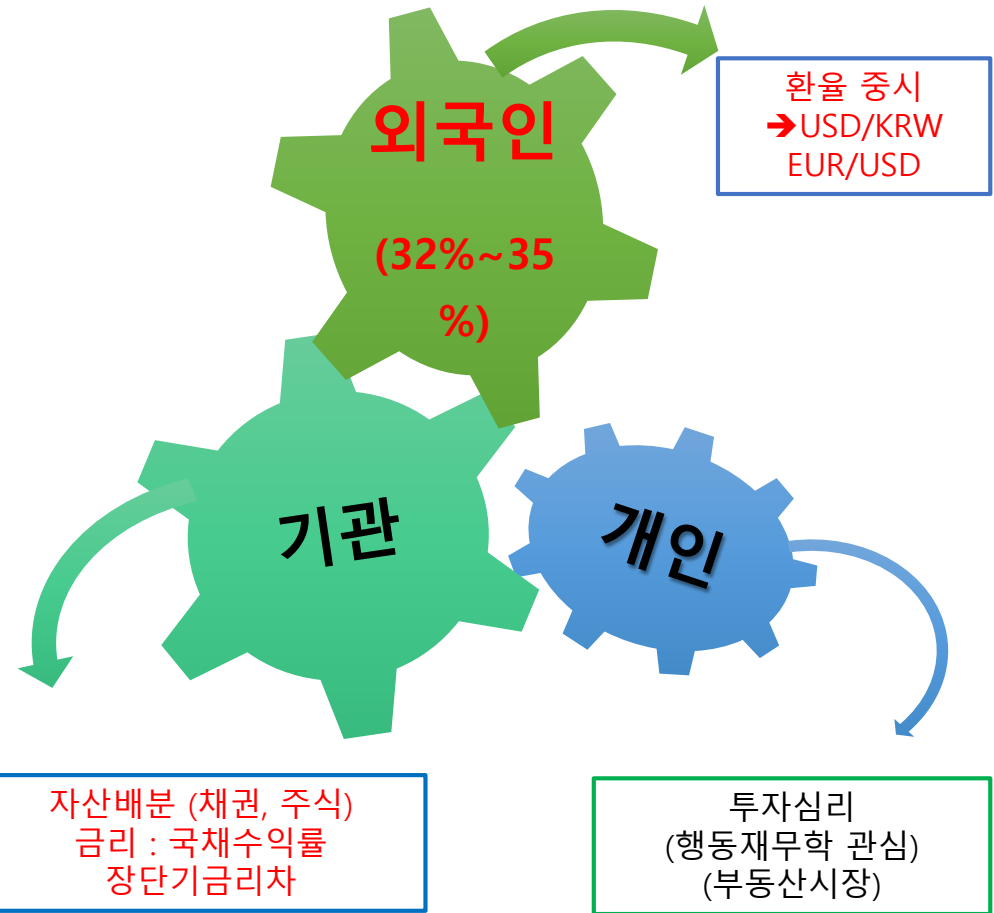
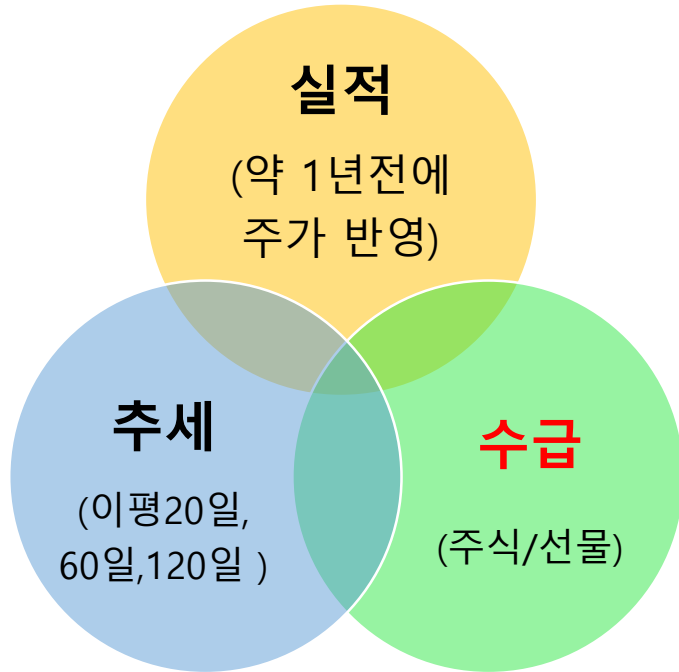


## **Domain Knowledge(5)**

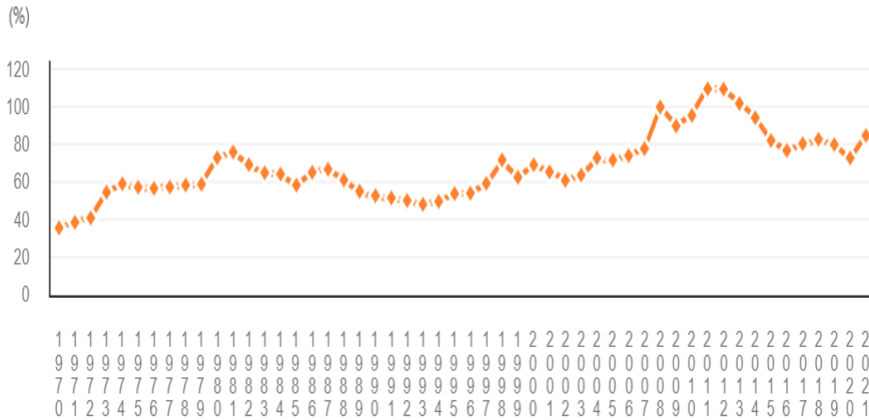
**퀀트(Quant) 투자**

# 주식 시장가격 결정 : 주가 예측



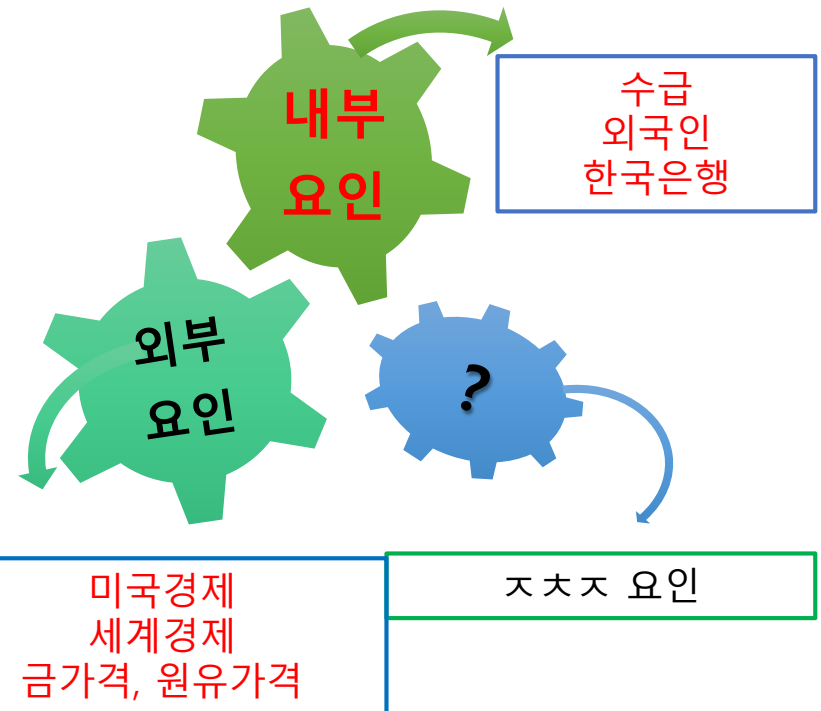
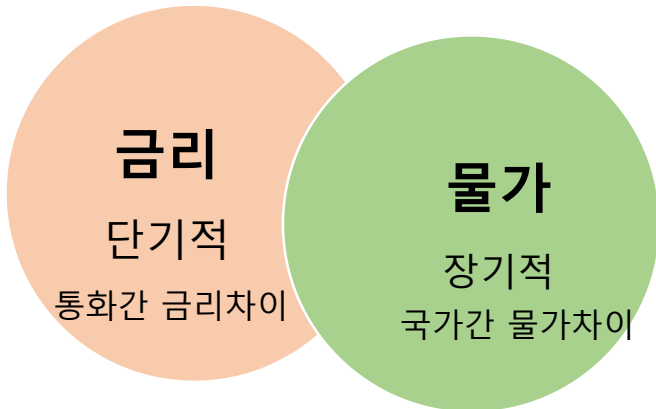
# 환율 결정

GDP 대비 수출입 비율



2020년 기준 한국의 GNI 대비 수출입 비율은 72.3%로 미국의 31.4%, 일본의 37.5%, 프랑스의 66.1%에 비해 높다.

<https://www.index.go.kr/unify/idx-info.do?idxCd=4207&clasCd=7>



# 퀀트(Quant) 투자 : 통계 vs 금융 머신러닝 vs 계량경제학

## ◆ 계량경제학

: 통계기법 + 경제, 금융, 다변량 선형회귀 ,  
: 학문적 관점에 집착(경제적 이론에 근거 해야 함)

## ◆ 금융머신러닝

: 머신러닝 알고리즘 : 고차원 공간--> 패턴 학습  
: 이론을 대체하는 것이 아니라 안내자 역할  
: 어떤 현상에 대한 어떤 특징(feature)이 예측 가능한지를 이해한 후 → 이론적 설명 수립하고 → 새로운 data set을 상대로 테스트를 해본다.

## 머신러닝 발전

머신러닝: 미국 국립연구소, 구글DeepMind Technologies Limited). 메타, 아마존, IBM, 넷플릭스, 테슬라 등

금융머신러닝 : Marcos Lopez de Prado( Cornell University )

<https://www.quantresearch.org/>

파이썬 코드 제공

[https://papers.ssrn.com/sol3/papers.cfm?abstract\\_id=1821643](https://papers.ssrn.com/sol3/papers.cfm?abstract_id=1821643)

The Probability of Backtest Overfitting

# 퀀트(Quant) 투자

Quantitative + Analyst 기반 투자

: 수치와 데이터를 바탕으로 만들어진 rule에만 근거하여 매매하는 투자법  
(예시: 저 PER, 저PBR투자)



해들리 위컴 (Hadley Wickham)([Grolemund and Wickham 2018](#))에 따르면 데이터 과학의 업무 과정

## 퀀트 투자 프로세스

파이썬(Python) Pandas 활용

1. 데이터를 수집. 정리  
재무비율 데이터, 재무제표 데이터, 투자지표 데이터
2. 데이터 가공
3. 모형 이용 : 종목선정, 백테스트(backtesting) 수행
4. 투자실행
5. 성과 평가

# 알고리즘 트레이딩(Algorithmic Trading)

- 데이터를 수집, 정리
- 재무비율 데이터, 재무제표 데이터, 투자지표 데이터
- 2데이터 가공



- 알고리즘
- 논리구조,
- 특정 매매기법

자동매매

수동매매

## 내부요인

매매가격, 거래량, 수급 매매비용 손익계산서,  
가중평균수익률

## 외부요인

경제지표 미국시장

## 특징

- 시장유동성 제공
- 지연시간에 민감
- 시간지연 문제

(투자자에게 정보 전달 걸리는 시간, 거래자 알고리즘이 정보 분석하는 데 걸리는 시간, 생성된 동작이 거래에 걸리는 시간)

# Machine Learning for Trading: From Idea to Execution

## The evolution of algorithmic strategies

Today, traders pursue a range of different objectives when using algorithms to execute rules:

- Trade execution algorithms that aim to achieve favorable pricing
- Short-term trades that aim to profit from small price movements, for example, due to arbitrage (long&short fund)
- Behavioral strategies that aim to anticipate the behavior of other market participants
- Trading strategies based on absolute and relative price and return predictions

## Use cases of ML for trading

- Data mining to identify patterns, extract features, and generate insights
- Supervised learning to generate risk factors or alphas and create trade ideas
- The aggregation of individual signals into a strategy
- The allocation of assets according to risk profiles learned by an algorithm
- The testing and evaluation of strategies, including through the use of synthetic data
- The interactive, automated refinement of a strategy using reinforcement learning

**Machine Learning for Algorithmic Trading:** Predictive models to extract signals from market and alternative data for systematic trading strategies with Python

# 알고리즘 트레이딩(Algorithmic Trading)

- **프로그램매매**

- 차익거래
- 비차익거래

- **고빈도 매매(HFT : High Frequency Trading)**,  
**초고빈도 매매(Ultra HFT)**

## 기술적 분석 Rule-based

- 추종 전략(Trend Following Strategy)
- 역추세 전략(Counter-Trend Strategy)

- **MA :SMA**
- **MACD, RSI, 볼린저밴드 기술적 지표**

## Arbitrage

- **통계적 차익거래(Statistical Arbitrage)** : Pairs Trading
- → 저평가된 종목을 매수하고, 동시에 고평가된 종목을 매도한 후 균형 상태에 이르렀을 때 청산
- → 평균 회귀(Mean Reversion) 현상 응용
- **Pair trading**
- **이벤트 차익거래**
- **팩터 인베스팅(Factor Investing)** : 모멘텀 팩터, 가치 팩터 등



# 알고리즘 트레이딩(Algorithmic Trading)

## 문제점

- BlackBox
- 시스템장애 문제
- 불공정매매 (주가조작)
- 매매비용 증가 (수수료(매수+매도), 세금(매도))

## 거래 비용 절감

- 대규모 주문을 작은 주문으로 나눠서 일정 간격으로 매매
- 주식의 유동성 문제

### ▪ 거래량 가중평균가격(VWAP; Volume Weighted Average Price)

→ 시장가격이 VWAP 위이면 오르고, 아래면 내릴 것으로 예측→ MA와 유사

### ▪ 시간 가중평균가격(TWAP; Time Weighted Average Price)

→ 시작과 종료 시점 사이의 평균 가격에 가깝게 주문을 넣어 시장의 영향을 최소화하고자 하는 것

### ▪ 거래량 비율(POV; Percentage of Volume)

→ 일정기간 주가 상승 일의 거래량과 주가 하락 일의 거래량과의 백분율로

→ 거래 주문이 완전히 체결될 때까지 미리 정의된 비율과 거래량에 따라 부분적인 주문을 지속해서 보내어 주식의 가격이 사용자가 정의한 수준에 도달할 때까지 거래량과 참여 비율을 증가시키거나 감소시키는 주문을 보내는 방법

# 알고리즘 트레이딩(Algorithmic Trading)

알고리즘을 컴퓨터 프로그램 → 필요 기술사항

1. Trading **Strategy**

2. Strategy **Backtesting**

3. Trading Software

4. Risk Management

## \*전통적 금융 + AI 기술

- 투자관리 (로보어드바이저)
- 트레이딩 서비스(종목 추천, 가격 예측) : Information Technology(정보기술) + Financial Engineering
- 플랫폼 서비스
- 리스크 평가 및 관리
- 금융 사기 탐지

### 트레이딩 서비스

- Advanced Execution Division, Algorithmic Trading Division
- 초개인화(니즈 부응) + 공학적 서비스 제공

# Fi- DATA

## Marcos Lopez de Prado 자료 수정

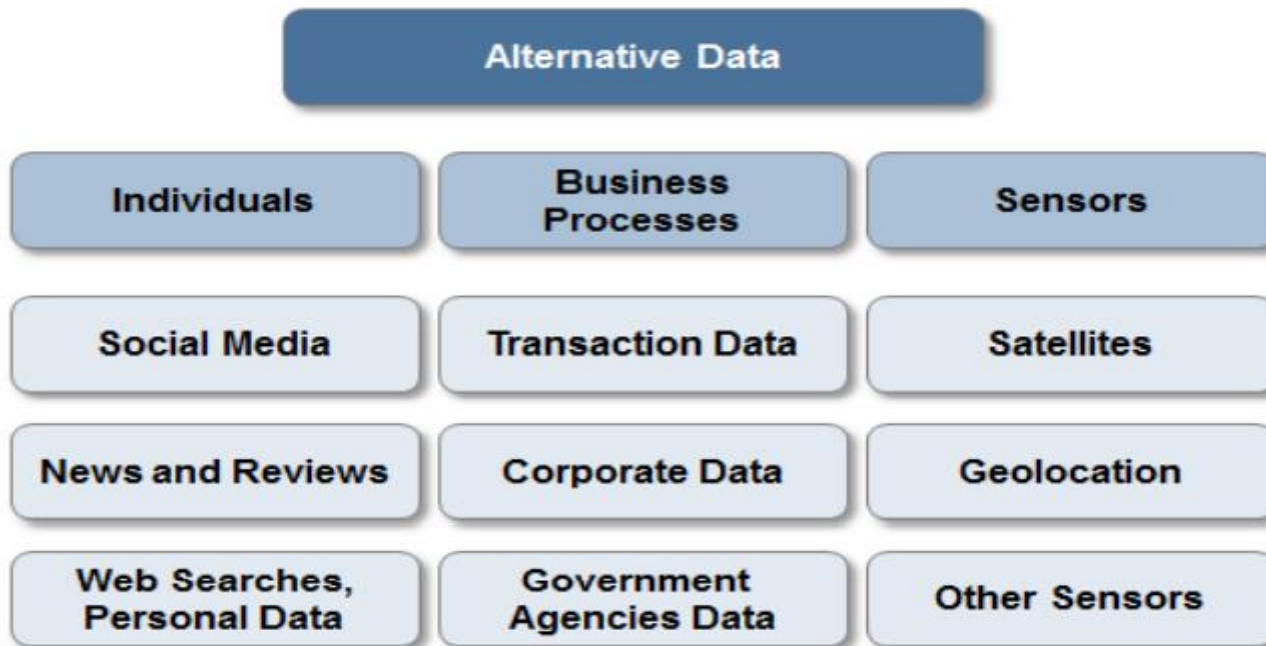
Open data				My data
Market data	F/S	analysis	alternatives	예금
주가,거래량 배당	B/S	애널리스트보고서(예측)	구글 검색어	대출
금리	I/S	신용등급	SNS	신용카드
환율	C/F	감성분석*	위성 이미지	위탁계좌
부동산가격		(파생자료)	뉴스기사	보험료.보험금

## DATA 활용 점검사항

- 신뢰성 문제
- 적시성(효용성)
- 세분화 여부
- 개인정보 위규 문제

# Fi- DATA

## Classification of Big/Alternative Data sources



Source: J.P.Morgan Macro QDS.

# Attributes of an alternative data set

Asset Class	Investment Style	Alpha (Net of Cost)	Known	Stage of Processing	Quality	Technical Aspects
Equity	Macro	Viable Stand alone	Public Free of cost	Raw	History	Frequency
Commodity	Sector Specific	Viable In a Portfolio	Well Known	Semi Processed	Outliers	Latency
Credit	Stock Specific	Not Viable	Lesser Known	Processed	Missing Values	Format
Rates	Risk Indicator	Capacity	Proprietary Not Known	Trading Signal	Methodology Transparency	Robust API
FX	Quant Signal	Orthogonality	Limited Sales Deals	Research Piece or Alert	Support Structure	Conflicts and Legal Risk



CIOs and  
Portfolio Managers

Quants and  
Data Scientists

Source: J.P.Morgan Macro QDS.

# 퀀트(Quant) 투자 : 필요데이터

- 야후 파이낸스: <https://finance.yahoo.com/>
- 네이버 금융: <https://finance.naver.com/>
- Quandl: <https://www.quandl.com/>
- tiingo: <https://www.tiingo.com/>

API를 이용한 Quandl 데이터 다운로드¶

데이터 제공업체 Quandl은 일부 데이터를 무료로 제공하며 API를 통해서 다운로드 가능

[https://www.quandl.com/api/v3/datasets/WIKI/AAPL/data.csv?api\\_key=xw3NU3xLUZ7vZgrz5QnG](https://www.quandl.com/api/v3/datasets/WIKI/AAPL/data.csv?api_key=xw3NU3xLUZ7vZgrz5QnG)

## DataReader() 함수를 데이터 다운로드¶

Quandl의 경우 무료로 얻을 수 있는 정보에 제한이 있으며, 다운로드 양에도 제한

이 방법으로 한두 종목의 데이터를 수집할 수 있지만, 전 종목의 데이터를 구하기는 사실상 불가능

→ 야후 파이낸스 역시 주가 데이터를 무료로 제공하며, pandas\_datareader 패키지의 DataReader() 함수는 해당 API에 접속해 데이터를 다운로드

## Pykrx 모듈 이용

### : 파이썬 소개

Naver(<http://www.naver.com>)와 한국 거래소(<http://www.krx.co.kr>)에서 유가 증권 데이터를 스크래핑 후 데이터 프레임으로 값을 반환

github(<https://github.com/sharebook-kr/pykrx>)를 통해 유지, 개발

# 퀀트(Quant) 투자 : 가치주 투자 사례

## Benjamin Graham : NCAV(Net Current Asset Value)전략

- NCAV(Net Current Asset Value 순유동자산)
- =유동자산 - 총부채



주식시장 환경??

