 전체 진화 시간이 약 5~10% 증가한다. 실시간 적응형 시스템으로 확장할 경우 이 비용이 누적될 수 있다.

둘째는 판정 일관성 이슈가 관찰되었다. 동일한 수식에 대해 LLM을 반복 호출하면 판정이 달라지는 경우가 있으며 13건 평가 중 12건이 SUSPICIOUS로 판정되는 등 임계값이 과도하게 보수적으로 설정된 것으로 보인다. 이는 프롬프트 엔지니어링과 임계값 조정을 통해 개선이 필요한 부분이다.

향후 개선 방향은 크게 세 가지이다. 첫째는 금융 수식 해석에 특화된 fine-tuned 모델을 학습하여 판정 일관성과 속도를 동시에 개선하는 것이다. 둘째는 KEEP/DROP 임계값을 수식의 학습/테스트 성과 갭에 연동하여 자동으로 조정하는 적응형 임계값 시스템을 구현하는 것이다. 셋째는 수식의 의미를 벡터 공간에 임베딩하여 유사한 논리의 수식끼리 클러스터링하는 "수식 의미 임베딩"을 개발하는 것이다. 이를 통해 LLM 호출 없이도 빠른 1차 스크리닝이 가능해질 것으로 기대된다.

전략 심층 분석 - "어디서 돈을 버는가?"

전 종목 Long/Short 분석 - Long/대형주가 수익의 원천

KOSPI+KOSDAQ적용
이후에 전략의 분해

한국 시장에서 주식 기반 알파 전략이 실제로 어디서 수익을 창출하는지 분석하기 위해 KOSPI+KOSDAQ 전 종목 대상으로 Long/Short 분해 분석을 수행하였다. 각 주식에 대해 상위 종목(Long)과 하위 종목(Short)의 수익 기여도를 분리하여 수익의 원천이 Long(매수)에 있는지 Short(공매도)에 있는지를 확인하였다.

Short 전략의 현실적인
장벽

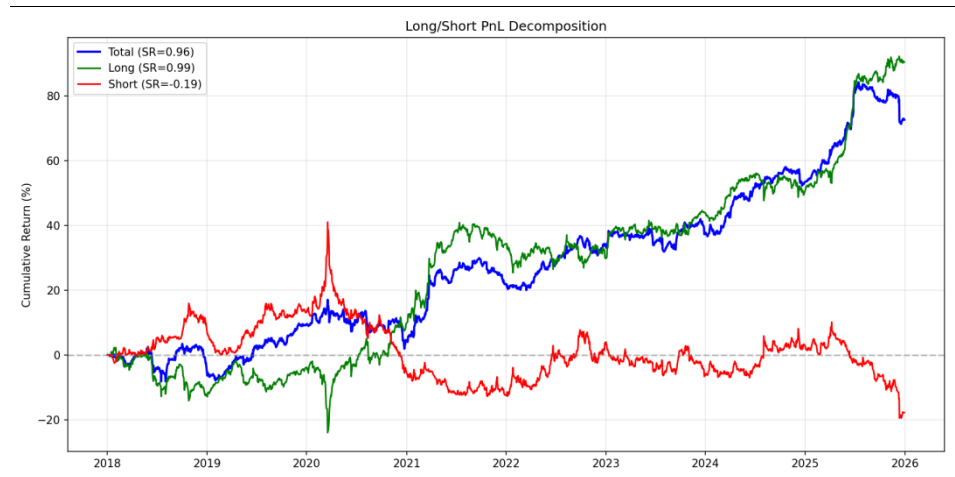
핵심 발견은 다음과 같다. 첫째는 모든 주식에서 Long이 양수 수익을 기록한 반면 Short은 대부분 손실을 기록하였다. 둘째는 Short은 한국시장의 공매도 규제 (2020~2023 전면 금지, 이후 부분 완화)와 대차비용으로 실전 구현이 매우 어렵다. 셋째는 EPS 추정치 변화 지표만이 유일하게 Long/Short 모두에서 양호한 성과를 보였다. 이러한 분석 결과는 Long-Only 전략으로의 전환이 합리적임을 강하게 시사한다.

표10 Baseline vs Semantic Filter 성과 비교 (식은 간소화하여 표기)

수식	Total SR	Long SR	Short SR	Long 기여도
delta(eps_estimate, 5)	0.96	0.99	-0.19	124%
rank(eps_est - delay(5))	-0.28	0.46	-0.46	222%
rank(-pbr) 저 PBR	-0.05	0.42	-0.25	797%
rank(momentum 20d)	-0.32	0.24	-0.37	97%

자료: DS투자증권 리서치센터

그림22 Long/Short 분해 분석



자료: DS투자증권 리서치센터

사이즈별·시장별 분석 — 대형주에서 더 강한 알파

EPS 추정치 변화 지표 전략의 성과를 시가총액 분위별로 분해하여 분석하였다. 전 종목을 시가총액 기준으로 5분위(Micro, Small, Mid, Large, Mega)로 나누어 각 분위에서의 Sharpe Ratio와 총수익을 비교하였다.