- *순유동자산 > 시가총액의 1.5배 이상 높은 경우 + 흑자기업인 경우 포트폴리오 구성*

- → The Intelligent Investor REV Ed. : 안전마진margin of safety' 개념
- (책 내용)
- 투자에서 안전마진도 마찬가지다. 1달러의 가치가 있는 투자 대상을 60센트에 사는 것은 40센트로 사는 것보다는 더 위험하다. 보상의 기대는 40센트로 산 경우에 더 많이 할 수 있다.
- 즉, 가치 포트폴리오에서 보상에 대한 잠재 가능성이 클수록 위험은 더 작아진다.
- 따라서 현명한 투자자는 기업의 기본적인 가치를 추정할 수 있는 지식도 갖추어야 한다.



# 퀀트(Quant) 투자 : 성장주 투자 사례

Peter Lynch : PEG(Price Earnings to Growth rate)전략

$$PEG = \frac{PER}{3\text{년 평균 EPS증가율}}$$

해당 수치가 낮을수록 주가가 저평가돼 있거나 성장률이 높다는 것을 의미



Jim Slater 줄루 주식투자법

성장주 투자와 금융지표인 PEG(주가수익비율을 주당순이익 성장률로 나눈 것) 대중화

부와 명성을 쌓은 영국의 투자 대가

→ 좁은 영역에 특화해서 투자하라.

$$PEG = \frac{1\text{년 후 예상 PER}}{1\text{년 후 예상 EPS증가율}}$$

→  $PEG \leq 0.75$ 이면 투자하기 좋은 기업

PEG =1 종목 , PER 12배 ~20배 사이 주식 , 연간 EPS 증가율 15% ~30%

- 이상적인 성장주의 특징
  - 강력한 이익증가율 실적
  - 낙관적인 미래 전망과 예상
  - 예상 이익증가율 대비 낮은 PER
  - 주당순이익을 훨씬 초과하는 강력한 현금흐름
  - 과도하지 않은 적절한 부채비율
  - 우수한 전년도 상대 주가 실적
  - 이사진의 회사 주식 매수

## \* 가치주와 성장주

구분	가치주	성장주
판단기준	<ul style="list-style-type: none"> <li>회사의 이익이나 자산에 대한 평가가 반영되지 않아 주가가 낮은 주식</li> <li><b>꾸준한 실적과 저변동성이 특징인 '가치주'</b></li> <li>PER, PBR이 낮으면서도 꾸준히 이익을 내고 있는 기업</li> </ul>	<ul style="list-style-type: none"> <li>매출액과 이익이 매년 크게 증가하고 있거나 앞으로도 증가할 것으로 예상되는 기업</li> <li><b>미래 가능성 높게 평가받는 '성장주'</b></li> </ul>
금리변화	<ul style="list-style-type: none"> <li>금리 인상기에 주목받는 가치주</li> <li>배당수익률 매력</li> </ul>	<ul style="list-style-type: none"> <li>금리 오를수록 불리한 성장주, 빅테크기업</li> </ul>
종목 선택 기준	<ul style="list-style-type: none"> <li>주가 &lt; 내재가치</li> <li>경기침체기 : 주식시장 회복기에</li> </ul>	<ul style="list-style-type: none"> <li>성장률 (매출액, 이익)</li> <li>경기 회복기 : 기업 실적이 회복될 때</li> </ul>
장점	<ul style="list-style-type: none"> <li>안정적 투자</li> <li>낮은 변동성</li> </ul>	<ul style="list-style-type: none"> <li>높은 종목 선정이 상대적으로 쉬움</li> </ul>
단점	<ul style="list-style-type: none"> <li>고수익 달성이 어려움</li> <li>중장기 관점에서 투자</li> <li>기업가치 산정의 주관성</li> </ul>	<ul style="list-style-type: none"> <li>높은 변동성</li> <li>유동성 위기</li> </ul>

# 퀀트(Quant) 투자 : 사례

## John B. Neff의 GYP비율 전략

$$\text{GYP 비율} = \frac{\text{당기순이익 증가율(Growth)} + \text{배당수익률(Dividend Yield)}}{\text{주가수익비율(PER)}}$$

### 2 이상인 경우를 투자 기준

→ 당기순이익 증가율 대용치 : ROE

사례 : 삼성전자 (네이버 금융 참조)

ROE 11.83% ,배당수익률 2.46% → 총수익률은 14.29%,

PER10.66

GYP 비율 =  $14.29 / 10.66 = 1.34$ 배

# 퀀트(Quant) 투자 : 사례

Willam J. O'Neil CAN SLIM 법칙

The Successful Investor

- **C (Current Earnings):** 현재 주당 분기 **순이익 (증가율)**
- 분기 순이익 증가율이 20~50% 이상인 기업에 투자
- **A (Annual Earnings increases) :** **연간 순이익 증가율**
- 연간 순이익 증가율이 25% 이상, 3년 순이익 증가율이 안정적인 기업, 자기자본이익률이 17% 이상인 기업에 투자
- **N (New) :** 신제품, **경영혁신**, 신고가
- 신제품, 신경영, 신고가를 형성하는 기업에 투자
- **S (Supply and Demand) :** **수요와 공급**
- 유동주식수가 적은 종목을 공략
- **L (Leader or Laggard) :** **주도주와 소외주**
- 시장 주도주를 사고 소외주를 피한다.
- **I (Institutional Sponsorship) :** **기관 관심주**
- 기관이 관심을 갖는 종목에 투자
- **M (Market Direction) :** **시장의 방향**
- 보조지표나 시장속보 등 전문가의 견해를 의지하는 것 보다는 매일 주가지수의 움직임을 체크하며 자신만의 노하우로 시장 움직임을 파악하는 것이 중요

(참고)

- 시가총액 100억 이상
- EPS 증가율 최근3년 평균 기준 25% 이상
- 3일봉 연속 외국인 순매수 발생 또는 3일봉 연속 기관 순매수 발생
- 최근 5일봉이내 260봉 신고가 갱신
- 제외종목 : 관리종목, 거래정지종목, 정리매매종목, 불성실공시종목, 투자위험종목, 우선주

# 퀀트(Quant) 투자 : 사례

Josef Lakonishok , Andrei Shleifer , Robert Vishny

저PER + 저PBR + 저PCR

- 순위 합의 순위가 낮은 종목 선정 , 순위 상위 50개 종목 포트폴리오 구성
- 매년 리밸런싱
- : 시가총액 10분위 하위 1분위 , PBR 0.2 미만 제외

# 퀀트(Quant) 투자 : 사례

F-Score : JOSEPH D. PIOTROSKI

2000 "Value Investing: The Use of Historical Financial Statement Information to Separate Winners from Losers"

구분	내용	점수
수익성 분석(4)	전년 당기순이익이 0 이상 인가? → 회계상	+1점
	전년 영업현금흐름이 0 이상 인가? → cf	+1점
	ROA가 전년대비 증가 했는가? (추세)	+1점
	영업현금흐름이 순이익보다 높은가?	+1점
성장성분석(1)	매출총이익률이 전년대비 증가했는가?	+1점
안정성분석(3)	부채비율이 전년대비 감소했는가?	+1점
	유동비율이 전년대비 증가했는가?	+1점
	금년 <b>신주 발행을 하지 않았다?</b>	+1점
활동성 분석(1)	자산회전율이 전년대비 증가했는가?	+1점

Backtesting



- CAGR
- MDD
- Sharpe

총점 0점~9점  
**일정점수 이상 Long**

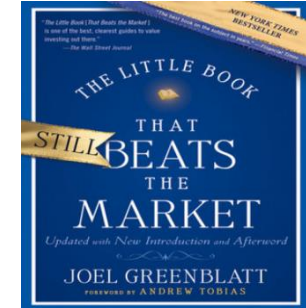
**generates a 23% annual return between 1976 and 1996,**

# 퀀트(Quant) 투자 : 사례

Joel Greenblatt

'Magic Formula

- Joel Greenblatt 헤지펀드 **Gotham Capital** 설립한 펀드매니저  
1985년 이후 2005년까지 20년 동안 연평균 수익률 40% 성과
- 가치투자자 'Magic Formula' 주목 "The Little Book that Beats the Market"(2005)  
마법공식을 사용 결과 시장평균(12.3%)과 S&P500(12.4%) 대비 **30.8%**의 복리수익률 달성



투자기준

Python 코드 별도 공유

## 1. 이익수익률(Earnings yield)

지불한 가격에 대해서 더 많이 높은 이익수익률을 가진 기업을 선택

## 2. 자본수익률(Return On Capital, ROC)

자본수익률은 회사의 자체의 성장성과 관련된 지표  
비용에 대해서 더 많은 자본수익률을 가진 기업을 선택

## 3. 종목 선정

이익수익률과 자본수익률에 따라 각 종목들의 순위(rank)를 매기고 두 값을 더하여 얻은 순위대로 종목을 선정  
Joel Greenblatt는 실험에서 상위 30개 종목 선정

## Magic Formula 적용

### 1. 주식 보유 기간

- 마법공식은 종목 선정하는 방법뿐만 아니라 선정된 종목을 매매하는 시점까지 중요하게 다루고 있다.
- 매달 특정한 매도/매수일을 정한다. (동일한 날)
- 선택된 자산은 1년(or 3년 등) 기간 동안 보유한다.

### 2. 주식선택범위

### 3. 마법공식 리스크

### (결과)

#### 1. 투자 보유기간에 대한 실증

보유기간에 차이 실험 결과 : 17년(1988년 1월 ~ 2004년 12월)

1년 기간으로 했을 때 4년에 한 번씩 시장 평균보다 좋지 못한 실적이 나타났지만,

2년 기간으로 했을 때 6번 중 1번 정도 시장 평균보다 실적이 좋지 못했고,

3년 기간으로 했을 때 총 차수의 5% 이하만이 시장 평균을 하회 하는 성과

3년 기간동안 보유 후 매매한 경우,

17년 동안 손실을 보지 않았다는 결과를 보여주고 있다.

→ 시장 평균 수익률이  $-(-46\%)$ 인 경우에도 마법공식의 최악의 수익률은  $+(+11\%)$ 였다

한편 어떤 투자 전략이 정말 이치에 맞을 경우 장기간 투자할수록 궁극적인 성공의 가능성이 높아지기 때문에, 투자 기간은 5년 또는 10년, 20년으로 하는 것이 이상적이라고 판단

#### 2) 주식 선택 범위

모두 시장 가치 기준 상위 1,000개 기업의 주식 중에서 선택했을 때의 결과로 상위 3,500개 기업에서 선택한다면 결과는 더욱 좋은 결과

3년 기간동안 보유 했을 시, 모든 포트폴리오들이 시장 평균 초과, 최악의 수익률도  $+35\%$ .

### 3) 마법공식의 리스크

위험에 대하여 장기적으로 훌륭한 성과를 내는 것으로 인정받은 index fund(ETF)(Index fund vs Active fund)와 마법공식을 비교

시장 평균을 따라가는 index fund는 모든 3년 기간들 중 12%에서 돈을 잃었지만,

마법공식은 모든 기간에서 손실을 보지 않았다.

마법공식 전략의 수익률은 시장 평균 수익률을 월등히 뛰어넘었다.

마법공식 전략은 시장 평균보다 더 작은 리스크로 더 나은 결과를 달성했다고 할 수 있다.

이것은 지난 17년의 실험기간에 한한 리스크를 평가한 것

Joel Greenblatt 역시 '미래에는 마이너스 성과를 낼 수도 있음이 거의 확실하다' 라고 했을 정도로 앞으로의 일을 보장 없다고 함..



### Joel Greenblatt 자기문답

1) 마법공식은 특정기업에만 적용 되는 것인가?

마법공식의 성과를 실험하기 위해 선택된 종목들은 모두 미국 주요 증권거래소에서 거래되는 3,500개 대기업 중에서 양호한 주식

- 최소 5,000만 달러(500억 원 이상) 이상의 시장가치 기업 → 30.8%(시장평균 12.3%)의 연간 수익률 달성
- 상위 2,500개 종목에 대해서도 마법공식을 적용 → 시장가치는 2억 달러(2,000억 원 이상) 이상 기업, 마법공식 적용 선택한 30개의 기업 포트폴리오를 구성하여 연간 23.7%(시장평균 12.4%)의 수익률 달성
- 상위 1,000개 종목(시장가치 10억 달러(1조 원 이상) 이상)에 대해서도 실험 → 결과 연 평균 22.9%(시장평균 11.7%)의 수익률

2) 우연히 발생한 염가 주식 30개를 발견했던 것이라면?

이전까진 믿을 수 없이 싼 소수의 주식들을 발견할 수 있어 마법공식이 작동할 수 있었지만 미래에는 그렇지 못하다면 마법공식은 무용지물이 될 것이다.

→ 이를 반박하기 위해 상위 2,500개 종목에 대한 실험을 재구성 실험.

먼저, 2,500개 기업을 등수에 따라 10개의 그룹(250개의 종목으로 구성)으로 나누고, 각각의 그룹을 하나의 포트폴리오로 구성하여 1년 동안 보유하고 그 결과를 계산 후 그 결과, 마법공식은 상위 30개 기업 주식에만 적용되는 것이 아님을 보여주었다.

마법공식은 순서에 따라 적용되는 것처럼 보였다.

공식이 매긴 등수 순서대로 좋은 성과를 내기 때문에 투자는 최상위 30개 기업의 주식에만 한정되지 않는다.

최고 등수를 받은 그룹의 주식들이 모두 좋은 성과를 내기 때문에 높은 성과를 내는 주식에 대한 선택의 폭도 넓다는 장점 또한 가지고 있다.

시장평균: 12.4% 30개 선택: 23.7% 그룹1: 17.9% 그룹2: 15.6% ,그룹3: 14.8%,그룹4: 14.2%,,,그룹5: 14.1%

그룹6: 12.7% , 그룹7: 11.3% , 그룹8: 10.1% , 그룹9: 5.2% , 그룹10: 2.5%

### MAGIC FORMULA INVESTING

[Welcome](#)[About the Book](#)[How it Works](#)[Stock Screener](#)[FAQs](#)

#### Screen for top-ranked stocks

1

Use the Stock Screener to select the top-rated stocks from the S&P CompuStat database. Choose the number of stocks to view, and choose the size of the companies you want in the list. Choosing more companies leads to greater diversification, and choosing larger companies generally leads to less volatility. Eliminate any companies you do not want to own for any reason; however, you should keep at least 20 stocks in an effort to properly manage risk.

#### Buy them

2

Use a cost effective way to purchase the stocks. If the amount you are investing represents a large percentage of your long-term investment portfolio, you may want to consider making multiple portfolio purchases over a 12 month period.

#### Hold for 1 year - then sell

3

In our opinion it is good idea to keep your stocks for approximately one year. The system is designed to maximize your after-tax returns; therefore, you want any gains to become long-term by holding them for a year, and you want to sell any losses before the one year holding period is reached.

#### Go back to - Step 1

4

Once you have sold any gains and any losses, select and purchase a new portfolio of screened stocks.

1. 투자자금 규모 설정 →

2. 투자대상 기업규모 설정 →  
(대형주, 중형주 위주)

3. 자본수익률 rank 매기기 →  
→ Proxy (총자산수익률: ROA)

4. 이익수익률 rank 매기기 →  
→ Proxy (주가수익배율 PER)

5. 자본수익률과 이익수익률 rank 합 값이 낮은 rank를 다시 매긴다. →

6. Rank 이 낮은 5개~7개기업을 선정하여 분할매수 (20~30% 매수) →

7. 2-3M 간격 반복 매수(long)하여 총 100% 매수

8. 매수 후 1년간 보유 후 매도 (short) →

9. 매도 이후 다시 종목 선정 과정 반복

- Python 코드는 별도 공유
- (ML.DLTrading).

<https://www.magicformulainvesting.com/>

# 퀀트(Quant) 투자 : 사례

## \*슬로안 비율

- University of Pennsylvania Richard Sloan이 1996년에 실시한 한 연구
- (1962년부터 2001 년까지)
- Sloan 비율이 가장 낮은 주식을 매수하고 Sloan 비율이 가장 높은 주식을 매도

## Sloan 비율 공식

$$\frac{\text{Net Income} - \text{Operating Cash Flows} - \text{Investing Cash Flows}}{\text{Total Assets}}$$

- 안전 지대 : -10 % ~ 10 % 사이
- 경고 영역 : -25 % ~ -10 % 또는 10 % ~ 25 %
- 위험 구역 : -25 % 미만이거나 25 %보다 큰 것

# \* Market microstructure

- 시장 미시구조(market microstructure)는 금융시장에서 거래와 가격 형성이 일어나는 과정과 메커니즘을 연구하는 분야
- 자산의 가격이 어떻게 결정되고 시장 효율성이 어떻게 영향을 받는지에 대해 설명하는 것이다.
- 시장 미시구조는 거래 규칙, 시장 참여자, 정보의 흐름 등이 자산 가격과 시장 효율성에 어떤 영향을 미치는지 이해하는 데 중점을 둔다.

## 1. 시장 참여자

딜러와 브로커, 외국인, 기관 및 개인 투자자

## 2. 주문 유형과 실행

주문 유형: 지정가 주문나 시장가 주문과 같은 다양한 주문 유형이 거래 방식에 영향

주문 실행: 주문 매칭의 우선순위(가격, 시간 등)가 거래 과정에 중요한 역할

## 3. 매수-매도 스프레드

두 가격의 차이를 매수-매도 스프레드는 유동성 및 거래 비용의 척도

## 4. 가격 발견

시장 미시구조는 정보가 거래를 통해 자산 가격에 어떻게 반영되는지를 연구하는 분야

가격 발견의 속도와 효율성은 거래량, 시장 구조, 정보의 흐름에 의해 영향을 받는다.

## 5. 시장 마찰 및 거래 비용

거래 비용: 거래와 관련된 수수료, 매수-매도 스프레드, 슬리피지(slippage) 등은 시장 효율성과 투자자의 행동에 영향을 미친다.

시장 마찰: 주문 실행의 지연, 공매도 제한 등과 같은 요소들이 거래의 효율성에 영향을 줄 수 있다.

## 6. 정보 비대칭성

정보 보유자 vs 비정보 보유자: 일부 거래자들이 다른 거래자들보다 더 많은 정보를 가질 때 발생하는 문제로, 정보 비대칭성은 거래 행동과 시장 결과에 영향을 미친다.

역선택: 정보 보유 거래자들은 더 많은 정보를 바탕으로 거래하기 때문에, 시장 조성자들은 정보 보유자와의 거래로 인한 손실을 줄이기 위해 매수-매도 스프레드를 조정한다.

## 7. 고빈도 거래(HFT)

고빈도 거래는 알고리즘과 매우 빠른 거래 속도를 사용해 가격 차이를 이용하는 방식으로, 유동성을 제공하는 동시에 시장 안정성에 대한 우려를 일으킨다.

## 9. 유동성과 시장 효율성

유동성: 자산을 가격에 큰 영향을 미치지 않고 쉽게 사고 팔 수 있는 능력을 말합니다. 시장 미시구조 분석은 유동성에 영향을 미치는 요인에 중점을 둔다.

시장 효율성: 효율적인 시장에서는 모든 정보가 가격에 반영됩니다. 시장 미시구조는 시장이 정보를 자산 가격에 얼마나 효율적으로 반영하는지를 연구한다.

# \* Market microstructure

## 연구 사례 : Continuous Auctions and Insider Trading - Econometrica John Kyle(1985)

### 역사

- 1985년 존 카일(John Kyle)이 제안한 **Kyle 모형**은 금융 시장의 시장미시구조를 설명하는 대표적인 모델
- 이 모형은 비대칭 정보 하에서의 거래 메커니즘과 가격 형성을 다루며, 이후 시장미시구조 연구와 실증 분석의 기반이 되었다.
- 현재까지도 정보의 흐름과 가격 형성을 분석하는 중요한 틀로 사용되고 있다.

### 모형 내용

Kyle 모형은 정보의 비대칭성이 시장에서 가격 변화에 어떻게 영향을 미치는지를 설명하며, 특히 정보 거래자와 유동성 공급자 간의 상호작용을 중심으로 분석한다.

이 모형에서는 시장 참여자를 세 가지로 나눕니다.

- **첫 번째는 내부자(정보 거래자)**로, 특정 자산에 대한 사적인 정보를 가지고 있어 가격 예측에 유리한 위치에 있다.
- **두 번째는 유동성 거래자(noise trader)**로, 정보 없이 단순히 자금의 필요에 따라 거래하는 사람들이다.
- **세 번째는 시장조성자(market maker)**로, 거래 주문을 받아들이며 가격을 결정하는 역할을 한다. 시장조성자는 내부자와 유동성 거래자의 주문을 관찰하지만, 그 주문의 의도를 알 수 없다.

Kyle 모형의 핵심은 가격 형성 방정식에서 내부자의 거래량이 가격에 미치는 영향을 통해 정보가 어떻게 가격에 반영되는지를 설명하는 것이다.

**내부자**는 자신의 정보가 시장에 미치는 영향을 최소화하며, 점진적으로 거래하여 이익을 극대화하려 한다.

반면, **시장조성자**는 거래량을 통해 내부자의 존재와 그가 가진 정보를 추론하려 한다.

**결과적으로, 정보는 시장에서의 거래를 통해 점차 가격에 반영되며, 시간이 지남에 따라 가격은 점점 효율적인 수준에 도달하게 된다.**

### 결론

- Kyle 모형은 정보의 점진적인 유입과 가격 반영 과정을 수량화하여, 정보의 비대칭성이 가격 형성에 미치는 영향을 정량적으로 분석할 수 있게 해준다.
- 이를 통해 금융 시장에서 정보 거래자의 전략과 그로 인한 시장 가격의 변화 과정을 이해하는 데 중요한 기여를 했다.
- 이 모형은 정보가 거래를 통해 시장 가격에 반영되는 과정을 설명함으로써, 시장의 효율성을 높이는 데 중요한 역할을 합니다.

## \* Market microstructure

연구 사례 : One Security, Many Markets: Determining the Contributions to Price Discovery Joel Hasbrouck  
Source: The Journal of Finance, Vol. 50, No. 4 (Sep., 1995)

1995년 Joel Hasbrouck) 발표한 논문으로 금융 시장의 가격 발견 과정을 다루는 중요한 연구  
이 논문은 동일한 증권이 여러 시장에서 동시에 거래될 때 각 시장이 가격 형성에 기여하는 정도를 분석하고, 이를 정량화하는 방법을 제시하였다.

이 연구는 자산의 본질적 가치를 반영하는 시장의 역할을 평가하는 데 중요한 기여를 했다.

### 모형 내용

- Hasbrouck 모형에서는 가격 발견(price discovery)을 자산의 "진정한 가격"에 대한 정보를 반영하는 과정으로 정의하며, 복수의 거래소나 시장에서 동일한 자산이 거래될 때 각 시장이 이 과정에 어떻게 기여하는지를 분석한다.
- 특히, 각 시장에서 거래되는 자산의 가격이 정보를 얼마나 빠르고 정확하게 반영하는지를 실증적으로 측정하기 위해 고빈도 거래 데이터를 사용한다. 이 모형은 여러 시장에서의 거래 데이터를 바탕으로 각 시장의 기여도를 평가하고, 어떤 시장이 정보에 더 민감하고 중요한 역할을 하는지를 밝히고자 한다.

### 주요 내용

- Hasbrouck는 이 연구에서 각 시장의 가격 발견 기여도를 측정하기 위해 "정보적 효율성"과 "상대적 기여도"를 분석한다. 이를 위해 시간 가중 평균 가격(TWAP)과 같은 통계적 도구를 활용해 시장 간의 가격 차이와 그 정보 반영 속도를 정량적으로 평가한다. 이 과정을 통해 특정 시장이 얼마나 빠르게 정보에 반응하는지를 측정하고, 이를 통해 가격 형성에 있어 중요한 시장을 식별한다.
- **정보적 효율성:**
  - 정보적 효율성은 시장이 새로운 정보를 얼마나 빠르고 정확하게 가격에 반영하는지를 의미한다.
  - 각 시장에서의 가격 변화가 해당 자산의 진정한 가치에 얼마나 근접하게 반영되는지를 평가하는 것이다.
  - 정보적 효율성이 높은 시장은 새로운 정보에 신속하게 반응하여 그 정보를 가격에 반영함으로써 가격 왜곡을 최소화하고, 시장의 효율성을 유지한다.
- Hasbrouck는 정보적 효율성을 통해 각 시장이 가격 발견 과정에서 중요한 역할을 하는지를 정량적으로 평가했다.
- **상대적 기여도:**
  - 상대적 기여도는 여러 시장이 동일한 자산의 가격 발견 과정에 기여하는 정도를 비교하는 개념이다.
  - 각 시장에서 발생하는 거래와 가격 움직임을 바탕으로, 각 시장이 전체 가격 형성 과정에 얼마나 기여하는지를 측정한다. Hasbrouck는 이를 위해 분산 분해 방법 등을 활용하여 각 시장의 가격 발견 기여도를 정량화하였다. (다음장 참조)
  - 상대적 기여도가 높은 시장은 해당 자산의 정보적 가치 반영에 있어 중요한 역할을 하며, 투자자들에게 더 신뢰할 만한 정보를 제공할 가능성이 크다.

## \* Market microstructure

연구 사례 : One Security, Many Markets: Determining the Contributions to Price Discovery Joel Hasbrouck  
Source: The Journal of Finance, Vol. 50, No. 4 (Sep., 1995)

### 정보적 효율성

정보적 효율성은 가격  $P_t$ 가 새로운 정보  $I_t$ 를 반영하는 정도를 수학적으로 설명

$$P_t = \lambda I_t + \epsilon_t$$

$\lambda$ : 정보가 가격에 얼마나 빠르게 반영되는지 나타내는 계수  
 $\epsilon_t$ : 시장에서의 오차

정보적 효율성이 높을수록  $\lambda$ 는 커지고  $\epsilon_t$ 는 작아진다.

### 상대적 기여도

상대적 기여도는 여러 시장이 가격 형성 과정에서 기여하는 정도

$$P_t = \sum_{i=1}^N w_i P_{t,i} + \epsilon_t$$

$P_{t,i}$ : 시장  $i$ 에서의 가격

$w_i$ : 각 시장의 상대적 기여도(가중치)

$w_i$ 가 클수록 해당 시장이 가격 발견에 더 크게 기여한다는 의미이다.

# \* Market microstructure

## 고빈도 매매 (High-Frequency Trading, HFT)

초단타 매매를 기반으로 하며, 고성능 컴퓨터와 정교한 알고리즘을 사용하여 매우 짧은 시간 안에 금융 상품을 사고파는 트레이딩 방식  
이는 알고리즘 트레이딩의 한 형태로, 특히 속도와 빈도에 초점을 맞추고 있다.

### 특징

- 초단타 매매: 거래는 밀리초(1/1000초) 또는 마이크로초(1/1,000,000초) 단위로 이루어진다.
- 자동화된 의사결정: 인간 개입 없이 알고리즘이 매매를 실행
- 투자 전략은 사전에 프로그래밍된 규칙에 따라 결정된다.
- 고빈도 거래량: 하루에도 수천 건에서 수백만 건의 거래가 이루어질 수 있다.
- 거래 속도 경쟁: 네트워크와 서버의 물리적 위치(거래소 근처에 서버를 두는 콜로케이션)가 거래 속도에 영향

### 사용 기술

- 알고리즘: 시장 데이터를 분석하고 매매 신호를 감지하는 핵심 소프트웨어.
- 저지연 네트워크: 거래 실행 속도를 최적화하는 초고속 네트워크.
- 머신러닝과 AI: 시장 변화를 학습하고 거래 전략을 개선.
- 고성능 하드웨어: 초고속 컴퓨팅이 가능한 서버와 인프라.

### 장점

- 시장 효율성 증대: 가격 차이를 줄이고 시장의 정보 비효율성을 제거.
- 유동성 공급: 대량의 거래를 통해 시장에서 매수와 매도의 수요를 맞추는 역할.
- 낮은 스프레드: 거래 비용 감소 효과로 이어질 수 있음.

### 단점과 논란

- 시장 변동성 증가: 알고리즘의 오류나 과도한 거래로 인해 시장 변동성이 커질 가능성.
- 공정성 문제: 일반 투자자들이 접근할 수 없는 기술적 우위를 활용하여 부당한 이익을 얻는다는 비판.
- 플래시 크래시(Flash Crash): 고빈도 매매가 급격한 시장 붕괴를 초래할 수 있음.

고빈도 알고리즘 매매의 데이트레이딩 성과 분석 Day Trading Performance in High Frequency Trading  
한국증권학회 우민철 (한국거래소) 2024.2



## \* Market microstructure

### 고빈도 매매 전략

1. **시장 조성 (Market Making)**: 매수와 매도 주문을 동시에 제공하여 스프레드(차이)에서 이익을 얻음.
2. **차익 거래 (Arbitrage)**: 동일한 상품이 다른 시장에서 다른 가격으로 거래될 때 차익을 이용.
3. **뉴스 기반 매매**: 뉴스 이벤트를 신속히 분석하고 그에 따라 거래.
4. **모멘텀 전략**: 가격 추세를 따라 매수 또는 매도.

### 일중 매매 (Day Trading)

일중 매매는 하루 동안 거래를 종료하는 방식으로, 하루 안에 포지션을 열고 닫아 가격 변동에서 수익을 얻는 매매 전략 - 주요 특징

- **거래 시간**: 일중 매매자는 보통 하루 동안 거래를 하고, 그날 거래 종료 시점까지 모든 포지션을 청산한다. 즉, **하루 내에 포지션을 유지하지 않고** 다음 날까지 보유하지 않는 것이 원칙이다.
- **시장 변동성 이용**: 일중 매매자는 주로 자산의 일중 변동성을 활용해 수익을 추구한다. 경제 뉴스, 기업 실적 발표, 중요한 경제 지표 등 다양한 이벤트가 가격 변동을 유발할 수 있으며, 이를 통해 수익 기회를 찾는다.
- **기술적 분석**: 일중 매매는 보통 **기술적 분석**에 의존한다. 차트, 지표, 패턴 등을 사용해 자산의 가격 움직임을 예측하고, 빠른 결정을 내리기 위해 정교한 트레이딩 도구를 사용한다.
- **짧은 시간의 거래**: 일중 매매자는 일반적으로 포지션을 **몇 분에서 몇 시간** 동안만 유지하며, 하루 동안 여러 차례 거래를 진행할 수 있다. 따라서 신속한 의사결정과 적절한 리스크 관리가 필수적이다.
- **레버리지 사용**: 일중 매매자는 종종 레버리지를 활용하여 자본 대비 높은 수익을 추구한다. 레버리지를 사용하면 수익이 커질 수 있지만, 동시에 손실의 위험도 커지기 때문에 위험 관리가 매우 중요하다.

## \* Market microstructure

### 고빈도 매매와 일중 매매의 비교

구분	고빈도 매매 (High-Frequency Trading, HFT)	일중 매매 (Day Trading)
거래 주기	초단위에서 밀리초, 마이크로초 단위 거래	분에서 시간 단위의 거래
거래 방식	자동화된 알고리즘 및 컴퓨터 프로그램 이용	트레이더의 수동 결정
목표	작은 가격 차이를 빠르게 이용해 이익 축적	하루 동안 변동성을 이용한 수익 창출
주체	금융 기관, HFT 전문 회사	개인 투자자 및 소규모 기관 투자자
주문 유지 시간	몇 밀리초에서 수초	수 분에서 몇 시간, 하루 내 종료
거래 빈도	하루 수천 ~ 수백만 건의 거래	하루 수 차례에서 수십 차례의 거래
주요 전략	시장 메이킹, 차익 거래, 주문 취소 등	스캘핑, 모멘텀 트레이딩, 기술적 분석
레버리지 사용	상대적으로 낮음	자주 사용하여 수익을 극대화
유동성 기여도	유동성 제공 역할	유동성에는 제한적 기여
기술 요구 사항	초고속 컴퓨터, 저지연 네트워크 등 고성능 인프라	거래 플랫폼, 차트 도구 등 일반적 기술
리스크 관리	빠른 속도의 주문 조정으로 리스크 관리	손절매 설정, 자본 배분 등 전략적 리스크 관리

# \* Market microstructure

## Market Making 전략

### 시장 조성(Market Making)?

- 시장 조성은 금융 시장에서 유동성을 제공하고, 매수와 매도 가격을 지속적으로 제시하여 거래가 활발히 이루어질 수 있도록 돕는 전략
- 이 역할을 수행하는 주체를 시장 조성자(Market Maker)라고 하며, 이들은 시장에서 일정한 수익을 목표로 매수와 매도 주문을 지속적으로 제시하여 자산의 매매를 원활하게 만든다.

### 시장 조성자의 역할

- 유동성 제공
- 가격 안정화: 매수와 매도 사이의 가격 차이인 스프레드(Spread)를 유지함으로써 가격이 지나치게 변동하는 것을 방지
- 매수-매도 스프레드로 인한 수익:  
예를 들어, 매수 가격이 100원이고 매도 가격이 102원이라면, 시장 조성자는 매도자에게 100원에 매수하고 매수자에게 102원에 매도하여 그 차이인 2원이 수익이 된다.

### 시장 조성의 메커니즘

- 시장 조성자는 거래를 빠르게 성사시키기 위해 항상 매도 호가와 매수 호가를 제시한다.
- 주문 제시: 시장 조성자는 매수자와 매도자가 자산을 손쉽게 거래할 수 있도록 특정 자산의 매수와 매도 주문을 동시에 제출한다. 이를 통해 매도자와 매수자가 바로 거래를 할 수 있는 환경을 제공한다.
- 리스크 관리: 시장 조성자는 계속해서 자산을 매수하거나 매도하기 때문에, 보유 자산의 재고 위험(Inventory Risk)을 관리하는 것이 중요하다. 이는 가격 변동으로 인한 손실 위험을 의미하며, 시장 조성자는 이를 줄이기 위해 포지션을 적절히 조정하거나 가격을 조정해 스프레드를 변경한다.
- 거래량과 유동성 관찰:
- 시장 조성자는 시장의 거래량과 유동성을 관찰하며, 필요에 따라 호가를 조정하여 시장에서의 이익을 극대화하거나 손실을 줄인다. 이 과정에서 시장의 가격 움직임과 주문 흐름을 모니터링하고, 이에 따른 가격 변동에 대응하여 유동성을 유지한다.

[https://github.com/freejyb/KDT/퀀트투자/Market microstructure](https://github.com/freejyb/KDT/퀀트투자/Market%20microstructure)

# \* Market microstructure

## 차익거래(Arbitrage)

- 차익거래는 시장 간 또는 상품 간의 가격 차이를 이용하여 무위험 이익을 얻는 투자 전략
- 특정 시장에서 싸게 사서 다른 시장에서 비싸게 파는 방식으로 이루어진다.

## 주식시장 + 파생상품 시장 : 차익거래

### 1. 주식 현물-선물 차익거래(Stock Index Arbitrage)

설명: 주식시장에서의 현물 가격과 파생상품(주가지수 선물) 간의 가격 차이를 이용한 차익거래

원리: 주가지수 선물 가격은 이론적으로 현물 주식 가격에 기반하여 계산됩니다.

이론 선물 가격 = 현물 가격  $\times$  (1 + 무위험 금리 - 배당 수익률)

실제 선물 가격이 이론 가격보다 높으면 선물을 매도하고 현물을 매수하며, 반대의 경우 선물을 매수하고 현물을 매도

### 2. 옵션 차익거래(Option Arbitrage)

설명: 주식 옵션과 기초 주식 간의 가격 차이를 이용한 거래입니다.

종류: 풋-콜 패리티 차익거래(Put-Call Parity Arbitrage)

옵션의 이론 가격과 실제 시장 가격이 불일치할 때 거래.

예: 콜옵션 가격 + 행사 가격(현재가치) = 풋옵션 가격 + 주식 가격.

커버드 콜 차익거래(Covered Call Arbitrage)

주식을 매수하고 동시에 해당 주식의 콜옵션을 매도하여 무위험 수익을 창출.

예시:

삼성전자 콜옵션의 가격이 고평가되었다면 콜옵션을 매도하고 기초 주식을 매수.

### 3. 배당 차익거래(Dividend Arbitrage)

설명: 특정 주식의 배당락일(배당 수익을 받을 권리가 없어지는 날짜)과 관련된 차익거래

전략: 배당락일 이전에 주식을 매수하고 배당금을 받은 후, 배당락일 이후 주식을 매도.

효과: 배당 수익과 주식 가격 변동 간의 차이를 이용해 수익 창출.

### 4. Volatility Arbitrage

설명: 주식 옵션의 내재 변동성(시장 가격에 반영된 변동성)과 실제 변동성(기초 자산 변동성) 간의 차이를 이용.

전략: 내재 변동성이 실제 변동성보다 높으면 옵션 매도.

반대로 내재 변동성이 낮으면 옵션 매수.

# \* Market microstructure

## 차익거래(Arbitrage)

## 주식시장 이용한 차익거래

### 1. 시장 간 차익거래(Cross-Market Arbitrage)

설명: 동일한 주식이 여러 시장에서 거래되는 경우, 시장 간 가격 차이를 이용한 거래.

예시:

삼성전자가 한국(KRX)과 미국(ADR, 미국 예탁증서)에서 동시에 거래되고 있을 때, 두 시장 간 환율을 고려한 가격 차이가 발생하면 이를 이용.

전략: 가격이 낮은 시장에서 매수하고, 가격이 높은 시장에서 매도.

조건: 거래 비용 및 환율 차이를 충분히 상쇄할 수 있는 가격 차이가 필요.

### 2. Pair Trading

설명: 동일한 산업군에 속하거나 상관관계가 높은 두 주식 간의 가격 차이를 이용한 전략

전략: 상관관계가 높은 두 주식 간 가격 차이가 정상 범위를 벗어났을 때, 고평가된 주식을 매도하고 저평가된 주식을 매수.

예시: 현대차와 기아의 주가가 비슷하게 움직이는 경우, 현대차가 상대적으로 고평가되었다면 현대차를 매도하고 기아를 매수.

장점: 시장 전체의 방향성과 무관하게 수익을 얻을 수 있음.

리스크: 두 주식의 가격 차이가 예상과 다르게 벌어질 경우 손실.

### 3. 지수 차익거래(Index Arbitrage)

설명: 특정 주가지수를 구성하는 주식들의 실제 가중평균 가격과 해당 주가지수의 움직임 간의 차이를 이용.

전략: 지수를 구성하는 개별 주식들을 매수/매도하여 지수 자체와의 가격 불균형을 활용.

예시: 코스피200 지수를 구성하는 개별 주식의 가중평균 가격이 실제 지수 가격보다 낮으면 주식을 매수하고, 반대의 경우 매도.

### 주식시장 차익거래의 실행 요건

- 실시간 데이터 및 분석 능력
- 수수료, 세금, 슬리피지(Slippage) 등 거래 비용이 차익거래 수익을 잠식할 수 있음.
- 유동성 높은 시장 : 충분한 거래량과 유동성이 보장되지 않으면 거래 실행이 어려움.
- 빠른 거래 시스템: 시장 간 가격 차이는 짧은 시간 안에 사라지므로 신속한 거래가 필수.

### 장점

- 파생상품을 사용하지 않으므로 구조가 단순.
- 시장 간 비효율성을 이용하여 상대적으로 안정적인 수익 가능.

### 한계

- 거래 비용이 높을 경우 차익거래의 이익이 감소.
- 유사한 주식 간의 관계가 갑작스레 변화할 경우 손실 위험.
- 고빈도 거래가 아닌 경우 실행 속도에서 밀릴 수 있음.

# \* Momentum effect : Momentum Anomaly

## Momentum effect

모멘텀 효과는 특정시점의 주식수익률(stock return)과 그 이후의 주식수익률 간의 상관성(correlation)을 의미하는데, 그 형태가 양(+) 또는 음(-)으로 존재하는 효과  
모멘텀 효과가 양(+)일 경우 과거에 수익률이 높은 주식은 미래에도 계속 높은 수익률을 보이거나, 음(-)일 경우 과거에 수익률이 높은 주식은 미래에는 반대로 낮은 수익률을 나타낼 것을 의미한다.

## market anomalies

- **Momentum investment strategy** : 과거에 높은 수익률을 기록하였던 승자(Winner)포트폴리오들이 계속해서 미래에도 좋은 성과를 기록하는 관성(Momentum)을 갖고 있을 경우 승자를 매입 하고 패자(Loser)를 매도하는 투자전략
- **Contrarian investment strategy** : 과거에 저조한 수익률을 기록하였던 포트폴리오가 가격이 상승하여 높은 수익률을 기록할 경우 패자 포트폴리오를 매입하고 승자 포트폴리오를 매도하는 전략

## Eugene F. Fama (EMH : Efficient Market Hypothesis)

금융경제학에서 모든 시장참여자가 완벽한 정보를 가지고 있을 때 자산가격이 균형에 도달한다는 가설

### Efficient Capital Markets: A Review of Theory and Empirical Work (1970)

Fama, E.F., "Efficient Capital Markets II," Journal of Finance 46, 1991

Fama, E.F. and French, K.R., "The Cross-Section of Expected Stock Returns," Journal of Finance 47, 1992

Fama, E.F. and French, K.R., "Common Risk Factors in the Returns on Stocks and Bonds," Journal of Financial Economics 33, 1993,

**Shiller, R.J.**, "Do Stock Prices Move Too Much To Be Justified by Subsequent Changes in Dividends?," American Economic Review 71, 1981, pp. 421~436.

100년간 주식가격은 그것의 형성에 근본을 이루는 배당금의 변화로는 설명할 수 없을 만큼 지나친 변동성을 가진다  
주식가격이란 펀더멘탈(fundamental)로 간주되는 배당이 아니라 심리적 요인 등에 의해 크게 영향을 받았다는 것

## \* Momentum effect : Momentum Anomaly

**Returns to Buying Winners and Selling Losers:** Implications for Stock Market Efficiency Author(s): **Narasimhan Jegadeesh and Sheridan Titman** Source: The Journal of Finance, Vol. 48, No. 1 (Mar., 1993)

모멘텀 투자가 주가의 상승 또는 하락 흐름을 따라 투자하는 전략이 장기적으로 성과를 낼 수 있다는 가설을 탐구  
미국시장에서는 강력하게 **모멘텀 전략이 유효하게** 나타나고 있음

1. 미국 주식시장을 대상으로 과거3개월 ~ 12개월의 구성기간 수익률을 기초로 Winners portfolio(상위 %)와 Losers portfolio(하위 %) 구성 한 후
2. 3개월 ~ 12개월의 보유기간별 수익률을 비교하였다  
(결과) 과거 수익률이 높은 W.P의 성과가 과거수익률이 낮은 LP포트폴리오 수익률보다 높음을 보였다.