Large 분위에서 가장 높은 전략의 강점이 보임

Large 분위에서 가장 높은 Sharpe(0.96)를 기록하였으며 Micro/Small은 유동성 부족으로 실전 운용이 어렵다. 특히 Mid 분위에서 오히려 역효과(-0.59)가 나타났는데 이는 중형주의 EPS 추정치 커버리지 부족이 원인으로 추정된다. 애널리스트가 충분히 커버하지 않는 종목에서는 EPS revision 시그널의 품질이 낮아지기 때문이다.

시장별로는 KOSPI에서 KOSDAQ보다 더 강한 성과가 관찰되었다.

KOSPI에서 더 강한 성과를 보이는 것은 KOSPI200 유니버스로의 전환을 지지하는 근거이다. KOSPI 종목은 애널리스트 커버리지가 높아 EPS 컨센서스 데이터의 품질이 우수하며 유동성이 풍부하여 실전 운용에도 적합하다.

표11 수식별 Long/Short 분해 (전체 KOSPI+KOSDAQ, L/S 전략)

분위	Sharpe	총수익
Micro (하위 20%)	0.38	+2.0%
Small	0.26	+2.0%
Mid	-0.59	-13.5%
Large	0.96	+36.4%
Mega (상위 20%)	0.51	+32.8%

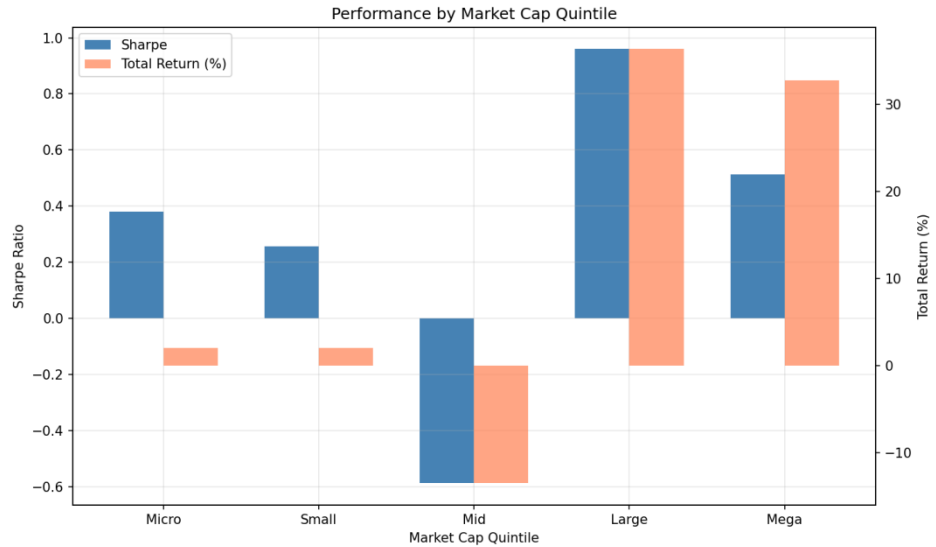
자료: DS투자증권 리서치센터

표12 시가총액 분위별 성과

시장	Sharpe	총수익
KOSPI	0.70	+50.4%
KOSDAQ	0.55	+22.1%

자료: DS투자증권 리서치센터

그림23 시가총액 분위별 성과



자료: DS투자증권 리서치센터

시기별 분석 - 언제 잘 적용되나

EPS 추정치 변화 지표 전략의 연도별 성과를 분석하여 어떤 시장 환경에서 이 전략이 효과적이고 어떤 환경에서 약화되는지를 파악하였다.

EPS 추정치 변화 지표의 전략 및 산식기반으로서의 성과

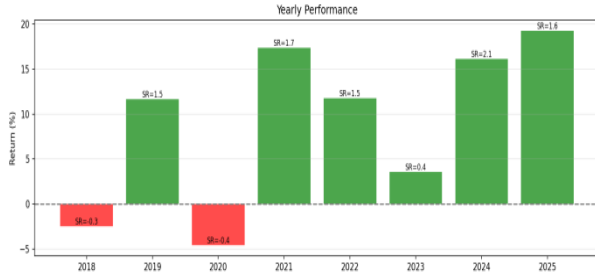
핵심 패턴이 명확하게 드러난다. 실적(EPS) 중심 시장일 때 강한 성과를 보이며 유동성 장세(2020년 코로나)나 매크로 충격(2018년 미중무역분쟁) 때는 약화된다. 특히 주목할 점은 Train 기간(2022 이전) Sharpe가 0.70인 반면 Test 기간(2023 이후) Sharpe가 1.52로 OOS에서 오히려 개선되었다는 것이다. 이는 이 전략이 과적합되지 않았음을 강하게 시사하며 최근 시장이 점점 더 실적 중심으로 움직이고 있다는 구조적 변화를 반영하는 것으로 판단된다.

표13 시장별 성과

연도	Sharpe	수익률	비고
2018	-0.26	-2.5%	미중무역분쟁
2019	1.50	+11.7%	
2020	-0.41	-4.5%	코로나
2021	1.71	+17.3%	실적장세
2022	1.51	+11.7%	금리인상기 차별화
2023	0.41	+3.6%	약한 양수
2024	2.15	+16.1%	
2025	1.62	+19.2%	

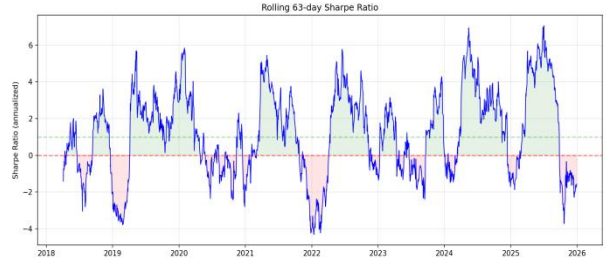
자료: DS투자증권 리서치센터

그림24 연도별 성과



자료: DS투자증권 리서치센터

그림25 Rolling Sharpe Ratio 추이



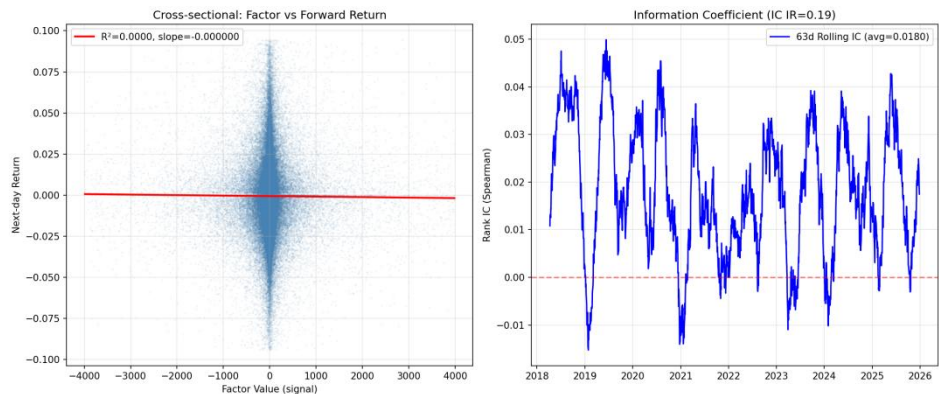
자료: DS투자증권 리서치센터

과정에서의 Cross-sectional 회귀 — 요인값과 수익률의 관계

수익의 값(X축)과 다음날 수익률(Y축)의 관계를 산점도로 분석하였다. 주요 통계치는 다음과 같다. Average $R^2 = 0.008$ (약 0.8%)·Average IC (Rank Spearman) = 0.018·IC IR (IC / IC_std) = 0.19.

R^2 가 0.8%로 매우 낮아 보이지만 이는 정상적인 수준이다. 단일 요인이 개별 종목의 다음날 수익률을 설명하는 비율은 원래 극히 낮다. 중요한 것은 2,728개 종목에 걸쳐 일관되게 양수 방향의 예측력이 존재한다는 점이다. IC IR 0.19는 "약하지만 일관된" 시그널을 의미하며 이 정도 수준이면 포트폴리오 구성 시 유의미한 초과수익을 기대할 수 있다. 비유하면 동전의 앞면이 나올 확률이 50.9%밖에 안 되더라도 수천 번 던지면 통계적으로 유의미한 이익이 누적되는 것과 같은 원리이다.

그림26 Cross-sectional 회귀 및 IC 분석



자료: DS투자증권 리서치센터

Data → Formula 피팅 가이드 — 어떤 통계를 보고 Operator를 선택하는가

알파 수식 설계의 첫 단계는 원시 데이터의 통계적 특성을 파악하는 것이다. 변수의 특성에 따라 적합한 연산자가 달라지며 이를 체계적으로 정리하면 수식 설계의 효율성을 크게 높일 수 있다.