**DeBont and Thaler(1985, 1987)**

DeBondt, W. and R. Thaler(1985), "Does the Stock Market Overreact?," *Journal of Finance*, 40

과거 3년 ~ 5년 동안 수익률이 높았던 주식은 이후 수익률이 떨어진다는 것을 발견하였다.  
이들은 이와 같이 주식수익률이 평균에 회귀한다(mean revert)는 수익률의 역전 현상(reversal)을 이용하여 **역행투자전략**의 유효함을 주장하였다

Conrad, J. and G. Kaul(1998)

"An Anatomy of Trading Strategies," *Review of Financial Studies*, 11,  
미국 주식시장을 대상으로

**단기 및 중기에는** 모멘텀 투자전략이 유리하나,  
**장기에는** 역행투자전략이 유리함

## \* Momentum effect : Momentum Anomaly

모멘텀 효과를 이용한 투자전략의 성과에 관한 연구 (2011)

Performance of Investment Strategies by Using Momentum Effect in Korea Stock Market

첫째, 주가 모멘텀에 따라 구성된 포트폴리오의 성과 분석결과, 전반적으로 패자 포트폴리오의 수익률이 승자 포트폴리오보다 높게 나타났다. 그리고 이에 기초한 역행투자전략(contrarian investment strategy)의 성과는 포트폴리오 구성기간과 보유기간별로 다소 차이는 있으나 대체로 유의한 양(+)의 값으로 확인되었다.

둘째, 장부가치/시장가치 비율(B/M)의 효과를 통제한 포트폴리오를 대상으로 역행투자전략의 성과를 분석한 결과, 앞서 확인된 역행투자의 성과가 거의 나타나지 않았다.

- **포트폴리오 구성기간(J)** : 주가모멘텀의 측정에 이용되는 과거 수익률의 기간수, 포트폴리오 구성시점에서 과거 3, 6, 9, 12개월 등으로 설정
- **보유기간(K)** : 포트폴리오 구성 이후 포트폴리오를 보유하는 기간수, 3, 6, 9, 12, 24, 36개월 등 → 수익률이 보유기간에 따른 차이를 고려
- 상위 30%, winner portfolio, 하위 30%를 loser portfolio)라고 구분
- t 기간 말 시점의 주가모멘텀(PM) 측정

$$PM = \sum_{j=1}^J R_{t-j+1}$$

J = 3, 6, 9, 12개월  
Rt: t 시점의 월별 주식수익률

포트폴리오의 성과는 포트폴리오의 구성 시점인 t시점 이후 특정 보유기간 K의 월평균수익률인  $R_{Pi}(K)$ 의 평균값인  $R_P(K)$ 을 이용하여 다음과 같이 측정  
N : 매 시점마다 반복하여 실시되는 포트폴리오의 구성 횟수

$$R_P(K) = \frac{\sum_{i=1}^N R_{Pi}(K)}{N}$$

(K = 3, 6, 9, 12, 24, 36개월)

$$\text{여기서 } R_{Pi}(K) = \frac{\sum_{k=1}^K R_{Pt+k}}{K}$$



# \* Momentum effect : Momentum Anomaly

모멘텀 효과를 이용한 투자전략의 성과에 관한 연구 (2011)

Performance of Investment Strategies by Using Momentum Effect in Korea Stock Market

투자전략의 성과는 승자 및 패자 포트폴리오 수익률의 차이의 평균값인  $ARP(K)$ 로 측정

예를 들어, 패자포트폴리오를 매입함과 동시에 승자포트폴리오를 매도하는

역행투자전략의 성과는 다음과 같이 측정

$$ARP_p(K) = \frac{\sum_{i=1}^N [R_{P_i}^L(K) - R_{P_i}^W(K)]}{N} \quad (3)$$

여기서  $R_{P_i}^L(K)$ : 보유기간 K의 패자포트폴리오 월평균 수익률

$R_{P_i}^W(K)$ : 보유기간 K의 승자포트폴리오 월평균 수익률

포트폴리오 및 투자전략의 성과에 대한 통계적 유의성 평가는 t 통계량을 이용한 검정

〈표 1〉 승자 및 패자 포트폴리오의 성과

K	J	3		6	
		W	L	W	L
3	return	0.0063	0.0145	0.0055	0.0139
	p-value	0.7332	0.4179	0.7471	0.4413
6	return	0.0056	0.0126	0.0056	0.0100
	p-value	0.5813	0.1682	0.5867	0.2688
9	return	0.0045	0.0117	0.0007	0.0108
	p-value	0.6135	0.1841	0.9392	0.1899
12	return	0.0127	0.0212	0.0091	0.0204
	p-value	0.0492	0.0067	0.1506	0.0098
24	return	0.0140	0.0201	0.0133	0.0190
	p-value	0.0004	0.0000	0.0002	0.0001
36	return	0.0143	0.0191	0.0131	0.0188
	p-value	0.0003	0.0000	0.0004	0.0000
K	J	9		12	
		W	L	W	L
3	return	0.0049	0.0131	0.0033	0.0111
	p-value	0.7790	0.4618	0.8487	0.5527
6	return	0.0056	0.0094	0.0026	0.0101
	p-value	0.5729	0.3204	0.7899	0.2986
9	return	0.0007	0.0095	-0.0007	0.0103
	p-value	0.9313	0.2561	0.9337	0.2125
12	return	0.0118	0.0165	0.0108	0.0169
	p-value	0.0720	0.0143	0.1387	0.0083
24	return	0.0148	0.0171	0.0144	0.0171
	p-value	0.0001	0.0000	0.0001	0.0002
36	return	0.0148	0.0174	0.0150	0.0168
	p-value	0.0003	0.0000	0.0004	0.0000

포트폴리오 구성기간과 보유기간에 상관없이 패자포트폴리오의 성과가 승자포트폴리오보다 일관되게 높은 것으로 나타났다. 이러한 결과는 패자포트폴리오를 매입하고 승자포트폴리오를 매도하는 **역행투자전략이 유효함을 시사**

## \* Momentum effect : Momentum Anomaly

모멘텀 효과를 이용한 투자전략의 성과에 관한 연구 (2011)

Performance of Investment Strategies by Using Momentum Effect in Korea Stock Market

〈표 2〉 역행투자전략의 성과

K \ J	J	3	6	9	12
		L-W	L-W	L-W	L-W
3	return	0.0082	0.0084	0.0083	0.0078
	p-value	0.2375	0.3221	0.3763	0.4321
6	return	0.0070	0.0044	0.0038	0.0075
	p-value	0.0406	0.2750	0.4271	0.1741
9	return	0.0072	0.0101	0.0088	0.0110
	p-value	0.0204	0.0047	0.0377	0.0197
12	return	0.0086	0.0114	0.0046	0.0061
	p-value	0.0254	0.0240	0.1893	0.2138
24	return	0.0061	0.0057	0.0024	0.0028
	p-value	0.0014	0.0374	0.2314	0.2739
36	return	0.0048	0.0057	0.0026	0.0018
	p-value	0.0082	0.0060	0.0537	0.3395

→ size로 통제된 포트폴리오 및 역행투자 전략의 성과  
 → B/M으로 통제된 포트폴리오 및 역행투자 전략의 성과

# 퀀트(Quant) 투자

## 일중 모멘텀(intraday momentum) 전략

- Market Intraday Momentum  
\* Lei Gao Iowa State University and Yufeng Han University of North Carolina at Charlotte and Sophia Zhengzi Li Rutgers University and Guofu Zhou Washington University in St. Louis† First Draft: March, 2014 Current Version: June, 2017
- Intraday Momentum: The First Half-Hour Return Predicts the Last Half-Hour Return\* Lei Gao Iowa State University

Intraday Momentum In the Korean stock market 서울대 학위논문 2020

한국 주식시장에서의 일중 모멘텀

한국거래소 시장에 상장되어 있는 ETF 중 대표성을 가지는 KODEX 200을 대상

- 전일 종가부터 거래일 첫 30분까지의 수익률로 거래일 마지막 30분 수익률에 대한 예측력을 검증 → 경기 수축기일수록, 시장의 유동성이 부족할수록 일중 모멘텀이 두드러진 것을 확인할 수 있었다.
- Out-of-sample forecasting 방법론을 추가로 활용 → 첫 30분간 변동성과 거래량, 기관의 순매수 비율이 높을수록 일중 모멘텀이 뚜렷하게 나타남을 확인

# 퀀트(Quant) 투자

## Hedging Demand and Market Intraday Momentum

Guido Baltussen\*, Z Da, S Lammers, MPE Martens

\*Corresponding author for this work 26th August 2020

equities, bonds, commodities, and currencies between 1974 and 2020

"market intraday momentum" everywhere. The return during the last 30 minutes before the market close is positively predicted by the return during the rest of the day (from previous market close to the last 30 minutes)

### In-Sample Forecast

: 학습으로 사용한 데이터부터 잘 예측 여부

→ Train set을 얼마나 잘 학습했는가를 판단하는 것

### Out-Of-Sample Forecast

: 학습 데이터 내부에 있었던 기간 외에 미래를 예측하는 것

→ 모델 평가: 절대오차평균(MAE)

1. 선물 10분봉 또는 30분봉 데이터 수집

2. 설명변수(x) : 전일 증가~당일 증가 30분전 까지 수익률

종속변수(y) : 당일 증가 30분간 수익률 계산

→ 일중 수익률 계산

3. 선형회귀분석

독립변수와 종속변수간 유의성 확인

4. ETF의 헤지수요와 일중 모멘텀 간의 관계 확인

전일증가~당일 증가 30분전 까지 수익률이 양수(+)인 경우 3시15분 매수 후 → 3시 45분 증가 청산하는 전략 : 변동성이 큰 경우 통계적 유의성

# 퀀트(Quant) 투자

## Back Testing

과거데이터를 이용해서 본인이 만든 특정 투자전략에 대한 성과테스트 해보는 것

- CAGR(Compound annual Growth Rate, 연복리수익률)
- MDD(Maximum Draw Down) : (최고가-최저가)/(최고가) → 최대손실가능 수익률
- Sharpe ratio(SR) : 해당(펀드수익률-벤치마크수익률)/(해당펀드수익률의 표준편차)

(추가 검토)

- 히트 비율 : 양(+)의 수익 결과를 낸 베팅의 비중
- 히트로부터의 평균 수익률 : 수익을 낸 베팅의 평균 수익률 벤치마크대비?
- 미스로부터 평균 수익률 : 음(-) 즉, 손실을 낸 베팅의 평균 수익률

(샤프지수 재검토)

Sharpe ratio : 정규분포 가정 (2시그마, 3시그마)

→ 변동성의 방향이 위 아래로 어느 정도 비슷하게 발생할 것이라는 가정에 기반 즉, 대칭적인(symmetric) 전략

→ 초과수익률은 보통이지만 변동성(분모값)도 낮은 전략 > 초과수익률이 높지만 변동성이 큰 전략

확률적 샤프지수(PSR: PROBABILISTIC SHARPE RATIO)

: 편향되고 양쪽 꼬리가 두터운 경우 부풀림 효과를 제거하기 위한 수정된 샤프지수  
줄어든 샤프지수

: 임계값이 시행의 다수성을 반영하여 조정된 PSR

$$\widehat{PSR}(SR^*) = Z \left[ \frac{(\widehat{SR} - SR^*)\sqrt{n-1}}{\sqrt{1 - \hat{\gamma}_3 \widehat{SR} + \frac{\hat{\gamma}_4 - 1}{4} \widehat{SR}^2}} \right]$$

THE SHARPE RATIO EFFICIENT FRONTIER (2012) Marcos M. López de Prado

- Sortino Ratio = (평균수익률 - 무위험수익률) / 손실 낸 경우의 표준편차(변동성)
- information ratio = (포트폴리오 수익률 - 벤치마크 수익률) / 초과수익률의 표준편차(추적오차)

# 퀀트(Quant) 투자

All that glitters is not gold.

<https://github.com/quantopian/zipline>

## Quantopian (2011~2020)

- 1) AlphaZipline : 이벤트 lens : 알파 주가 예측 팩터의 성과 분석을 위한 파이썬 라이브러리
- 2) 기반 백테스팅 시스템,
- 3) Pyfolio : 포트폴리오 성과와 리스크 분석을 위한 파이썬 라이브러리 와 같은 일련의 오픈소스 툴을 공개

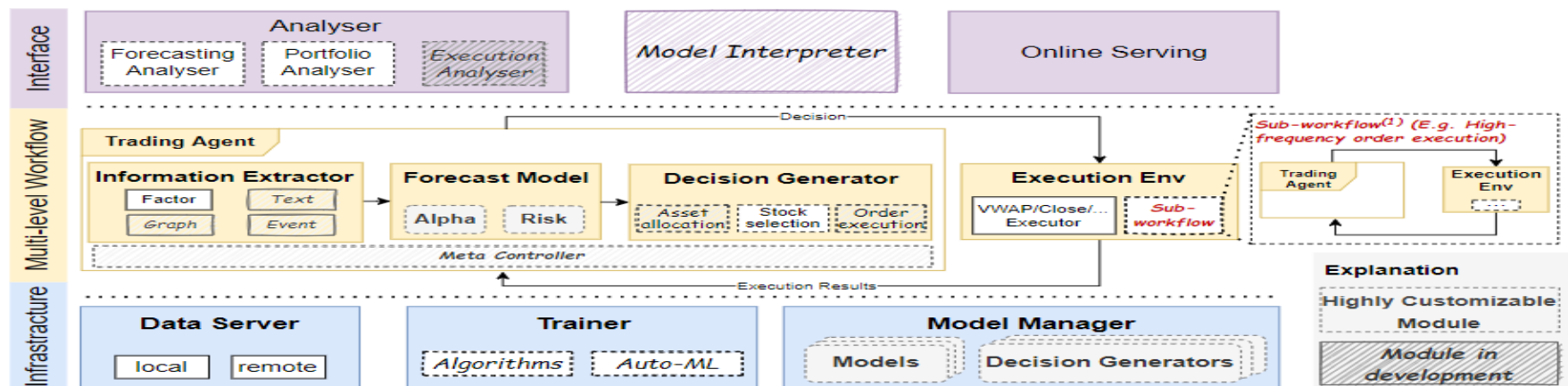
Zipline is a Pythonic algorithmic trading library. It is an event-driven system for backtesting.  
Zipline is currently used in production as the backtesting and live-trading engine powering [Quantopian](#)

Microsoft 중국 연구소, 오픈소스 퀀트 플랫폼인 **Qlib**

Qlib is an AI-oriented quantitative investment platform, which aims to realize the potential, empower the research, and create the value of AI technologies in quantitative investment.

<https://github.com/microsoft/qlib#framework-of-qlib>

## Framework of Qlib



# 퀀트(Quant) 투자

## 성과지표 관련

- **Relative Return (%)**:

→ 운용포트폴리오의 Benchmark 대비 상대수익률. (초과수익률)

→ abnormal return, excess return

- **Tracking Error (%)**

→: Relative Return의 표준편차.  $\sigma(R - RB)$

- **Information Ratio (IR) =: Relative Return/Tracking Error.**

- → 해당 포트폴리오의 벤치마크 대비 초과수익률을 Tracking Error로 나눈 지표. 벤치마크 대비

- **Information Coefficient (IC):**

→ 포트폴리오의 예상수익률과 실제수익률의 상관계수. IC가 높을수록 투자능력이 뛰어나다는 의미

→ 스피어만 상관계수 활용

Hit ratio : 전체 기간 중에서 특정기간 기준 벤치마크 대비 성과가 우수한 기간 횟수의 비율

# 퀀트(Quant) 투자

## Back Testing

<http://pmorisette.github.io/ffn/>

### bt 패키지 백테스트 과정

1. 가격 데이터의 수집
2. 전략 사전 정의
3. 데이터를 이용한 **전략**의 백테스트
4. 결과에 대한 평가 -- 예측

### 팩터모델 백테스팅의 방법론

- 종목군 :매월 해당시점 기준의 시가총액 상위 150개 종목을 투자 유니버스로 선정→
- 포트폴리오 구성: 각 Factor별로 수치가 존재하는 종목군을 대상으로 하여 최상위 20%의 종목군은 1분 위 (Long) 포트폴리오에 배정→
- 시뮬레이션 기간 설정→
- 포트폴리오 수익률 소속된 종목들에 비중입력 투자하여 포트폴리오 수익률을 계산→
- 포트폴리오 리밸런싱 투자 Horizon을 일정 개월단위 별로 , 6개 리밸런싱 방식으로 동시 테스트
- 상승구간/하락구간 구분



# 퀀트(Quant) 투자

## Back Testing : Measuring backtest performance

### 1. Backtesting.py

Backtest trading strategies in Python

### 2. bt - Flexible Backtesting for Python

3. backtrader

4. Zipline

Pyfolio

5. PyAlgoTrade

Zipline 사례 추가 코드 공유

<https://blog.est.ai/2021/05/bayesian-optimization-in-trading/>

<https://pmorisette.github.io/bt/>

<https://www.backtrader.com/>

<https://github.com/quantopian/zipline>

<https://gbeded.github.io/pyalgotrade/>

# Back Testing : backtrader 라이브러리 사용 사례

## zipline 라이브러리 (3.5)

<https://www.backtrader.com/docu/>

### 전략 만들기

- 잠재적인 조정 가능한 매개변수 결정
- 전략에 필요한 지표 인스턴스화
- 시장 진입/퇴출 논리를 작성

### Cerebro 엔진 만들기

- 첫째: 전략 (또는 신호 기반 전략) 주입
- 데이터 피드 로드 및 주입 (한 번 생성되면 사용 cerebro.adddata)
- 실행 : cerebro.run()
- 시각적 피드백: cerebro.plot()

```
from datetime import datetime
import backtrader as bt
```

```
class SmaCross(bt.Strategy) # bt.Strategy를 상속한 class로 생성해야 함.
    params = dict( pfast=5, # period for the fast moving average
                    pslow=30 # period for the slow moving average)
    def __init__(self):
        sma1 = bt.ind.SMA(period=self.p.pfast) # fast moving average
        sma2 = bt.ind.SMA(period=self.p.pslow) # slow moving average
        self.crossover = bt.ind.CrossOver(sma1, sma2) # crossover signal
    def next(self):
        if not self.position: # not in the market
            if self.crossover > 0: # if fast crosses slow to the upside
                close = self.data.close[0] # 종가 값
                size = int(self.broker.getcash() / close) # 최대 구매 가능 개수
                self.buy(size=size) # 매수 size = 구매 개수 설정
            elif self.crossover < 0: # in the market & cross to the downside
                self.close() # 매도
        cerebro = bt.Cerebro() # create a "Cerebro" engine instance
        # 삼성전자의 '005930.KS' 코드를 적용하여 데이터 획득
        data = bt.feeds.YahooFinanceData(dataname='005930.KS',
                                           fromdate=datetime(2019, 1, 1),
                                           todate=datetime(2019, 12, 31))
        cerebro.adddata(data)
        cerebro.broker.setcash(1000000) # 초기 자본 설정
        cerebro.broker.setcommission(commission=0.00015) # 매매 수수료는 0.015% 설정
        cerebro.addstrategy(SmaCross) # 자신만의 매매 전략 추가
        cerebro.run() # backtesting 실행
        cerebro.plot() # 그래프
```

<https://wendys.tistory.com/181>

# Back Testing : backtrader 라이브러리 사용 결과

## Calmar Ratio

- Start
- End
- Duration
- Exposure Time [%]
- Equity Final
- Equity Peak
- Return [%]
- Buy & Hold Return [%]
- Return (Ann.) [%]
- Volatility (Ann.) [%]
- Sharpe Ratio
- Sortino Ratio
- Calmar Ratio
- Max. Drawdown [%]
- Avg. Drawdown [%]
- Max. Drawdown Duration
- Avg. Drawdown Duration
- # Trades
- strategy SmaCross

## Calmar Ratio (Drawdown 비율)

$$\text{Calmar Ratio} = \frac{\text{Annualized}(R_P)}{\text{MaxDD}_P}$$

**Annualized( $R_P$ )** : Annualized portfolio return

**MaxDD<sub>P</sub>** : Portfolio maximum drawdown

- ➔ 안정적인 절대 수익률을 추구하는 헤지펀드 성과평가 지표로 주로 활용
- ➔ Terry W. Young 1991년 무역 저널 Futures에서 처음 출판