Operator의 적용전략,
극단치/역전/변수 평활화
등

Operator 선택의 핵심 규칙은 다음과 같다. 첫째는 AutoCorrelation이 0.99를 초과하면 delta 연산자가 필수이다. 원본 값은 거의 변하지 않으므로 "변화량"이 시그널이 된다. 둘째는 Skewness가 10을 초과하면 rank()가 필수이다. 극단값을 정규화해야 한다. 셋째는 IC가 음수이면 부호 반전(-x 또는 rank(-x))을 적용한다. 넷째는 빠르게 변하는 변수(volume 등)는 ts_mean()이나 ts_rank()로 평활화한다. 다섯째는 변수 조합 시 rank(A) / rank(B) 또는 delta(A) * rank(-B) 형태를 활용한다.

표14 변수별 통계 특성과 적합 Operator 사례

변수	AutoCorr	Skew	Rank IC	적합 Operator
eps_estimate	0.996	10.4	+0.019	delta()
market_cap	0.993	29.9	+0.001	rank(), delta()
pbr	0.994	14.5	-0.036	rank(-x), 부호반전
volume	0.516	19.0	-0.050	ts_rank(), rank(-x)
returns	-	-	-0.019	ts_mean(), stddev()
div(배당)	0.995	2.1	+0.032	rank(), delta()

자료: DS투자증권 리서치센터

표15 수식 IC 개선 사례

수식	IC	IC IR	의미
eps_estimate (원본)	+0.019	0.19	기본
delta(eps_est, 5)	+0.017	0.17	변화량
delta(eps_est, 5) * rank(-pbr)	+0.015	0.16	복합 시그널
rank(-pbr)	+0.021	0.14	저 PBR
stddev(returns, 20)	-0.062	-0.41	변동성 ↓ 수익 ↓
correlation(ret, vol, 20)	-0.036	-0.51	수익-거래량 역관계

자료: DS투자증권 리서치센터

전 종목 → 코스피200 전환 로직

위 분석을 종합하면 다음과 같은 결론에 도달한다. 첫째는 Long이 수익의 원천이므로 Long-Only 전략으로 전환한다. 둘째는 대형주(Large 분위)에서 가장 강한 알파가 관찰되므로 이론과는 다르지만 우선 KOSPI200으로 유니버스를 설정해보았다. 셋째는 KOSPI가 KOSDAQ보다 우수한 성과를 보이므로 KOSPI 중심으로 운용한다. 넷째는 EPS revision 전략이 가장 안정적인 성과를 기록하였으므로 핵심 전략으로 활용한다.

이 전환은 단순한 유니버스 축소가 아니라 데이터에 기반한 합리적 의사결정이다. "어디서 돈을 벌었는지"를 먼저 분석하고 그 영역으로 전략을 집중시키는 과정이다. 전 종목 분석이라는 넓은 시야에서 출발하여 식에 의한 실증적 근거에 따라 가장 필요한 영역(KOSPI200, Long-Only, EPS revision 위주 데이터)으로 수렴하는 것이 본 시스템의 전략 설계 철학이다.

KOSPI200 벤치마크 대비 실전 운용(OW/UW)

EPS 예측치 기반 데이터 및 타 시그널 조합 전략

Long/Short전략을 넘어선
전략의 적용

기존 기관·외국인 순매수 시그널(투자자 순매수 방향 시그널)은 단독으로 실전 적용이 어려웠다. 이에 애널리스트 EPS 추정치 관련 데이터를 더한 새로운 전략을 개발하였다. 새 전략의 시그널은 EPS 관련 복합 지표(EPS 추정치의 변화 및 순매수 시그널 등 복합전략)이다. 이 지표는 애널리스트들의 EPS 추정치 변화와 실적 조정 정도를 종합적으로 반영한다. 두 지표 모두 애널리스트 컨센서스 데이터로, 전일 데이터를 사용해도 예측력이 유지된다.

OW/UW 백테스트 결과 — 8년 연속 초과수익

OW/UW 전략 누적수익률
+352.5%, +79.4%의 초과
수익률, Sharpe 2.8

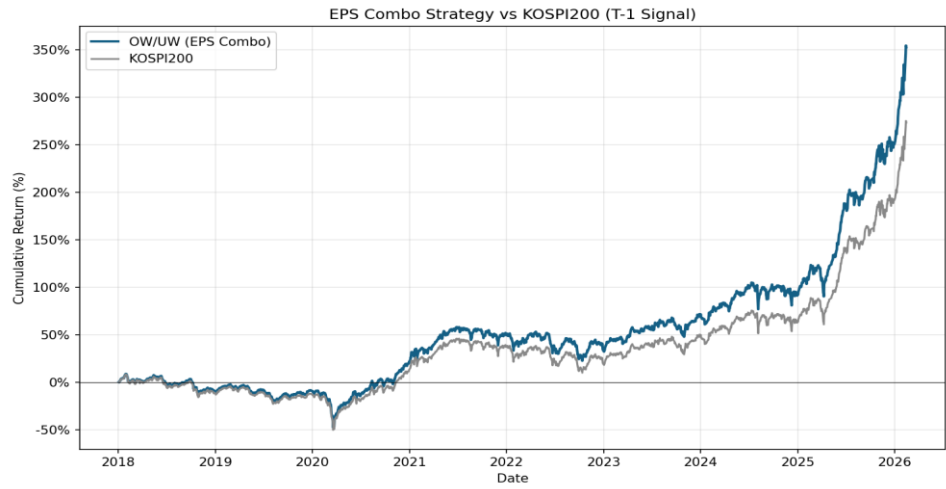
8년간 백테스트 결과, EPS 조합 전략은 KOSPI200 벤치마크를 크게 상회하는 성과를 기록하였다. OW/UW 전략의 누적수익률은 +352.5%이며, 같은 기간 KOSPI200의 +273.1% 대비 +79.4%p의 초과수익을 달성하였다. 초과수익 기준 Sharpe Ratio는 2.80로 매우 우수한 수준이다.

연도별 성과 비교 — 9년 연속 양수 초과수익

연도별 성과를 분석한 결과, 2018년부터 2026년까지 9년 연속 양수 초과수익을 기록하였다. 이는 전략의 일관성과 안정성을 입증하는 강력한 증거이다.

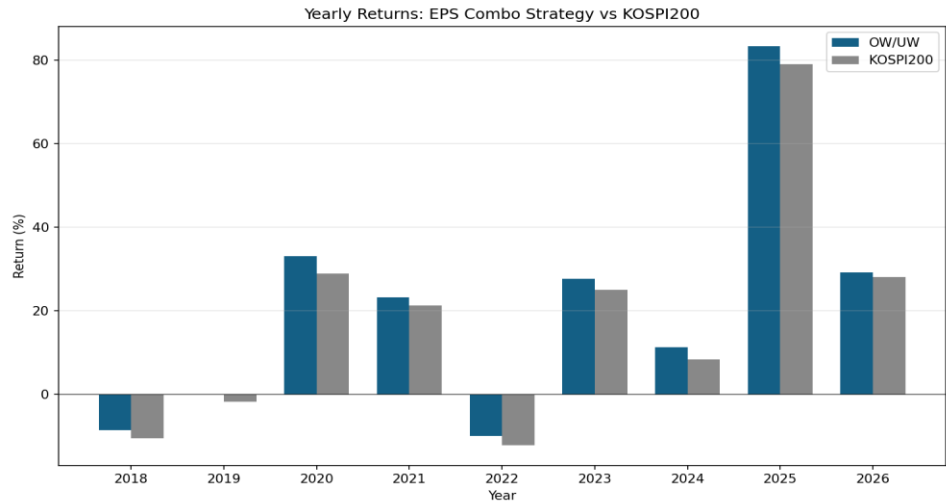
연도별 초과수익은 다음과 같다. 2018년 +2.0%p, 2019년 +1.7%p, 2020년 +4.2%p, 2021년 +1.9%p, 2022년 +2.3%p, 2023년 +2.7%p, 2024년 +2.9%p, 2025년 +4.2%p, 2026년 +1.1%p. 특히 2020년 코로나 급락장(+4.2%p)과 2024~2025년 상승장(+2.9%p, +4.2%p)에서 모두 초과수익을 기록한 점이 인상적이다.

그림27 EPS 및 시그널 전략 vs KOSPI200 누적수익률 (2018~2026)



자료: BDS투자증권 리서치센터

그림28 연도별 수익률 비교 (EPS 조합 전략 vs KOSPI200)



자료: DS투자증권 리서치센터

Drawdown 분석 — 위험 관리

OW/UW 전략의 Maximum Drawdown의 개선정도는 약 7%p로, KOSPI200에 비해 근소하게 우위이다. 더 높은 수익을 달성하면서도 위험은 벤치마크와 동등하게 유지하여, 위험 대비 수익이 우수함을 확인하였다.