# 퀀트투자 : Backtesting 통계량 검토

## 1. 대상(투자)기간 및 운용주기 설정

: 충분한 기간 확보, 월간 리밸런싱. 투자전략

## 2. 운용자산의 평균 규모 고려

## 3. 최대 포지션 운용 규모 고려

## 4. Long과 short position 보유

## 5. 비율

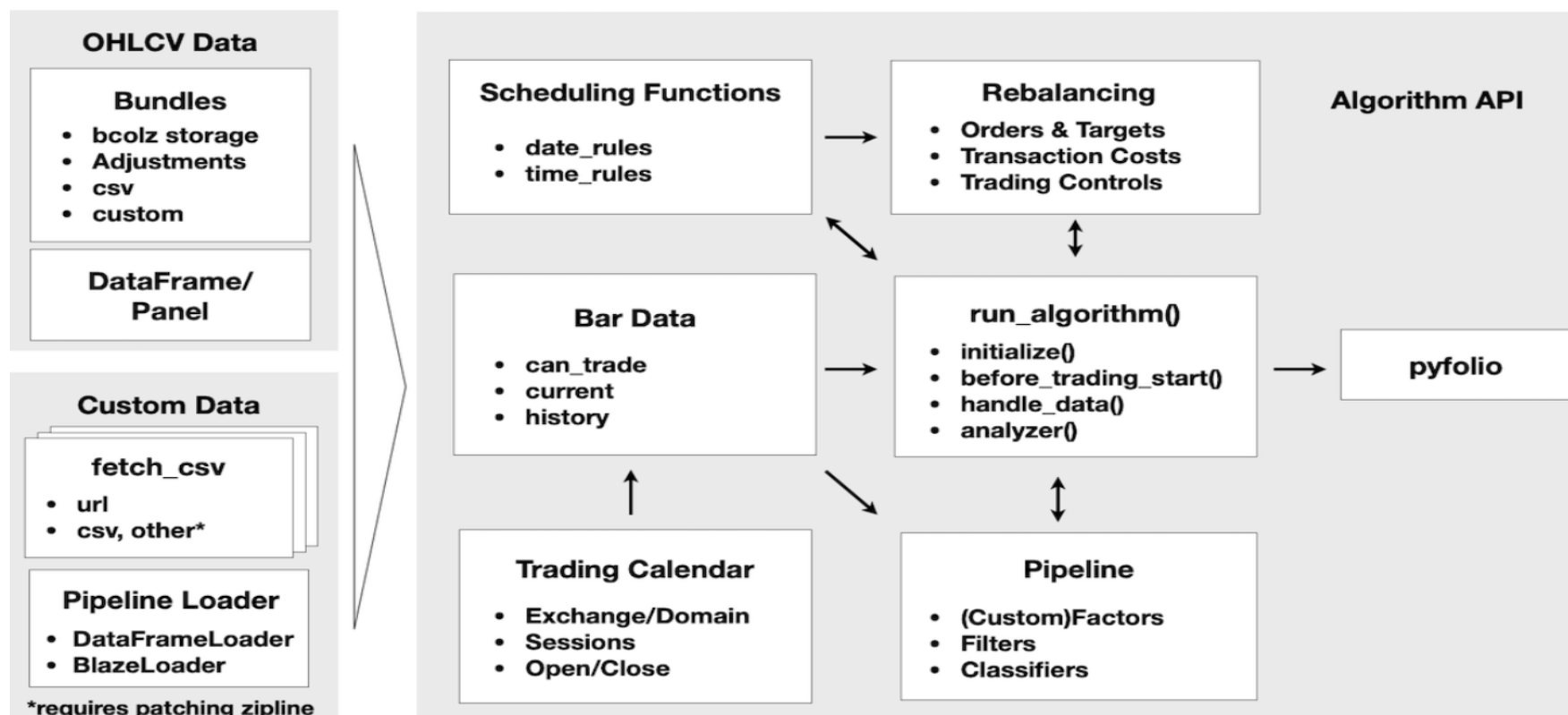
## 6. 평균 보유기간 : 1년 또는 3년, 5년

'상승' 시그널 : KOSPI 매수 유지

'보합' : 현금 보유 포지션 유지 (0% 수익률)

'하락' : KOSPI (공)매도 포지션 유지

### Zipline Architecture



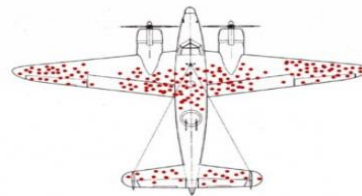
**Machine Learning for Algorithmic Trading:** Predictive models to extract signals from market and alternative data for systematic trading strategies with Python

# 퀀트투자 유의할 점

- 데이터 품질 : 투명성, 객관적 검증
  - 정확성: 백테스팅에 사용되는 데이터는 정확해야 한다. 잘못된 데이터는 잘못된 결과를 초래할 수 있다.
  - 완전성: 누락된 데이터가 없는지 확인해야 한다. 특히 시간의 흐름에 따라 일관성이 있어야 한다.
- Overfitting 문제점 : 금융 머신러닝 한계점
- 백테스팅 :→ 미래를 담보하지는 않는다.!!!
- Benchmark 설정(대표적 시장지수)
- 전략 분산 필요 (알고리즘 다양)
- 슬리피지(Slippage) 및 거래 비용
  - 문제: 실제 거래에서는 매수/매도 시점의 가격과 차이가 발생할 수 있으며, 거래 비용이 발생합니다.
  - 해결책: 백테스팅 시 슬리피지와 거래 비용을 고려하여 조정
- 편향(bias) 문제 발생
  - 1) 생존자 편향'(Survivorship bias)의 오류,
  - 2) Look ahead-bias
- Out of Sample Testing : Train/Test/Validation 데이터 셋 구분 , Walk-forward Testing , K-fold Cross Validation

- 생존편의(Survivorship bias) : 현재 존재하는 종목들만 사용하면, 과거에 상장폐지된 종목을 무시하게 되어 발생하는 편향을 말한다.
- 선견편의(Look ahead bias) : 아직 공시되지 않은 미래의 정보를 사용하여 백테스팅을 수행하는 경우 발생하는 편향을 말한다.

생존자 편향'(Survivorship bias) 사례  
2차 대전 중, 미 해군분석센터 연구원  
임무를 마치고 돌아온 전투기들의 총탄 자국에 대해 연구



## \* 백테스팅 관련 : bias

### 1. 생존편의(Survivorship bias) 제거

분석기간 동안 연속적으로 상장된 기업들뿐만 아니라 중간에 상장폐지되거나 신규 상장된 기업들을 모두 표본으로 선정

### 2. 선견편의 (예견 편향 : Look ahead bias) 제거

- 그 때 당시에는 알 수 없었던 정보나 데이터를 가지고 백테스팅 함으로써 발행하게 되는 오류
- 기업 재무제표 공시 관련 : 회계기간 기준월(12월)과 실제 공시 시점과는 3개월내 소요
- 기업 재무제표 수정하게 되는 경우 : 과거 데이터가 최종 수정된 버전의 데이터로만 저장되어 있다. 이런 경우 원래 데이터를 고려하지 않은 백테스팅이 발생하기에 재차 예견 편향이 발생
- 결국 백테스팅이 수정 전 데이터를 사용한 것인지? 수정 후 데이터를 사용한 것인지? 오류

#### (해결 방법 사례)

1. PIT(Point in time) 데이터 베이스를 사용  
: 백테스팅 날짜에 맞게 원래 보고된 데이터와 수정된 데이터를 모두 담고 있는 데이터
2. 재무제표자료는 3월 말부터 이용할 수 있는 것으로 가정 :  
재무변수 기준으로 정렬한 기업의 수익률은  $t$ 년 4월부터  $t+1$ 년 3월까지의 주기로 계산

## \* 백테스팅 관련 : bias

### 선견편의 (예견 편향 : Look ahead bias) : 12월 결산기업 기준 해결 사례

#### 1. 데이터 시점의 적절한 설정

(문제) 12월 결산기업의 경우, 연말 재무제표 3월 말에 공시(공시 의무기일)

(해결책) ① 데이터 타임스탬프 정확히 설정하기

각 데이터 항목이 실제로 이용 가능한 시점을 기준으로 타임스탬프를 설정한다.

(예시) 2022년 재무제표는 2023년 3월 31일 이후에나 이용할 수 있으므로, 이 데이터를 2023년 3월 31일 이후의 시점에서만 사용할 수 있도록 한다.

② 지연 데이터 사용: 백테스트에 사용되는 모든 데이터는 실제로 공시된 이후 시점부터 사용할 수 있게 설정한다.

(예시) 2022년 재무 데이터는 2023년 3월 이후의 투자 결정에만 반영되도록 한다.

#### 2. 테스트 기간 설정 및 데이터 접근

(문제) 데이터 접근 시점의 부정확성은 LAB (Look ahead bias)를 초래한다. 데이터가 실제로 언제 공개되었는지를 고려하지 않고 백테스트에 사용하면, 미래 정보를 미리 아는 것처럼 잘못된 결과가 나올 수 있다.

(해결책)

① 롤링 윈도우 방식: 데이터의 이용 가능 시점을 정확히 반영하여, 예를 들어 2015년부터 2020년까지의 데이터를 이용해 2021년을 예측하는 방식이다. → 이렇게 하면 각 연도별로 데이터를 실제 공개 시점 이후로만 사용할 수 있다.

② 실제 공시 일정 반영: 각 결산자료의 실제 공시 일정을 반영하여 데이터 접근을 설정한다. 예를 들어, 2022년 12월 결산 자료는 2023년 3월 31일 이후 접근 가능한 것으로 설정한다.

#### 3. 백테스트 구현 예시

예제 상황: 기업 A의 2022년 재무제표가 2023년 3월 31일에 공시되었다. 백테스트를 통해 2023년 1월부터 12월까지의 투자 전략을 검증하려고 한다.

1. 데이터 수집: 2022년 재무제표 데이터를 수집하고, 2023년 3월 31일을 타임스탬프로 설정합니다.

2. 데이터 적용 시점 설정: 2023년 3월 31일 이후 시점에서만 2022년 재무제표 데이터를 사용할 수 있도록 코드를 작성한다.

투자 전략 백테스트:

(사례1)

2022년 결산 자료를 기준으로 2023년4월1일~2024년 3월말까지로 투자하는 것으로 검증 실행한다.

(사례2) 2023년 1월부터 3월까지의 2021년 데이터를 기반으로 전략을 검증 실행한다.

2023년 4월부터 12월까지의 2022년 데이터를 기반으로 전략을 검증 실행한다.

#### 4. 검증 및 재검증

(문제) LAB를 완전히 제거하기 위해서는 철저한 검증이 필요하다.

(해결책) -모델 검증: 백테스트 결과를 철저히 검토하여 LAB가 없는지 확인한다.

-크로스 검증: 다양한 시점과 조건에서 모델을 테스트하여 일관된 성과를 확인한다.

-독립적 검토: 제3자의 검토를 통해 편향 여부를 추가로 확인한다.

LAB를 해결하기 위해서는 데이터의 공시 시점을 정확히 반영하고, 백테스트 시 이를 엄격하게 준수하는 것이 중요하다.

이를 통해 보다 신뢰성 있는 투자 전략을 개발할 수 있기 때문이다.



## \* 코스콤 테스트 베드

<https://www.ratestbed.kr:7443/portal/main/main.do>



로보어드바이저에 퇴직연금 맡긴다"...187개 전략 출격 준비 입력2024.02.24.

[https://n.news.naver.com/article/003/0012390443?cds=news\\_my](https://n.news.naver.com/article/003/0012390443?cds=news_my)

# Machine Learning 활용 Quant : 단기예측

- Linear Regression
- Logistic Regression
- Penalized Regression : LASSO, Ridge, Elastic Net regression
  
- Decision Tree : 20
- Ensemble
- Bagging : Random forest
- Boosting : Ada , GBM...

Python/sklearn

overfitting !!!

모델 학습 (설명) : training data

모델의 검증과 튜닝을 위해 사용되는 데이터 : validation data

→ (Cross-Validation) : training , validation data 분리 → K-fold

성능평가(예측): out-of-sample → test data

# Machine Learning 활용 Quant

## 단기예측 AI 모델 : R.F 알고리즘

- 기간(샘플수): 과거 일정시점 영업일별 ~ 현재 시계열 자료  
→ Up 데이터, down 데이터, neutral 데이터
- 피처(feature):  
Per 10배미만? → vkospi 20? →
  - 매크로 : 환율 1200 미만? → 장단기금리차이 양(+)  
여부? →
  - 밸류에이션
  - 어닝스 등의 투자지표 일별 수치
- 레이블(label): 해당일의 사후 KOSPI 방향성 예측  
→ up ( 4%이상), neutral (4%~ 0%) down ( -% 미만)  
→ y (종속변수) 라벨링을 통해 지도학습 모델링  
→ KOSPI의 사후 60영업일 수익률이 4% 이상이면 '상승', 4%~0% 사이면 '보합', 0% 미만이면 '하락' 결과로 기록.  
(T일의 KOSPI 방향성은 'T+1일~T+61일 간의 사후 수익률'을 기준으로 삼음. 현실적인 투자 가능성을 위해 하루 뒤부터 투자하는 것으로 가정)

데이터셋의 '불순도' (impurity) 를 가장 크게 낮추는 방향으로 하나의 특성(변수)과 그의 임계값을 순차적으로 찾아가는 방식

$$G_i = 1 - \sum_{k=1}^n p_{i,k}^2$$

### 피처 데이터 리스트

리스트	리스트
Citi G10 ESI	한국 P/E의 선진국 대비 할인율
Citi EM ESI	KOSPI Trailing P/B
원/달러 환율	FY2 이익조정비율 (1m)
한국 수출 증가율	FQ1 이익조정비율 (1m)
WTI 유가	FQ1 이익조정비율 (1w)
US Core PCE	투자의견 점수 변화율 (1w)
KOSPI Forward P/E	VKOSPI
Yield Gap	VIX
한국 P/E의 이머징 대비 할인율	Put-call volume ratio (20d)

참고: 기존 단기예측 모델의 데이터를 동일하게 사용. 다만, 일별 데이터 사용  
자료: 삼성증권

# Machine Learning 활용 Quant

## 모델 학습 프로세스

### 1. 기본 라이브러리 호출

2. Raw data 가져오기 (train, test data 나누기), 인코딩
3. 최적 하이퍼파라미터를 찾기 위한 모델 튜닝
4. K-fold cross validation 실행으로 모델 성능 확인
5. 최종 모델 학습
6. test data로 최종 모델 성능 확인
7. 모델 저장

```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import sklearn.metrics as mt
from sklearn.tree import export_graphviz
from sklearn.tree import DecisionTreeClassifier
from sklearn.ensemble import RandomForestClassifier
from sklearn.model_selection import train_test_split
from sklearn.model_selection import
StratifiedShuffleSplit
from sklearn.model_selection import cross_val_score
from sklearn.model_selection import
RandomizedSearchCV
from sklearn.externals import joblib
from sklearn.metrics import confusion_matrix
```

# Machine Learning 활용 Quant : 단기예측

## 2. 데이터 가져오기, 인코딩

각 일자별로(샘플수가 n개면 n개 ROW )  
레이블 정보와 피쳐 정보가 각 COLUMN 에 표시  
→ 레이블 정보 인코딩

\* encoding

### 1) label encoding

→ 범주형 데이터를 숫자로 일대일 매핑해주는 인코딩 방식  
범주형 데이터를 숫자로 치환

사이킷런의 LabelEncoder로 구현

명목형 데이터를 인코딩하면 모델 성능이 떨어질 수 있다.

머신러닝 모델이 서로 가까운 숫자를 비슷한 데이터라고 판단하기 때문

X 변수 저장 : 피쳐 데이터  
Y 변수 : 레이블 데이터

train, test 나누기  
: StratifiedShuffleSplit 객체 사용

### 2) one-hot encoding

여러 값 중 하나(one)만 활성화(hot)하는 인코딩

<실행절차>

1. 인코딩하려는 피쳐의 고유값 개수를 구한다.
2. 피쳐의 고유값 개수만큼 열을 추가한다.
3. 각 고유값에 해당하는 열에 1을 표시, 나머지 열에는 0을 표시

# Machine Learning 활용 Quant : 단기예측

## 3. 모델 세부 튜닝: Hyperparameter tuning

1. **GridSearchCV** : 모든 조합에 대해 교차검증 후 가장 좋은 성능을 내는 하이퍼파라미터 조합을 찾음

### 2. **RandomizedSearchCV**

hyperparameter의 최소, 최대값을 정해두고 범위 내에서 무작위 값을 정해진 횟수만큼 반복적으로 추출하여 최적의 조합을 찾는 것을 의미

GridSearch 와 동일한 방식으로 사용하지만 모든 조합을 다 시도하지는 않고 랜덤하게 적절한 수의 조합으로 최적 파라미터 테스트를 해주는 객체

Bergstra, James, and Yoshua Bengio. "Random search for hyper-parameter optimization." Journal of machine learning research 13.2 (2012).

### **GridSearchCV**

#### 주요 매개변수

- > estimator : 모델 객체를 지정 변수
- > param\_grid : 하이퍼파라미터 목록을 dictionary로 전달
- > scoring : 평가 지표
- > cv : 교차검증 시 fold 개수
- > n\_jobs : 사용할 CPU 코어 개수 (1: 기본값, -1: 모든 코어 다 사용)

#### 메소드

- > fit(X, y) : 학습
- > predict(X) : 제일 좋은 성능을 낸 모델로 예측
- > predict\_proba(X) : 제일 좋은 성능을 낸 모델로 predict\_proba() 호출

#### 결과 조회 변수

- > cv\_results\_ : 파라미터 조합별 결과 조회
- > best\_params\_ : 가장 좋은 성능을 낸 parameter 조합 조회
- > best\_estimator\_ : 가장 좋은 성능을 낸 모델 반환

### **RandomizedSearchCV**

#### 주요 매개변수

- > estimator : 모델 객체 지정
  - > param\_distributions : 하이퍼파라미터 목록을 dictionary로 전달
  - > n\_iter : 파라미터 검색 횟수
- 이하 GridSearchCV 와 동일

# Machine Learning 활용 Quant : 단기예측

## 4. K-fold cross validation 실행으로 모델 성능 확인

- K-fold cross\_validation을 진행하여, 모델의 일반적인 예측력을 확인하는 단계
- cross\_val\_score 교차 검증 함수

## 5. 최종 모델 학습

## 6. test data로 최종 모델 성능 확인

- test data를 통한 out-of-sample 테스트 과정

confusion matrix 확인

confusion\_matrix 함수를 통해서 혼동행렬을 생성하는 코드

## Feature importance :

트리기반 모델에서 사용 → 중요도를 구분하는 데에 트리의 분리 밀접한 관련

- 특정 feature가 트리를 분리하는데 얼마나 기여를 했는지에 따라 중요도가 결정되는 것<sup>12</sup>



# Machine Learning 활용 Quant : 단기예측

## 예측 프로세스

모델 로드

최근 데이터로 예측하기

→ 전일자까지의 각 지표 데이터가 포함된 최근 데이터셋

→ 일자 지나면서 데이터가 쌓일수록 최근 데이터를 계속 포함 하는 식으로 훈련 데이터를 늘리는 방식이 유효, 재학습

현재(마지막) 데이터 표시

연중 일별 전망치의 확률 변화 : predict\_proba 메서드 활용

# Machine Learning 활용 Quant

## 텐서플로우를 이용한 주가 예측에서 가격-기반 입력 피쳐의 예측 성능 평가 Performance Evaluation of Price-based Input Features in Stock Price Prediction using Tensorflow (2017)

학습데이터: 총 5430개로 2005/1/3 ~ 2015/12/30 동안의 일별 주가 데이터로 구성

테스트데이터 : 2016/1/4 ~ 2016/12/30 동안의 총 482개의 일별 데이터

검증데이터: 2017/1/2 ~ 2017/7/6 동안의 총 242개의 데이터

(입력 피쳐)

1. 시가, 고가, 저가, 종가, 거래량 5개 → 47.51%의 적중률, 52.69%의 예측 성능
2. 종가이동평균인 MA피쳐 5개를 추가 + 총 10개의 피쳐 → 약 51.01%, 56.01% 예측 성능
3. 거래량 이동평균 VMA 5개 피쳐 추가 + 총 15가지의 피쳐 → 48.31%의 적중률, 56.84% 예측 성능

(적중률 : hit ratio)

5일 후 종가가 오를 경우 1, 내릴 경우를 0으로 설정

출력 결과가 1 과 1로 또는 0과 0으로 일치할 시 적중했다는 표현

## 퀀트(Quant) 투자 : 통계 vs 금융 머신러닝 연구 자료

<https://www.quantresearch.org/>



**Marcos López de Prado**

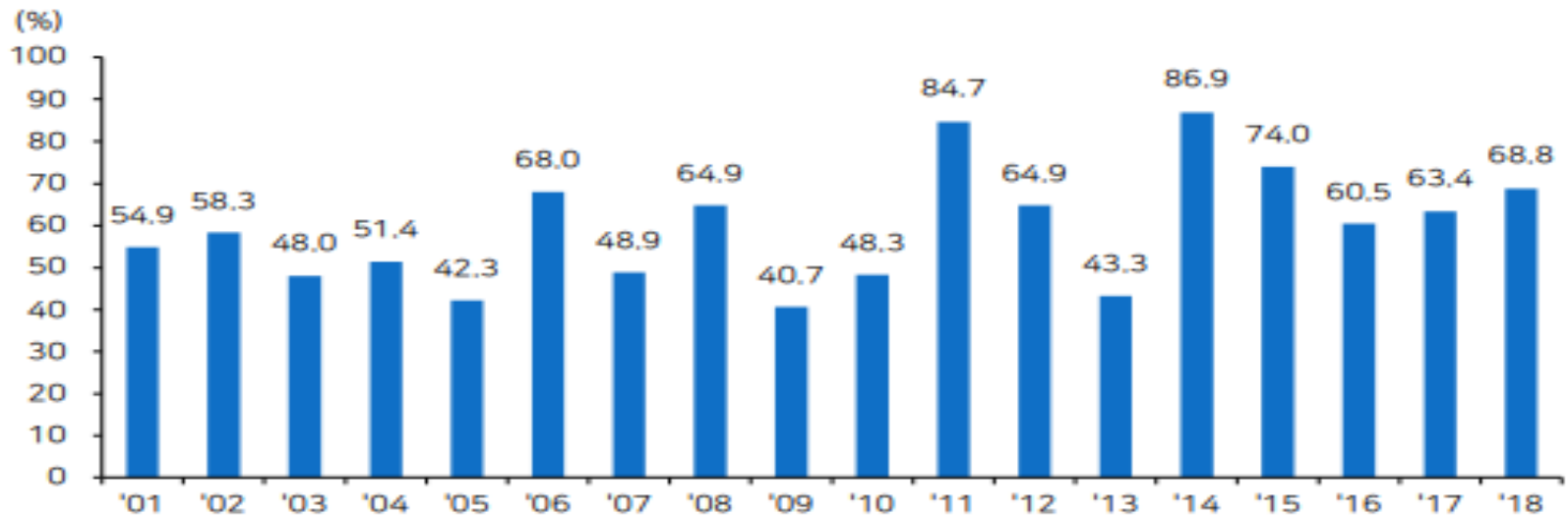
연구 도중 백테스팅을 하는 것은 음주운전과 같다.