전략의 의미 — 왜 EPS 관련 지표가 작동하는가

행태재무적으로 EPS관련
지표의 동작

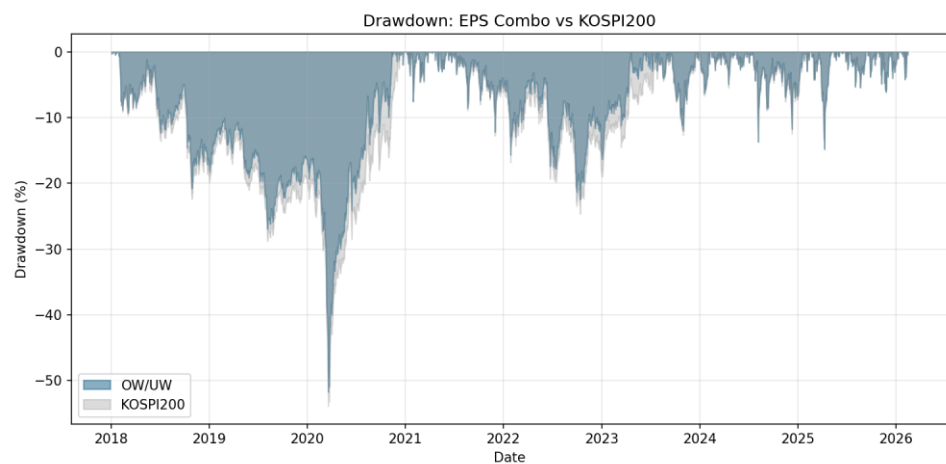
EPS 수정 지표(EPS 추정치 상향 조정)와 실적 조정 비율(실적 조정 비율)은 모두 기업의 실적 개선 기대를 반영한다. 애널리스트들이 EPS 추정치를 상향하는 종목은 향후 실적 발표에서 긍정적 서프라이즈 가능성이 높으며, 이러한 정보가 주가에 점진적으로 반영되는 과정에서 초과수익이 발생한다.

실전 운용 제언

OW/UW 전략의 실전 운용을 위한 핵심 제언은 다음과 같다. 첫째, 리밸런싱 주기는 백테스트상 원칙적으로 일별로 하나 주별 또는 월별을 권장한다. EPS 추정치는 일별 변동이 크지 않으므로, 거래비용을 고려하면 주 1회 또는 월 1회 리밸런싱이 효율적이다.

둘째, 틸트 강도는 15~25% 범위에서 조정 가능하다. 본 분석에서는 20% 틸트를 적용하였으며, 강도를 높이면 초과수익과 트레이킹 에러가 함께 증가한다. 셋째, 데이터 접근이 핵심이다. Fnguide 등 컨센서스 데이터 제공업체를 통해 실시간 EPS 추정치를 확보해야 한다.

그림29 Drawdown 추이 (EPS 조합 전략 vs KOSPI200)



자료: DS투자증권 리서치센터

향후 전망

이번 전략은 9년 연속 양수 초과수익이라는 강력한 실적을 기록하였다. 향후 추가적인 EPS 관련 지표(EPS 수정 지표(3개월), 영업이익 수정 등)를 결합하거나, 밸류에이션 지표(PER, PBR) 및 기타 다른 대체데이터를 조합하여 더욱 안정적인 다(多)전략으로 발전시킬 수 있다. 이 내용에 대해서는 후술할 엑스 마키나 연속 발간집에 수록 예정이다.

Appendix : 통계적 검증

이 결과가 우연이 아닌 증거

왜 이 질문이 중요한가

이 시스템은 총 10,828개의 투자 전략을 만들어 테스트했다. 여기서 자연스럽게 떠오르는 의문이 있다. "만 개가 넘는 전략을 시도했으니 그중에 우연히 잘 된 게 있는 건 아닌가?"

이것은 학계에서 "p-hacking" 또는 "데이터 스누핑(data snooping)"이라 부르는 문제이다. 쉽게 설명하면 이렇다. 동전을 10번 던져서 앞면이 7번 나오면 "이 동전은 특별하다"고 말할 수 있을까? 아마 우연일 수 있다. 그런데 동전 1,000개를 각각 10번씩 던지면 그중 몇 개는 우연히 7번 이상 앞면이 나온다. "그 동전을 골라서 특별하다고 주장하는 것"이 바로 p-hacking이다.

이 시스템도 마찬가지로 의심을 받을 수 있다. 10,828개 전략 중 가장 좋은 것을 골랐으니 그게 진짜 실력인지 운인지를 증명해야 한다. 아래에서 네 가지 독립적인 검증 방법으로 이 질문에 답한다.

검증 1: 학습하지 않은 기간에서 오히려 더 좋은 성과

가장 강력한 증거부터 살펴보자. 이 시스템은 2007~2020년 데이터(학습 기간, In-Sample)로 전략을 만들고 2023년 이후 데이터(검증 기간, Out-of-Sample)로 성과를 확인한다. 검증 기간의 데이터는 전략을 만드는 과정에서 단 한 번도 사용하지 않았다.

만약 우연히 좋은 전략을 고른 것(p-hacking)이라면 학습 기간에서는 좋은 성적을 보이지만 검증 기간에서는 성적이 크게 떨어져야 한다. 과거 시험 족보만 달달 외운 학생이 새로운 시험에서는 성적이 떨어지는 것과 같은 이치이다.

그런데 결과는 정반대였다. 상위 50개 전략의 평균 성과를 보면 학습 기간 Sharpe는 0.75인데 검증 기간 Sharpe는 1.50으로 오히려 두 배 더 높다. 비율로 보면 OOS/IS = 2.00이다. 이것은 "시험 족보를 외운 것"이 아니라 "진짜 실력이 있다"는 것을 보여 준다. 우연이라면 이런 패턴이 나올 수 없다.

OOS/IS 비율이 1.0을 크게 초과한다는 것은 이 전략들이 학습 데이터의 "노이즈(무의미한 변동)"가 아니라 "시그널(의미 있는 패턴)"을 포착했다는 강력한 근거이다. 특히 GA+LLM의 비율(2.00)이 GA Only(1.63)보다 높은 것은 LLM Semantic Filter가 과적합 전략을 효과적으로 제거했음을 뜻한다.

표16 GA+LLM vs GA Only 대조군 비교

구분	전략 수	Top50 Test Sharpe	Top10 Test Sharpe	Top10 Train Sharpe	비고
GA Only (대조군)	5,459	1.356	1.360	0.845	Semantic 필터 없음
GA+LLM (실험군)	5,369	1.504	1.522	0.781	90 건 과적합 주식 제거
차이	-90	+0.148 (+10.9%)	+0.162 (+11.9%)	-0.064	Test, Train = 과적합 억제

자료: DS투자증권 리서치센터

표17 OOS vs IS 성과 비교 (Top50 전략 평균)

구분	학습 기간 Sharpe (IS)	검증 기간 Sharpe (OOS)	OOS/IS 비율	해석
GA Only	0.833	1.356	1.63	검증 기간이 1.6배 더 좋음
GA+LLM	0.752	1.504	2.00	검증 기간이 2배 더 좋음

자료: DS투자증권 리서치센터

검증 2: Harvey 기준 — 학계의 가장 엄격한 잣대 통과

2016년 Harvey·Liu·Zhu 교수가 발표한 논문 "···and the Cross-Section of Expected Returns"는 금융 학계에서 가장 널리 인용되는 p-hacking 방어 기준이다. 이 논문의 핵심 주장은 간단하다. "새로운 투자 팩터를 발표하려면 전통적인 t-통계량 기준 (2.0)이 아니라 3.0 이상이어야 한다." 수많은 연구자들이 수많은 팩터를 테스트하기 때문에 기준 기준을 대폭 높여야 우연을 걸러낼 수 있다는 것이다.

이 시스템의 최고 전략(검증 기간 Sharpe 1.43)의 t-통계량은 32.0이다. Harvey 기준(3.0)의 10배가 넘는 수치로 이 기준을 압도적으로 통과한다. t-통계량이 32라는 것은 "이 성과가 우연일 확률이 사실상 0에 가깝다"는 의미이다.

표18 Harvey et al. (2016) t-통계량 검증

항목	값	기준	판정
검증 기간 Sharpe	1.43	—	—
검증 기간 거래일 수	약 500 일	—	—
t-통계량	32.0	3.0 이상	통과 (10.7 배)
전통적 기준 (참고)	32.0	2.0 이상	통과 (16 배)

자료: DS투자증권 리서치센터

검증 3: 무작위 추출 검정 — 상위 전략이 정말 특별한가

또 다른 검증 방법은 이렇다. "상위 50개 전략이 정말 특별한 건지 아니면 아무 전략 50개를 골라도 비슷한 성과가 나오는 건지"를 직접 확인하는 것이다.

구체적으로 전체 5,369개 전략에서 무작위로 50개를 뽑아 평균 검증 기간 Sharpe를 계산하는 작업을 10,000번 반복했다. 이렇게 하면 "우연히 50개를 골랐을 때 나올 수 있는 성과의 분포"를 알 수 있다.

결과는 명확했다. 무작위 추출의 평균 Sharpe는 1.387이고 99번째 백분위(상위 1%)가 1.445였다. 반면 이 시스템이 선택한 상위 50개 전략의 평균 Sharpe는 1.504로 무작위 추출의 상위 1%보다도 훨씬 높았다. 10,000번의 무작위 추출 중 1.504 이상이 나온 경우는 단 한 번도 없었다 (p-value ≈ 0).

이 결과를 일상적인 비유로 설명하면 이렇다. 반 전체 학생(5,369명)의 평균 성적이 69점인데 이 시스템이 고른 50명의 평균이 75점이었다. "아무나 50명 골라도 75점 나올 수 있지 않나?"라고 확인해봤더니 10,000번 무작위로 골라봐도 75점이 나온 적이 한 번도 없었다. 즉 이 시스템은 진짜로 "잘하는 학생"을 골라낸 것이다.