# 퀀트투자 진화 : 퀀터멘탈(Quantamental)

퀀터멘탈(Quantamental) = Quant + Fundamental

계량분석(퀀트투자)과 펀더멘탈 분석(액티브 투자)을 모두 사용하는 투자방법론을 의미

그림1 미국내 주식형 펀드 중 벤치마크 지수 성과를 하회한 비율(Underperform 펀드 비중)



주: 벤치마크 지수로는 S&P500 지수를 사용

자료: S&P Dow Jones Indices LLC, 메리츠증권증권 리서치센터

미국 최대 자산운용사인 Blackrock

SAE(Systematic Active Equity)

Blackrock에서 퀀터멘탈 투자를 실행하는 본부로 기존의 액티브 매니 저들의 경험과 노하우를 알고리즘으로 시스템화하고, 과학적인 데이터 분석, 투자 시뮬레이션, 리스크 관리 등에 퀀트 기법을 융합하는 조직

# 퀀터멘탈(Quantamental)

퀀터멘탈(Quantamental) = Quant + Fundamental

## 표1 퀀터멘탈 투자 방법론이 주목받는 배경

기존 액티브 투자의 성과 부진

ETF등 패시브 펀드의 부상

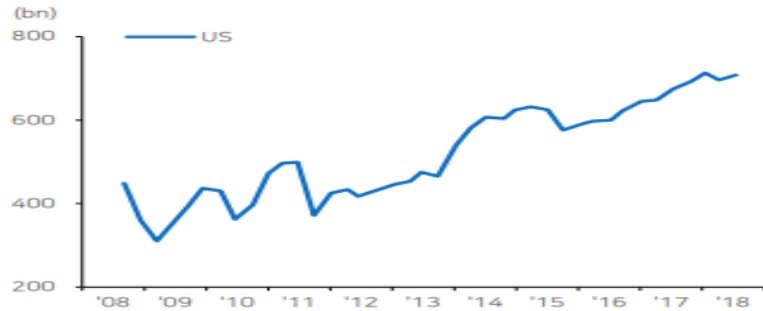
이용 가능한 데이터의 증가와 데이터 가격의 하락

AI, 머신러닝 분석 기술의 발전

비용 절감 욕구 : 사람이 하는 작업의 일정부분은 알고리즘으로 시스템화 하는 것이 비용상 유리

자료: 메리츠증권증권 리서치센터

그림4 퀀터멘탈 방법론을 쓰는 펀드 규모 추이\_미국



주: 액티브 롱온리 펀드 기준

자료: eVestment, 메리츠증권증권 리서치센터

그림5 퀀터멘탈 방법론을 쓰는 펀드 규모 추이\_미국 외



주: 액티브 롱온리 펀드 기준

자료: eVestment, 메리츠증권증권 리서치센터

# 퀀터멘탈(Quantamental)

## 퀀터멘탈 투자방법론

그림6 퀀터멘탈 방법론 설명 사례 : 퀀트 스크리닝 후 중소형주 애널리스트가 최종 종목 선별



자료: 한화투자증권, 메리츠증권 리서치센터

# 퀀터멘탈(Quantamental)

## 퀀터멘탈 최근

최근에는 투자 의사 결정 전반을 알고리즘으로 만들고 대체 데이터의 사용 및 분석, 리스크 관리 등에 적극적으로 퀀트 분석(머신러닝을 포함)을 활용하는 정도까지 퀀터멘탈 투자가 적용

표2 대체 데이터 사용 및 분석 사례1



자료: Twitter, Amenty Analytics, US BEA

표3 대체 데이터 사용 및 분석 사례2



자료: CEUR

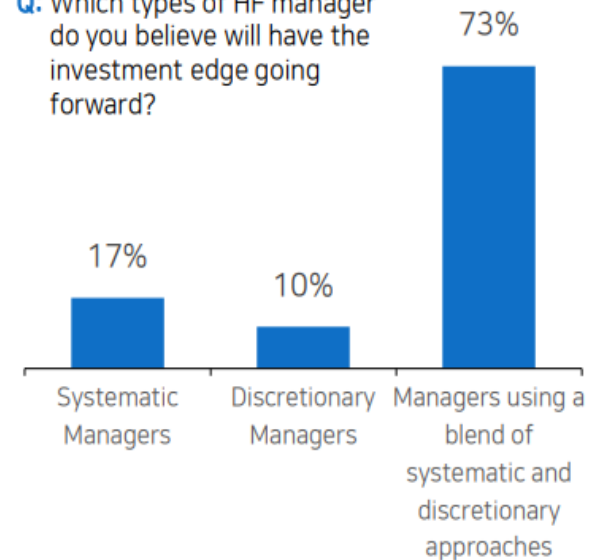
# 퀀터멘탈(Quantamental)

## 퀀트투자 방식(Systemic advantage)과 펀더멘탈 투자 방식(Human advantage)

Areas of Comparative Advantage		
Systematic Advantage		Human Advantage
Liquid	Liquidity of Underlying	Illiquid
Large	Sample Size	Small
Broad	Area of Focus	Narrow
Diversified	Portfolio Construction	Concentrated
Shorter	Investment Horizon	Longer
Generic	Investment Themes	Idiosyncratic / Event Specific
Generalist ELS / EMN, Macro	Key Strategies	Distressed / stressed credit, concentrated / focused ELS, activist, convert arb

### Expectations for Manager Edge Going Forward<sup>1</sup>

Q: Which types of HF manager do you believe will have the investment edge going forward?



Barclays 증권

**Discretionary investment management** is a form of investment management in which buy and sell decisions are made by a portfolio manager or investment counselor for the client's account.



# 퀀트(Quant) 투자 : 투자자 편향

## \*Behavioral Finance Bias

- Overconfidence Bias
- Hindsight Bias(사후 확신 편향)
- Confirmation Bias (확증 편향)
- Anchoring Bias
- Loss Aversion Tendency
- Herd Mentality Bias

과신하지 말라 !!

# 퀀트(Quant) 투자 : 사례

\*\* 인공지능(ML/DL) Trading

Python 파일 공유(별도)

- Backtesting : buy & hold 전략
- Bollinger Bands 사례
- Joel Greenblatt 마법공식

# Fi- machine learning

## True Positive Technologies 창업

Quant of the Year 2019" Marcos Lopez de Prado

*Q: Isn't it true that some **econ-quant firms** have started to experiment with **machine learning**?*

*A: According to the press, yes. But I suspect that many of these public announcements are primarily a **marketing ploy**, in an attempt to appease investors after years of mediocre results and pressures to reduce fees. There are two reasons for this reluctance to modernise. **First**, for years some quant-econ firms have publicly criticised machine learning, based on their false belief that it is a black box. It is hard for them to acknowledge their obsolescence. **Second**, young economists are willing and eager to apply modern statistical tools, **however the old guard within quant-econ firms has an agency problem. They perceive modernity as a threat to their authority and status. In this internal struggle, the leadership may undermine or backpedal the modernisation effort, in order to preserve the balance of power.***

## THE FUTURE OF FINANCE

<https://www.truepositive.com/>



# 퀀트(Quant) 투자 : 금융 머신 러닝 논제

## 금융머신러닝 논제

: 비구조화된 데이터--분석, 유용한 정보 추출 - 정규화된 형태로 저장

1. 금융데이터 구조 : 기초데이터,시장데이터,분석데이터,대체데이터

2. 레이블링

- 지도학습 알고리즘

-퀀터멘탈 투자전략

3. 모델링 문제점

4. 백테스팅 문제점

# Marcos Lopez de Prado

## 투자 알고리즘을 테스트

## Tactical investment algorithms1

the walk-forward 방법(과거 데이터로 시뮬레이션)

→ all-weather 가설

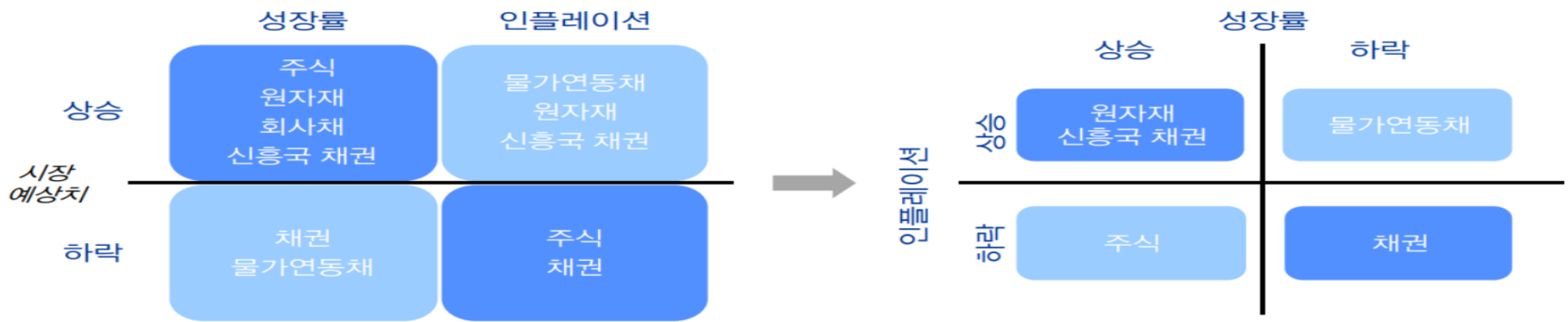
→ 특별한 시장 환경에서 최적인 투자 알고리즘(Tactical investment algorithms)

[금융이 과학이 될 수 없는 한계]

- 반증가능성(Popper's falsifiability)이 성립하지 않는다.
- 금융의 비정상성(non-stationarity) 때문이다. → 금융시장이 매우 동적이고 복잡하고 시간에 따라 빠르게 변하며, 인과관계도 변하며, 경기사이클, 시장 환경적인 변수들이 수시로 변한다.
- → 이러한 한계로 인해 투자 알고리즘 개발하고 back testing 검증

## 실전은 이론과 다르다

그림 5. All weather – 경제 상황을 4개 국면으로 구분 (4 Season)



자료: Bridgewater, IBK투자증권

# Marcos Lopez de Prado

New frontiers: Marcos Lopez de Prado on Machine Learning for finance

<https://www.youtube.com/watch?v=MP-jMMOgdjQ>

The 7 Reasons Most Machine Learning Funds Fail Marcos Lopez de Prado from QuantCon 2018

<https://www.youtube.com/watch?v=BRUISm4gdQ4>

Quantum Computing | Marcos López de Prado | Exponential Finance

믹스 - Quantopian

<https://www.youtube.com/watch?v=kU7vk9jmQC8>

# 유튜브 : 믹스 - Quantopian

Quant 워크 플로우

[https://www.youtube.com/watch?v=RomfTrm5\\_7g](https://www.youtube.com/watch?v=RomfTrm5_7g)

Intro and Getting Stock Price Data - Python Programming for Finance p.1

<https://www.youtube.com/watch?v=2BrpKpWwT2A&list=PLQVvvaa0QuDcOdF96TBtRtuQksErCEBYZ>

# mlFinLab Dr. Marcos Lopez de Prado

Machine Learning Financial Laboratory (MlFinLab)

개발한 회사 [Hudson and Thames Quantitative Research](#)

다양한 논문을 기반으로 하여 금융관련 ML을 집중 연구하는 그룹.

We source all of our implementations from the most elite and peer-reviewed journals.

Including publications from:

1.[The Journal of Financial Data Science](#)

2.[The Journal of Portfolio Management](#)

3.[The Journal of Algorithmic Finance](#)

4.[Cambridge University Press](#)

*Python 라이브러리 설치하셔서*

*ML 를 이용 금융 분석시 활용해보세요*

<https://www.mlfinlab.com/en/latest/index.html>



## Popular Modules

### **Financial Data Structures**

*Standard: Tick, Volume, Dollar bars. Information-Driven Bars: Imbalance and Run Bars (Tick, Volume, Dollar).*

### **Labelling Techniques**

*Triple-Barrier, Meta-Labeling, Trend Scanning, Tail-Sets, Matrix Flags, Excess Over Mean/Median, Return Vs. Benchmark.*

### **Feature Engineering**

*Fractionally Differentiated, Structural Breaks (CUSUM, Explosiveness Tests), Market-Microstructural.*

### **Machine Learning**

*Sampling, Sequentially Bootstrapped Ensembles, Feature Importance (MDI, MDA, Model Fingerprint), Cross-Validation(Purged, Embargo), Bet Sizing (EF3M).*

### **Portfolio Optimization**

*Mean-Variance, Black-Litterman, Hierarchical Risk Parity, Hierarchical Equal Risk Contribution, Nested Clustered Optimization.*

### **Risk Estimators**

*Min Cov Determinant, MLE Covariance Estimator, Shrinkage, De-noising and De-toning, Hierarchical Cluster Filtering, Theory Implied Correlation.*

### **Online Portfolio Selection**

*Benchmarks, Momentum, Mean Reversion (PAMR, OLMAR), Pattern Matching (CORN, SCORN, FCORN, FCORN-K), Universal Portfolios.*

### **Pairs Trading**

*Codependence, Co-integration (Engle-Granger, Johansen), Optimal Timing of Trades (Entry, Take Profit, Stop Loss), Simulations (OU, XOU).*

### **Synthetic Data Generation**

*Related to Correlation Matrices: CorrGAN, Vines (R, C, D, Partial Correlation), Extended Onion Method, Hierarchical Correlation Block Model.*

# Quant : 트레이딩을 위한 소셜 빅데이터 분석 모델 송성환1), 외 참고

소셜빅데이터로부터 추출된 감성을 전통적인 투자기법을 효율 적으로 적용하여 인공지능(기계학습)을 이용한 주식트레이딩을 구현하고 운용함으로써 사람이 배제된 인공지능과 빅데이터만으로 구성되는 주식트레이딩의 유효성을 입증

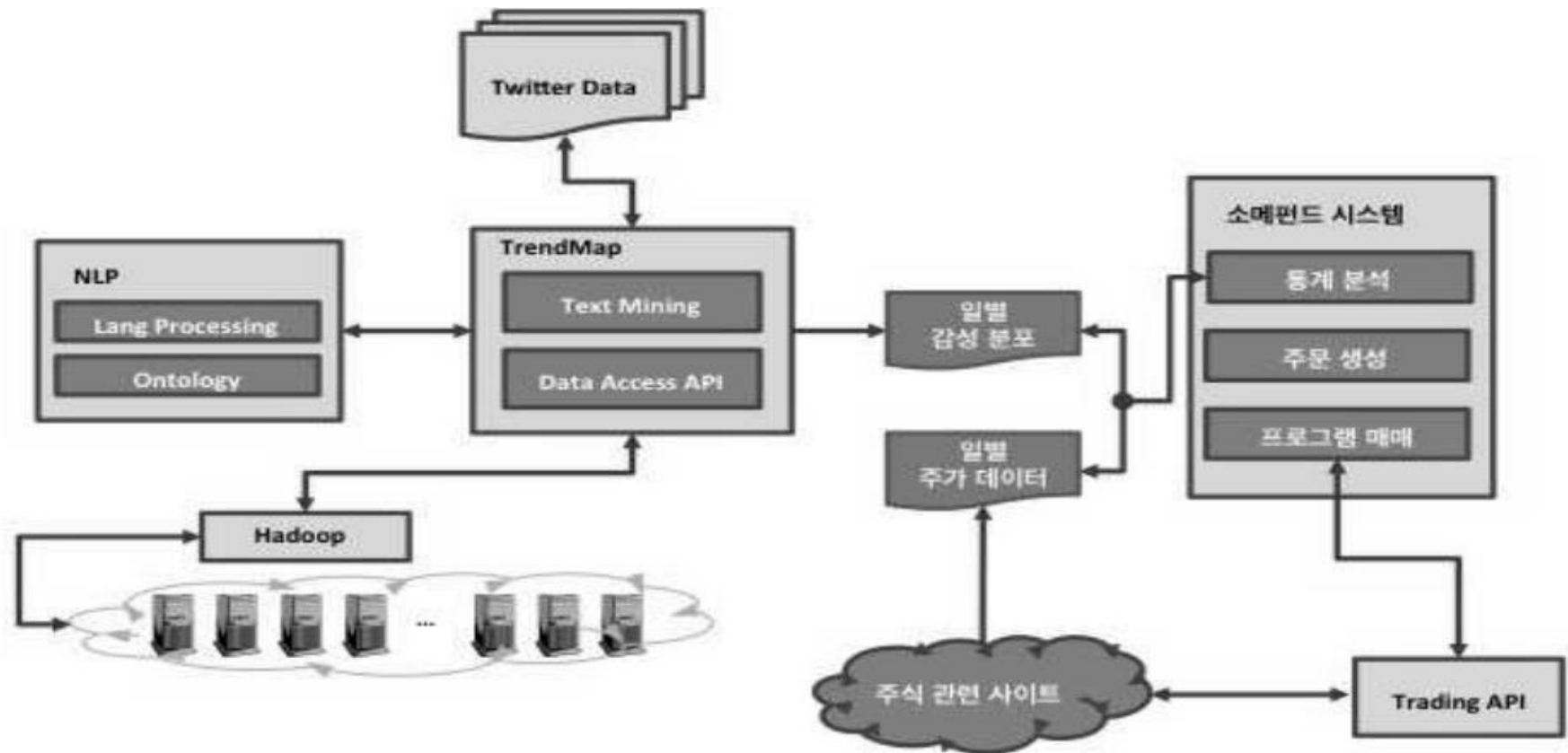
- Mean Reversion
- (Pairs Trading) (Statistical Arbitrage Strategy)
- Black-Litterman Mean
- 트위터 데이터를 수집하기 위하여 트위터사에서 제공하는 REST API와 Streaming API를 사용
- 일별 감성 데이터와 일별 주가 데이터로 상관관계 분석
- 최근 100일 동안의 감성 데이터와 주식 데이터에 대해 회귀분석 실시. 독립변수(x)는 감성카운트, 종속변수(y)는 매일의 해당 종목의 주가

트레이딩시스템 운용결과

벤치마크인 KOSPI200 대비 12%를 초과하는 성과

# Quant : 트레이딩을 위한 소셜 빅데이터 분석 모델 송성환1), 외 참고

## 트위터 데이터를 이용한 트레이딩 시스템 구성



트레이딩시스템 실제 운용결과

벤치마크인 KOSPI200 대비 12%를 초과하는 성과

## Quant : OO 사례-1

- 시가총액 큰 종목만 투자하는 방식은 효과적이지 않음. 반대로, 단순히 소형주에만 투자할 경우에도 수익률은 개선되지 않음
- 주가 Size와 투자수익률의 관계는 로직 상 아무런 연관이 없음. 결과도 마찬가지
- 투자의견 평균점수의 상향조정이 큰 종목에 투자. 중간 수준의 성과.
- 고ROE는 고PBR과 연결되며 고밸류에이션 상황일 가능성이 높음
- ROE, ROA 등 재무지표는 좋은 회사를 찾는 기준은 될 수 있으나 좋은 주식을 찾는 기준은 될 수 없음
- 재무지표(ROE,ROA,OP Margin) 중 가장 나쁜 성과 기록. 고마진에 의한 과도한 밸류에이션 적용이 문제가 됨
- 이익의 질이 높은 종목에 투자 (회계상의 순이익보다 현금흐름이 좋은 기업). 하락장에서 좋은 성과를 보이는 특징을 가짐
- 목표주가의 변화 자체만 보는 방식으로, 현주가의 고려가 없다는 것이 단점. 낮은 수익률
- 애널리스트의 최종 투자의견을 활용. 이익모멘텀 팩터보다는 성과가 떨어짐
- 현주가 대비 목표주가가 높은 종목 투자. 중간 수준의 성과

## Quant : OO 사례-2

- 경기민감주와 경기방어주의 차별적인 팩터반응도를 감안, 업종별로 factor weight를 다르게 적용. 단순한 "저P/E + 이익모멘텀" 조합 전략보다 향상된 수익률
- 밸류에이션, 이익모멘텀, 성장성과 수급 등 다양한 팩터를 활용하는 멀티팩터 전략. 우수한 성과
- 종목별 EPS를 연초부터 정확히 안다고 가정, 저PER에 꾸준히 투자하는 가상전략. 투자가능전략보다 훨씬 높은 성과 기록
- BPS를 연초부터 정확히 안다고 가정, 저PBR에 꾸준히 투자하는 가상전략. 가상PER 전략보다는 낮음
- 종목별 EPS를 연초부터 정확히 안다고 가정, 고성장주에 투자하는 가상전략. 투자가능전략보다 훨씬 높은 성과 기록
- 1달후의 EPS컨센서스를 미리 정확히 안다는 가정, 상향조정 상위주에 투자하는 가상전략. 가상PER 전략보다 더 높은 수익률
- 향후 1달간 외국인 비중이 가장 크게 높아질 종목을 미리 알고 투자하는 가상전략. 높은 수익률 보임
- 향후 1달간 기관이 가장 많이 살 종목을 미리 알고 투자하는 가상전략. 가상전략 중에서 가장 높은 수익률. 가상전략 중 수급전략이 가장 뛰어남

# Financial Feature Engineering: How to research Alpha Factors

## Index Fund 와 알파펀드

John Bogle, "The Index Mutual Fund: 40 Years of Growth, Change, and Challenge", Journal of Financial Analyst, Jan/Feb 2016).

### 'Smart beta'

투자 포트폴리오 구축시 risk factor에 대한 overweight 또는 underweight을 통해 시장 대비 초과수익을 추구하는 규칙 기반 투자전략

→ 특정 Factor에 노출된 Alpha 전략 일환 정의

(Josh Barrickman, "Why smart beta can't win the indexing race", Vanguard Blog for Advisor, Feb. 2015)

→ 2006년 Willis Towers Watson

→ 2019년 기준 : ETF.com에 따르면 스마트 베타 펀드에 약 8,800억 달러가 투자, 현재 시장에는 1000개 이상의 스마트 베타 ETF

# Factor 투자 전략

**팩터(Factor)** : 투자자산의 전체를 구성하는 성분 요소를 의미하며, 각 요소는 독립적인 특징

## Factor의 종류

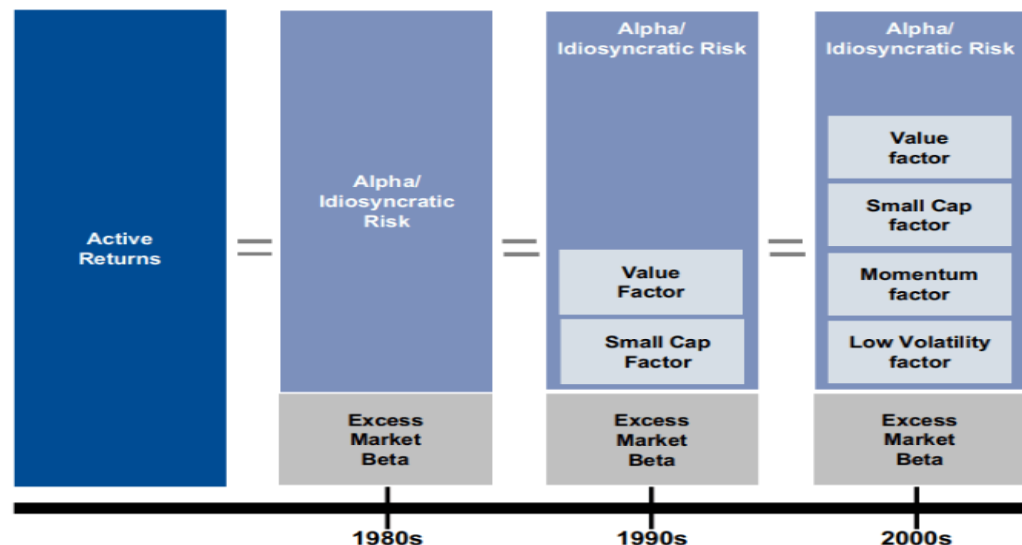
1. 매크로 Factor : 경기성장 및 인플레이션과 관련된 팩터 → 경기민감업종
2. 통계적 Factor : 통계적으로 주가와 유사하게 움직이는 팩터.
3. 내재 /스타일 Factor : 가치주, 성장주 & 모멘텀 등

Eugene Fama+ Kenneth French

: 밸류, size, 퀄리티 + 저변동성, 배당, 모멘텀 팩터 : 6개의 주요 팩터

"Size, Value, and Momentum in International Stock Returns," *Journal of Financial Economics* 105 (September 2012), with **Eugene F. Fama.**

## SmartBeta 전략



## \*\* 추가

<https://www.msci.com/documents/10199/71b6daf5-9e76-45ff-9f62-dc2fcd8f2721>

Well-Known Systematic Factors from the Academic Research

Systematic Factors	What It is	Commonly Captured by
Value	➤ Captures excess returns to stocks that have low prices relative to their fundamental value	➤ Book to price, earnings to price, book value, sales, earnings, cash earnings, net profit, dividends, cash flow
Low Size (Small Cap)	➤ Captures excess returns of smaller firms (by market capitalization) relative to their larger counterparts	➤ Market capitalization (full or free float)
Momentum	➤ Reflects excess returns to stocks with stronger past performance	➤ Relative returns (3-mth, 6-mth, 12-mth, sometimes with last 1 mth excluded), historical alpha
Low Volatility	➤ Captures excess returns to stocks with lower than average volatility, beta, and/or idiosyncratic risk	➤ Standard deviation (1-yr, 2-yrs, 3-yrs), Downside standard deviation, standard deviation of idiosyncratic returns, Beta
Dividend Yield	➤ Captures excess returns to stocks that have higher-than-average dividend yields	➤ Dividend yield
Quality	➤ Captures excess returns to stocks that are characterized by low debt, stable earnings growth, and other “quality” metrics	➤ ROE, earnings stability, dividend growth stability, strength of balance sheet, financial leverage, accounting policies, strength of management, accruals, cash flows



## \*\* 추가 : MSCI Foundations of Factor Investing

<https://www.msci.com/documents/10199/71b6daf5-9e76-45ff-9f62-dc2fcd8f2721>

Theories Behind the Excess Returns to Systematic Factors

Systematic Factors	Systematic Risk-based Theories <sup>17</sup>	Systematic Errors-based Theories <sup>18</sup>
Value	<ul style="list-style-type: none"> <li>➤ Higher systematic (business cycle) risk</li> </ul>	<ul style="list-style-type: none"> <li>➤ Errors-in-expectations</li> <li>➤ Loss aversion</li> <li>➤ Investment-flows-based theory</li> </ul>
Low Size (Small Cap)	<ul style="list-style-type: none"> <li>➤ Higher systematic (business cycle) risk</li> <li>➤ Proxy for other types of systematic risk</li> </ul>	<ul style="list-style-type: none"> <li>➤ Errors-in-expectations</li> </ul>
Momentum	<ul style="list-style-type: none"> <li>➤ Higher systematic (business cycle) risk</li> <li>➤ Higher systematic tail risk</li> </ul>	<ul style="list-style-type: none"> <li>➤ Underreaction and overreaction</li> <li>➤ Investment-flows-based theory</li> </ul>
Low Volatility	<ul style="list-style-type: none"> <li>➤ N/A</li> </ul>	<ul style="list-style-type: none"> <li>➤ Lottery effect</li> <li>➤ Overconfidence effect</li> <li>➤ Leverage aversion</li> </ul>
Dividend Yield	<ul style="list-style-type: none"> <li>➤ Higher systematic (business cycle) risk</li> </ul>	<ul style="list-style-type: none"> <li>➤ Errors-in-expectations</li> </ul>
Quality	<ul style="list-style-type: none"> <li>➤ N/A</li> </ul>	<ul style="list-style-type: none"> <li>➤ Errors-in-expectations<sup>19</sup></li> </ul>

## \*\* 추가 : MSCI Foundations of Factor Investing

사례 보기 : <http://www.kodex.com/factor.do?mobile>

KODEX는 방대한 양의 주식 및 재무 데이터를 기반으로 한 팩터 공식을 이용하여 종목을 선정하고, 선정된 종목을 바탕으로 다양한 팩터 기반 ETF를 운용

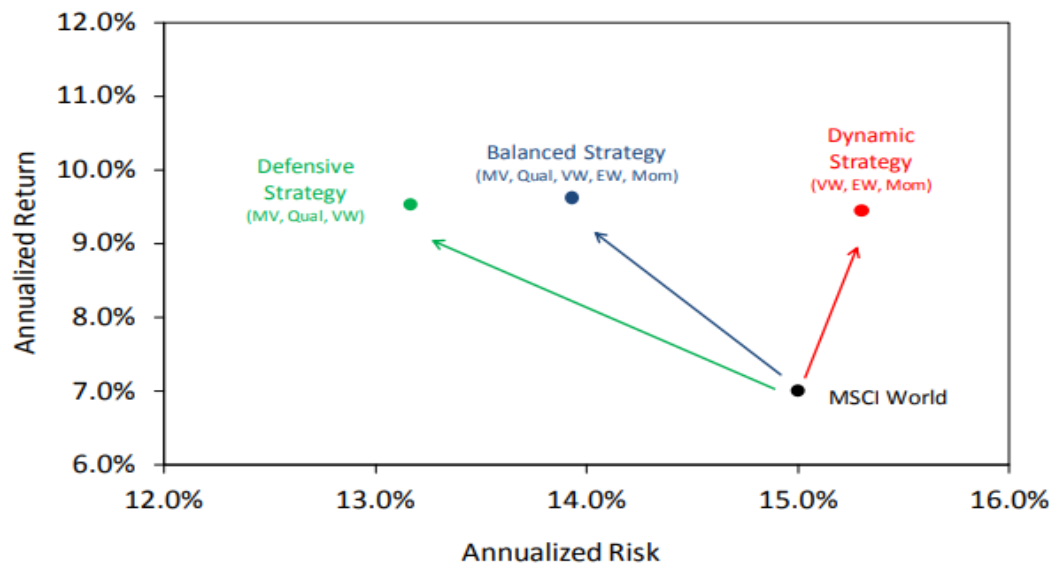


## \*\* 추가 : MSCI Foundations of Factor Investing

<https://www.msci.com/documents/10199/71b6daf5-9e76-45ff-9f62-dc2fcd8f2721>

Factor Investing versus Market Cap Investing

Factors Represent Strategic Bets Away from the Market Capitalization Weighted Index (Return and Risk, June 1988 to June 2013)



In sum, a market capitalization weighted index is the only appropriate candidate for a truly passive, macro consistent, buy and hold investment strategy that aims to capture the long term equity risk premium with structurally low turnover, very high trading liquidity and extremely large investment capacity.

In contrast, investing in factors represents active views away from the market portfolio and investors must form their own belief about what explains the premium and whether it is likely to persist. Thus, like traditional active strategies, factor index strategies should be assessed in the long run against a market capitalization weighted benchmark.

# 자산배분 : 선형 Factor Model

THE JOURNAL OF FINANCE \* VOL. XLVII, NO. 2 \* JUNE 1992

The Cross-Section of Expected Stock Returns

EUGENE F. FAMA and KENNETH R. FRENCH\*

Fama - French (FF)의 모형은 CAPM(Capital Asset Pricing Model)의 한계점을 극복하기 위해 제시된 이론

$$R_{it} - R_{ft} = \alpha_{it} + \beta_1(R_{Mt} - R_{ft}) + \beta_2SMB_t + \beta_3HML_t + \epsilon_{it}$$

where:

$R_{it}$  = total return of a stock or portfolio  $i$  at time  $t$

$R_{ft}$  = risk free rate of return at time  $t$

$R_{Mt}$  = total market portfolio return at time  $t$

$R_{it} - R_{ft}$  = expected excess return

$R_{Mt} - R_{ft}$  = excess return on the market portfolio (index)

$SMB_t$  = size premium (small minus big)

$HML_t$  = value premium (high minus low)

$\beta_{1,2,3}$  = factor coefficients

- Fama French 모형 → 팩터투자전략
- 다중회귀분석

1)  $R_m - R_f$  = Excess return on the market → 시장수익률 : KOSPI 지수 수익률

- 무위험이자율 : 정기예금금리, CD91일물 수익률

2) **SMB** : 소규모기업(시가총액 작은 기업) 평균수익률 -

대규모기업평균수익률 ( $SMB = 1/3(\text{Small Value} + \text{Small Neutral} + \text{Small Growth}) - 1/3(\text{Big Value} + \text{Big Neutral} + \text{Big Growth})$ )

- 기업규모; 시가총액) : 보통주 주가 × 보통주 발행주식수

3) **HML**: Book to market equity 가 높은 기업평균수익률 - 저가치주 평균수익률

▪  $HML = 1/2(\text{Small Value} + \text{Big Value}) - 1/2(\text{Small Growth} + \text{Big Growth})$

▪ - BM(장부가/시가) : 주식의 장부가치(Book value 자본총계-우선주)을 시가총액으로 나눈 값 =  $1/PBR$

Market Risk' 'Size Risk', 'Value Risk"

→ 시가총액이 작거나 (소형주), Value가 높은 기업 (장부가치/시가총액이 낮은 종목 →  $\frac{1}{PBR}$ ) 일수록 초과수익을 내기 유리하다는 것을 증명

# 자산배분 : 선형 팩터 모델

## 'Fama French Three Factor Model'(1992)

### SMB와 HML 포트폴리오를 구성 사례 : 2 x 3

- 1) 기업의 시가총액 기준으로 상위 50%와 하위 50%에 속하는 기업들로 2개의 포트폴리오(Big, Small)를 구성한다.
- 2) 장부가(BE)/시장가(ME) 비율 기준으로 상위 30%, 중위 40%, 하위 30%에 속하는 기업들로 3개의 포트폴리오(Value, Neutral, Growth)를 구성
- 3) 기업 규모 기준 2그룹과 장부가치/시장가치 기준 3그룹을 서로 교차해 6개의 포트폴리오 군 구성
- 4) SMB와 HML 측정

구분		장부가치(BE)/시장가치(ME)		
		높은 비율	중간비율	낮은 비율
기업규모	소규모	SH (Small Value)	SN (Small Neutral)	SL (Small Growth)
	대규모	BH (Big Value)	BN (Big Neutral)	BL (Big Growth)

### t-1년말 기준 세부포트폴리오 구성 (10분위수 : 10 × 10)

→ t년 7월 ~ t+1년 6월말까지 수익률 (월수익률 평균을 연율화)

→ 이때 , 포트폴리오 수익률은 소속 주식들의 월별수익률을 동일가중평균수익률

→ 이상치 조정 : 월별 수익률 데이터들에 winsorization을 적용

# 자산배분 : 선형 팩터 모델

## 'Fama French Three Factor Model'(1992)

### Factor 계산 사례

1.  $R_m - R_f = \text{Excess return on the market}$  → 시장수익률 : KOSPI 지수 수익률 → 상장된 모든 기업들의 횡단면수익률을 전기말의 횡단면시가총액으로 가중평균(value-weight) 한 것
2.  $SMB = \frac{(SH + SN + SL)}{3} - \frac{(BH + BN + BL)}{3}$ , S :기업규모가 작은 포트폴리오, B:기업규모가 큰 포트폴리오  
H :장부시장가치비율이 높은 포트폴리오, L: 장부시장가치비율이 낮은 포트폴리오  
B/M(가치주, 성장주, 중립주)와 시가총액의 크기(소형, 대형)로 나눠 소형가치주, 소형중립주, 소형성장주와 대형가치주, 대형중립주, 대형성장주
  - ME (market equity) =보통주 종가 X 보통주 발행주식수, BE (book equity) =자본총계-우선주자본금
  - B/M ( book-to-market equity ratio) =  $\frac{BE}{ME}$  → PBR의 역수로 이해
3.  $HML = \frac{(SH + BH)}{2} - \frac{(SL + BL)}{2}$   
[대형가치주(BH)와 소형가치주(SH)의 수익률 평균] - [대형성장주(BL)와 소형성장주(SL)의 수익률 평균]
  - HML은 가치주와 성장주 간의 수익률 스프레드를 설명  
HML 베타 계수(회귀계수)부호 의미 : 양(+)의 베타는 포트폴리오가 가치 프리미엄과 양의 관계를 가지고 있거나 포트폴리오가 가치주에 노출된 것처럼 행동한다는 것을 의미  
베타가 음수(-)이면 포트폴리오가 성장주 포트폴리오처럼 작동

# 자산배분 : 선형 팩터 모델

Fama, E.F., K.R. French, 2015, A **Five-Factor** Asset Pricing Model, Journal of Financial Economics, 116

Fama-French five-factor model In 2015, Fama and French extended the model, adding a further two factors  
→ 3factor + **수익성요인 (operating profitability)** + **투자도 요인(CMA)**

$$R_{it} - R_{Ft} = a_i + b_i(R_{Mt} - R_{Ft}) + s_iSMB_t + h_iHML_t + r_iRMW_t + c_iCMA_t + e_{it}.$$

구분		수익성요인 (operating profitability)		
		높은 비율	중간비율	낮은 비율
기업규모	소규모	SR (Small Robust)	SN (Small Neutral)	SW (Small Weak)
	대규모	BR (Big Robust)	BN (Big Neutral)	BW (Big Weak)

구분		투자도 요인(CMA)		
		높은 비율	중간비율	낮은 비율
기업규모	소규모	SC (Small Conservative)	SN (Small Neutral)	SA (Small Aggressive)
	대규모	BC (Big Conservative)	BN (Big Neutral)	BA (Big Aggressive)

# 자산배분 : 선형 팩터 모델

$$R_{it} - R_{Ft} = a_i + b_i(R_{Mt} - R_{Ft}) + s_iSMB_t + h_iHML_t + r_iRMW_t + c_iCMA_t + e_{it}.$$

## 4. **RMWt (Robust minus Weak) 수익성 요인 ( OP; operating profitability)**

(average return on the 2 robust operating profitability portfolios) - (average return on the 2 weak operating profitability portfolios) → **영업 수익성(OP : operating profitability) : 매출에서 비용을 뺀 금액**

$$OP = \frac{\text{매출액} - \text{매출원가} - \text{판관비} - \text{비영업 이자비용}}{\text{장부가치}(BE)}$$

$$= \frac{(SR + BR)}{2} - \frac{(SW + BW)}{2}, \text{ R : 수익성이 높은 포트폴리오, W : 수익성이 낮은 포트폴리오,}$$

## 5. **CMAt Conservative Minus Aggressive ) 투자도 요인(investment)**

(average return on the 2 conservative investment portfolios) - (average return on the 2 aggressive investment portfolios) 투자를 얼마나 보수적/공격적으로 하는지

$$Inv = \frac{\text{금년 회계기말 자산총계}}{\text{전년 회계기말 자산총계}} - 1$$

$$= \frac{(SC + BC)}{2} - \frac{(SA + BA)}{2}, \text{ A : 투자도(Inv)가 높은 포트폴리오, C : 투자도(Inv)가 낮은 포트폴리오}$$



# 자산배분 : 선형 팩터 모델

Fama, E.F., K.R. French, 2015, A **Five-Factor** Asset Pricing Model, Journal of Financial Economics, 116

- I/A(투자액/자산) :  $(\text{유형자산 변화분} + \text{재고자산 변화분}) \div \text{전년도 총자산}$
- ROE(자기자본이익률) :  $\text{당기순이익} \div \text{전년도 장부가치}$  혹은  $\text{경상이익} \div \text{전년도 장부가치}$
- ROA(총자산이익률) :  $\text{당기순이익} \div \text{전년도 총자산}$
- E/P(이익-주가 비율) :  $\text{당기순이익} \div \text{시가총액}$
- OCF/P(영업현금흐름-주가 비율) :  $\text{영업현금흐름} \div \text{시가총액}$
- CF/P(현금흐름-주가 비율) :  $\text{현금흐름}(\text{당기순이익} + \text{감가상각비}) \div \text{시가총액}$

# 자산배분 : 선형 팩터 모델

\*\*\*[http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data\\_library.html](http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html)

In this [paper](#) Fama and French explain how they produce the U.S. factor returns in their Data Library and they estimate the effect of the two changes in their process and five major CRSP data-improvement projects on the average values of SMB and HML.