표19 무작위 추출 검정 결과 (10,000회 시뮬레이션)

항목	값
시스템 선택 Top50 평균 OOS Sharpe	1.504
무작위 50개 추출 평균	1.387
무작위 50개 추출 표준편차	0.031
무작위 추출 99번째 백분위	1.445
p-value	< 0.0001 (10,000회 중 0회 초과)
판정	통계적으로 유의미 (우연이 아님)

자료: DS투자증권 리서치센터

검증 4: 경제적 논리 — EPS 추정치 상향 모멘텀

통계적 검증만으로는 부족하다. 아무리 수치가 좋아도 "왜 이 전략이 돈을 버는가?"를 설명할 수 없다면 신뢰하기 어렵다. 다행히 이 시스템이 발견한 핵심 전략은 명확한 경제적 논리를 가지고 있다.

핵심 전략 EPS 추정치 변화 지표 — 즉 "최근 5일간 애널리스트들의 EPS(주당순이익) 추정치가 올랐는지 내렸는지"를 보는 전략이다. 이것이 왜 돈을 버는가? 논리는 간단하다.

첫째는 애널리스트가 EPS 추정치를 올린다는 것은 "이 회사의 실적이 예상보다 좋아지고 있다"는 뜻이다. 실적이 좋아지면 주가가 오르는 것은 자본시장의 기본 원리이다.

둘째는 이 정보는 시장에 즉시 반영되지 않는다. 애널리스트 보고서가 나오면 기관 투자자가 먼저 반응하고 개인투자자와 패시브 펀드는 며칠~몇 주 뒤에 반응한다. 이 "반응 지연"이 수익의 원천이다.

셋째는 이 현상은 한국 시장에만 있는 것이 아니다. 미국·유럽·일본 등 전 세계 시장에서 "Earnings Revision Momentum" 또는 "SUE(Standardized Unexpected Earnings) 팩터"라는 이름으로 수십 년간 학술적으로 연구되고 실무에서 활용되어 온 검증된 현상이다.

정리하면 이 시스템이 발견한 전략은 첫째 경제적 논리가 명확하고 둘째 전 세계적으로 검증된 현상에 기반하며 셋째 통계적으로도 우연이 아님이 확인된다. 이 세 가지가 동시에 충족될 때 "p-hacking이 아니다"라고 자신 있게 말할 수 있다.

검증 요약 — 네 가지 독립적 증거

위 네 가지 검증은 서로 독립적이다. 한 가지만 통과해도 의미가 있지만 네 가지 모두 통과한다는 것은 이 시스템이 발견한 전략이 단순한 우연(p-hacking)이 아니라 실제 시장의 패턴을 포착한 것임을 강하게 시사한다.

표20 통계적 검증 종합

검증 방법	결과	의미
OOS > IS (검증 기간 성과)	OOS/IS = 2.00	"진짜 실력" — 우연이면 불가능한 패턴
Harvey t-통계량	32.0 (기준: 3.0)	학계 최엄격 기준의 10배 초과
무작위 추출 검정	p < 0.0001	10,000 번 시도 중 0 번 — 우연 가능성 제거
경제적 논리	EPS Revision Momentum	전 세계 시장에서 검증된 현상

자료: DS투자증권 리서치센터

참고문헌

Allen, Franklin, and Risto Karjalainen. "Using Genetic Algorithms to Find Technical Trading Rules." *Journal of Financial Economics* 51, no. 2 (1999): 245-271.

Harvey, Campbell R., Yan Liu, and Heqing Zhu. "...and the Cross-Section of Expected Returns." *The Review of Financial Studies* 29, no. 1 (2016): 5-68.
<https://doi.org/10.1093/rfs/hhv059>

Holland, John H. *Adaptation in Natural and Artificial Systems: An Introductory Analysis with Applications to Biology, Control, and Artificial Intelligence*. University of Michigan Press, 1975.

Compliance Notice

- 동 자료는 기관투자가 등 제 3자에게 사전 제공한 사실이 없습니다.
 - 동 자료에 게시된 내용들은 본인의 의견을 정확하게 반영하고 있으며, 외부의 부당한 압력이나 간섭 없이 작성되었음을 확인합니다. (작성자: 김남일)
 - 동 조사자료는 고객의 투자에 참고가 될 수 있는 각종 정보제공을 목적으로 제작되었습니다. 이 조사자료는 당사의 리서치센터가 신뢰할 수 있는 자료 및 정보로부터 얻어진 것이나, 당사가 그 정확성이나 완전성을 보장할 수 없으므로 투자자 자신의 판단과 책임하에 종목 선택이나 투자시기에 대한 최종 결정을 하시기 바랍니다. 따라서 이 조사 자료는 어떠한 경우에도 고객의 증권투자 결과에 대한 법적 책임소재의 증빙자료로 사용될 수 없습니다.
 - 동 조사자료의 지적재산권은 당사에 있으므로 당사의 허락 없이 무단 복제 및 배포 할 수 없습니다.
-