**Eugene F. Fama**  
The Robert R. McCormick Distinguished Service Professor of Finance at the University of Chicago Booth School of Business



**Kenneth R. French**  
The Roth Family Distinguished Professor of Finance at the Tuck School of Business at Dartmouth College

	January 2024	Last 3 Months	Last 12 Months
<b>Fama/French 3 Research Factors</b>			
Rm-Rf	0.70	15.05	14.66
SMB	-5.07	0.72	-13.39
HML	-2.39	4.25	-11.56
<b>Fama/French 5 Research Factors (2x3)</b>			
Rm-Rf	0.70	15.05	14.66
SMB	-5.73	0.90	-16.13
HML	-2.39	4.25	-11.56
RMW	0.68	-6.67	8.52
CMA	-0.95	-0.72	-15.47
<b>Fama/French Research Portfolios</b>			
<b>Size and Book-to-Market Portfolios</b>			
Small Value	-5.15	20.92	-0.74
Small Neutral	-4.24	16.91	5.60
Small Growth	-4.11	18.30	-1.99
Big Value	-2.00	22.07	6.03
Big Neutral	1.98	15.71	6.61
Big Growth	1.74	16.18	30.40

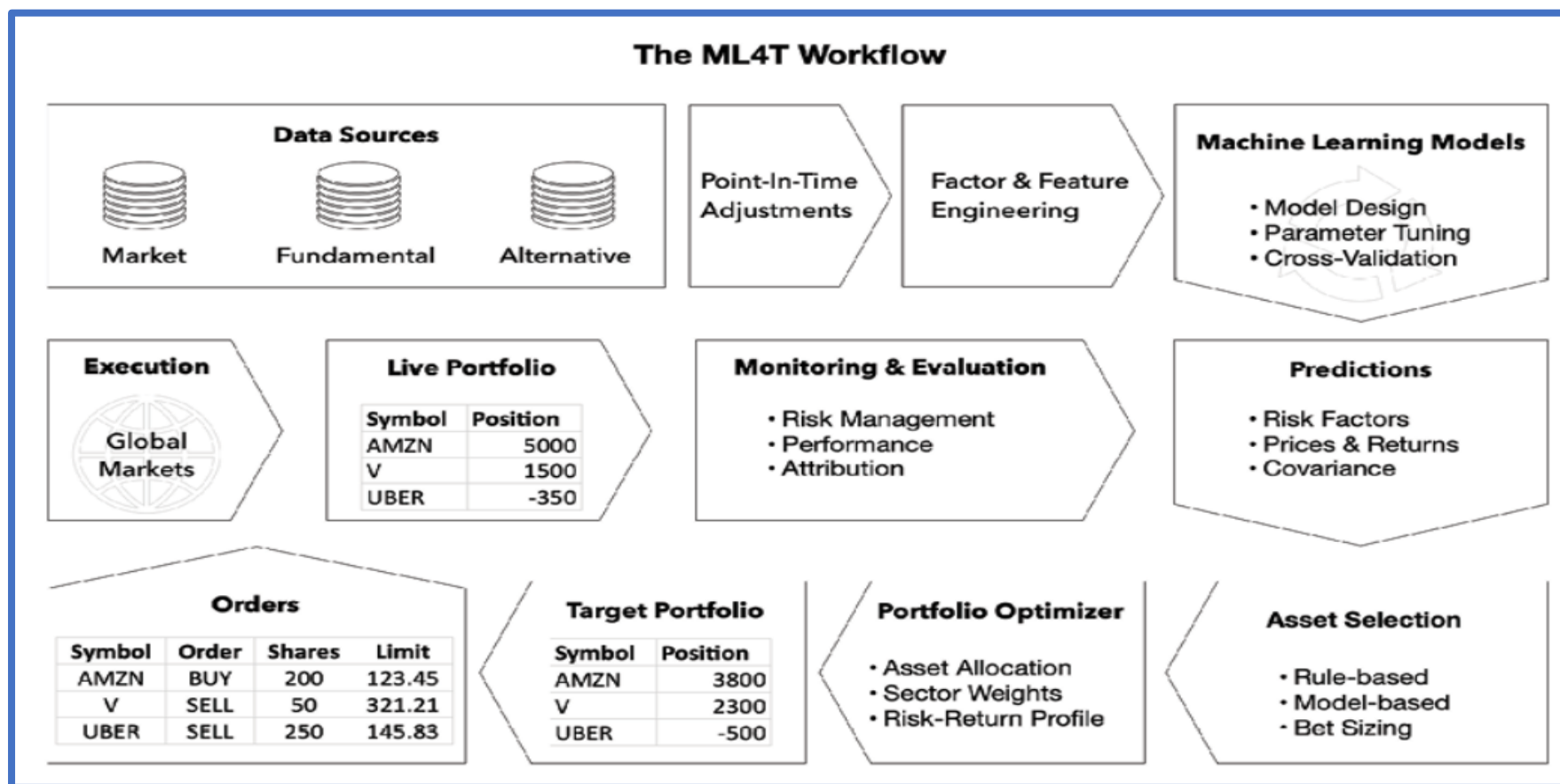
# 리스크 줄이기: 중립화 전략

전통적 Risk factor : 베타 factor, 업종 factor, 사이즈 factor

- 퀀트에서는 중립화 전략이 필요.
- 베타 중립화 전략: 5개 베타군별로 나눠서
- 소속 베타군의 평균 P/E보다 Forward P/E 할인율이 큰 종목 순으로 종목 선택.
- 디스카운트 큰 종목 Long, 프리미엄 높은 종목 Short.
- 베타군은, 전체 유니버스의 현재 베타수치를 기준으로 정렬하여 20%씩 그룹을 형성

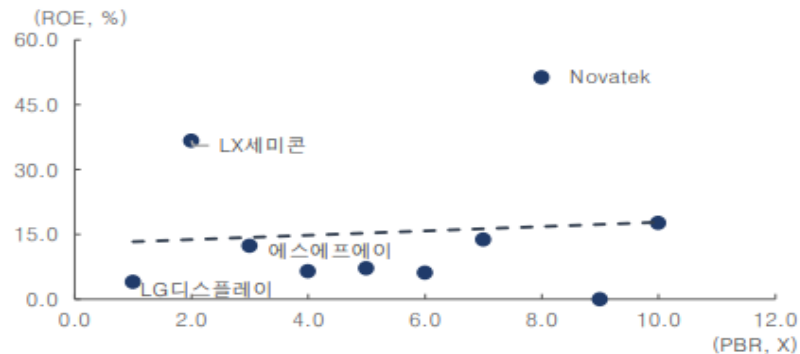
# Machine Learning for Trading: From Idea to Execution

## Designing and executing an ML-driven strategy



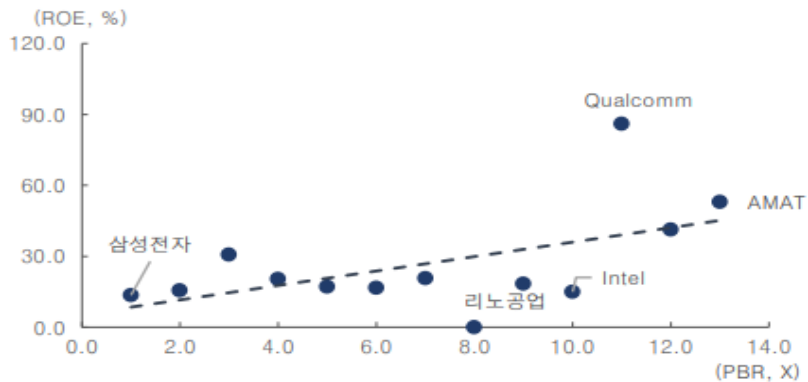
**Machine Learning for Algorithmic Trading:** Predictive models to extract signals from market and alternative data for systematic trading strategies with Python

디스플레이 22년 ROE-PBR



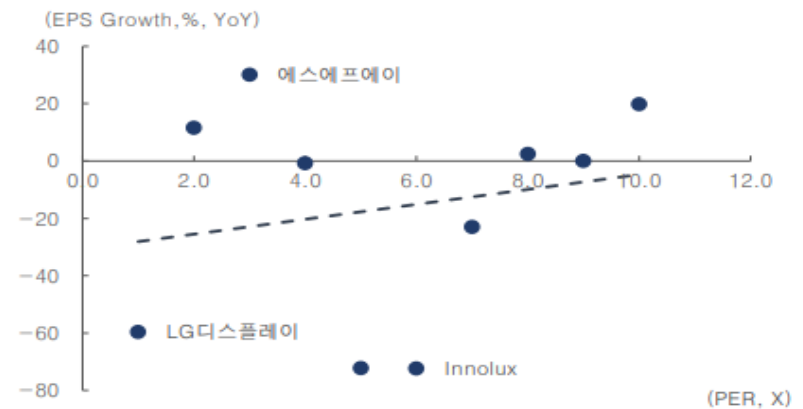
자료: 교보증권 리서치센터

반도체 22년 ROE-PBR



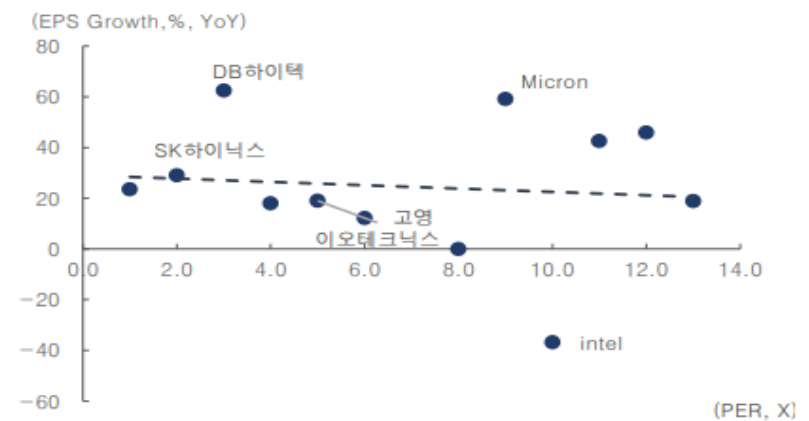
자료: 교보증권 리서치센터

디스플레이 22년 PER-EPS Growth



자료: 교보증권 리서치센터

반도체 22년 PER-EPS Growth



자료: 교보증권 리서치센터

# 포트폴리오 성과 비교

## 1. 가치투자전략(Value Strategy)

: 기업의 기본적 실적을 현재 주가와 비교하여 가치비율을 구하고 이 가치비율의 높고 낮음에 따라 포트폴리오를 분류하여 저평가된 가치주 포트폴리오를 매입하고 고평가된 성장주 포트폴리오를 매도하는 투자전략

## 2. Momentum Strategy

: 과거에 높은 수익률을 기록하였던 승자(Winner)포트폴리오들이 계속해서 미래에도 좋은 성과를 기록하는 관성(Momentum)을 갖고 있을 경우 승자를 매입하고 패자(Loser)를 매도하는 투자전략

## 3. Contrarian Strategy

과거에 저조한 수익률을 기록하였던 포트폴리오가 가격이 상승하여 높은 수익률을 기록할 경우 패자 포트폴리오를 매입하고 승자 포트폴리오를 매도하는 전략

# 포트폴리오 성과 비교

## Momentum Strategy 성과 분석

- 절대 모멘텀 전략 : 주식 시장 자체의 수익률(예: KOSPI)이 과거 3개월, 6개월, 또는 12개월 이전에 비해서 상승하면 투자하고 하락했다면 보유
- 상대 모멘텀 전략 : 주식 시장에 상장된 종목 중 최근에 가장 많이 오른 종목을 매수한 후 일정 기간 보유한 후 매도

### 1. 일정기간 수익률에 의한 포트폴리오를 구성한 후

- 성과가 좋은 기업들의 주식을 매입하고 성과가 좋지 않은 기업들의 주식을 매도하는 전략
- 양(+)의 투자성과가 발생하는지 여부 검증
- 기업규모,장부가치 대 시장가치 비율을 결합하여 일주일 수익률 투자전략이 수익률의 움직임 에서 일관성 있게 유지되는 모멘텀 현상을 갖는지 아니면 반전하는 현상을 갖는지 검증
- 일주일 수익률에 기초하여 매주 말에 10개의 포트폴리오를 구성한 후 1주일 간,2주일간,3주일간,4-52주일간,1-52주일간 보유하는 전략을 취한 후 성과를 검증
- 일주일 수익률을 기초로 주식수익률을 정렬→ 정렬한 후 포트폴리오 구성일 현재 성과가 가장 좋은 포트폴리오와 성과가 가장 좋지 않은 포트폴리오 사이의 차이를 검증

$$\text{*모멘텀(수익률)} = [(\text{금일주가} - \text{과거 특정일 주가}) / (\text{과거 특정일 주가})] \times 100$$

Jegadeesh & Titman(1993) : 'Returns to Buying Winners and Selling Losers: Implications for Stock Market Efficiency'  
상대 모멘텀 전략

# 포트폴리오 성과 비교

## 2. Fama and French(1992)의 3요인

### 1) 기업규모에 대한 위험프리미엄 (SMB)

동일가중 포트폴리오 수익률인 SL, SM, SH의 평균수익률에서 BL, BM, BH의 평균수익률을 차감하여 계산

$$SMB_t = \frac{1}{3} \sum_{i=1}^3 R_{S_{it}} - \frac{1}{3} \sum_{i=1}^3 R_{B_{it}}$$

### 2) 장부가치/시장가치비율에 대한 위험프리미엄 (HML)

동일가중 포트폴리오 수익률인 SH, BH의 평균수익률에서 SL, BL의 평균수익률을 차감하여 계산

$$HML_t = \frac{1}{2} \sum_{i=1}^2 R_{H_{it}} - \frac{1}{2} \sum_{i=1}^2 R_{L_{it}}$$

이와 같이 계산된 SMB와 HML은 식 (9)의 회귀분석에서 X변수가 수익률 R을 설명하는지의 여부를 검증할 경우에 통제변수로 사용

$$R_t = a + sSMB_t + hHML_t + b(X_t) + e_t$$



# 포트폴리오 성과 비교

## 3. Fama-MacBeth의 방법론을 이용한 횡단면 회귀분석

장부가/시가비율과 기업규모를 통제하고 → 포트폴리오 구성일 이 전 6개월의 정보가 이후 미래 1년간의 주식수익률에 미치는 영향을 살펴볼 수 있다

$$R_{[t+1,t+52]} = \gamma_0 + \gamma_1 r_t + \gamma_2 r_{[t-26,t-1]} + \gamma_3 BE/ME + \gamma_4 \ln(Size) + e_i$$

- 1) 표본기간을 하위 표본기간 15년씩으로 분류
- 2) 이 기간을 각각 5 년씩 포트폴리오 구성기간, 포트폴리오 베타 추정기간, 검증기간으로 구분
- 3) 추정된 베타의 크기에 따라 증권들을 정렬한 다음, 동일한 기업수로 10개의 포트폴리오들을 구성  
가장 높은 베타들로 구성된 증권들을 첫 번째 포트폴리오로 하고, 가장 낮은 베타들로 구성된 증권들을 마지막 10번째 포트폴리오로 한다.
- 4) 두 번째 5년의 기간 동안에 포트폴리오의 수익률(포트폴리오를 구성하고 있는 증권들의 동일 가중수익률)을 종속변수로 하고 시장수익률을 설명변수로 하여 회귀분석을 이용하여 포트폴리오들의 베타들을 추정
- 5) 포트폴리오의 베타와 주식수익률 간의 관계를 검증하는 단계 : 추정된 15개의 포트폴리오의 베타와 포트폴리오에 속하는 개별증권의 1개월간 의 주식수익률을 횡단면 회귀분석을 실시하여 회귀계수 추정 → 이러한 검증을 이용하여 한국 주식시장에서 양(+)의 위험-수익률간의 상반관계 (trade-off)를 검증

# 금융권 빅데이터 활용 사례 : 주가예측

## 파이썬 Tensorflow LSTM RNN을 이용하여 인공지능 주가 예측하기

- Long - Short - Term - Memory (LSTM) 네트워크는 시간경과에 따른 Backpropagation을 사용하여 훈련되는 인공신경망 네트워크
- 기존 순환신경망(RNN: **Recurrent Neural Networks**)에서 이전의 데이터를 통해 학습된 셀의 상태 정보가 다음 데이터를 이용하여 학습 시킬 때 사용 된다는 의미→ 이는 특히 시계열데이터를 처리할 때 적합
- 기존 STM 한계 극복 : 기본적인 순환신경망은 단기 기억을 저장→ 데이터들의 연관정보를 분석하기 위하여 보다 장기적인 기억 알고리즘 탄생(LSTM)
- OHLCV (open, high, low, close, volume)로 당일 시가, 고가, 저가, 종가, 거래량에 대한 데이터

## Keras를 활용한 주식 가격 예측

A deep learning framework for financial time series using stacked autoencoders and longshort term memory Wei Bao1 , Jun Yue2 \*, Yulei Rao 2017

# 금융권 빅데이터 활용 사례 : 주가예측

## 시계열 예측을 위한 Facebook Prophet 사용하기

<https://facebook.github.io/prophet/>

[https://facebook.github.io/prophet/docs/quick\\_start.html](https://facebook.github.io/prophet/docs/quick_start.html)

## Forecasting at Scale

Sean J. Taylor\*†

Facebook, Menlo Park, California, United States

sjt@fb.com

and

Benjamin Letham†

Facebook, Menlo Park, California, United States

bletham@fb.com

### Abstract

Forecasting is a common data science task that helps organizations with capacity planning, goal setting, and anomaly detection. Despite its importance, there are serious challenges associated with producing reliable and high quality forecasts – especially when there are a variety of time series and analysts with expertise in time series modeling are relatively rare. To address these challenges, we describe a practical approach to forecasting “at scale” that combines configurable models with analyst-in-the-loop performance analysis. We propose a modular regression model with interpretable parameters that can be intuitively adjusted by analysts with domain knowledge about the time series. We describe performance analyses to compare and evaluate forecasting procedures, and automatically flag forecasts for manual review and adjustment. Tools that help analysts to use their expertise most effectively enable reliable, practical forecasting of business time series.

*Keywords:* Time Series, Statistical Practice, Nonlinear Regression

## PROPHET

### Documentation

#### Installation

Using R

Using Python

#### Quick Start

Python API

R API

#### Saturating Forecasts

Forecasting Growth

Saturating Minimum

#### Trend ChangePoints

Automatic changepoint detection in Prophet

Adjusting trend flexibility

Specifying the locations of the changepoints

## Installation

Prophet has two implementations: [R](#) and [Python](#).

## Installation in R

Prophet is a [CRAN package](#) so you can use `install.packages`.

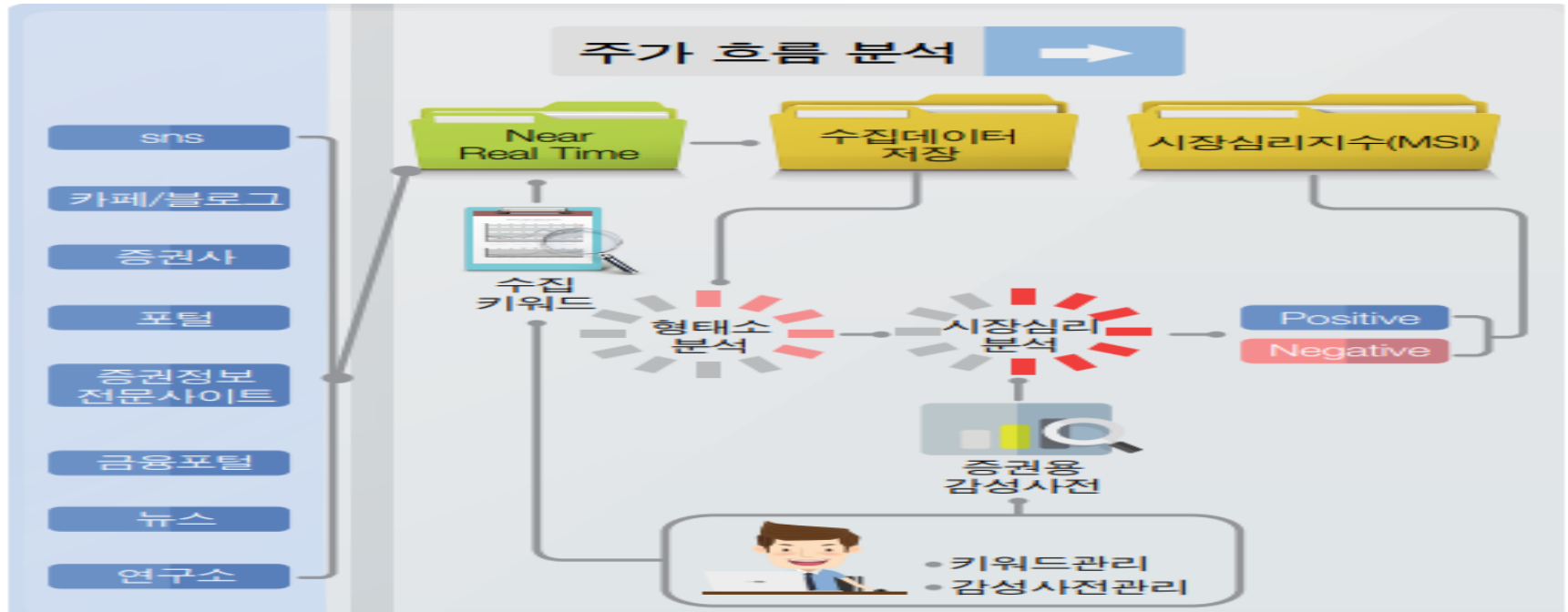
```
1 # R
2 install.packages('prophet')
```

## 금융권 빅데이터 활용 사례 : 주가예측

- simple regression analysis
- multiple regression analysis
- Logistic Regression
- RandomForest Regressor

## \* 테스트 마이닝 : 감성분석(시장심리분석)

뉴시스-코스콤'빅데이터 시장심리지수(MSI)'



데이터 건수-MSI(시장심리지수)스코어-MSI지수 레벨-버즈워드  
주가예측 위한 4대 정성평가 데이터가 핵심

## \* SNS 와 주가 금리

### Bank Of America

: 트럼프 대통령의 트윗 양과 스탠다드앤드푸어스(S&P) 500 지수의 상관관계

→ 트럼프의 트윗과 주가수익률이 역(-)의 상관관계

### JP모건

트럼프 대통령의 트윗이 채권시장의 변동성에 미치는 영향을 분석한 '볼피피지수'(Volfefe Index)

볼피피란 트럼프 대통령이 지난 2017년 자신의 트위터에 "늘 부정적인 언론의 '코브피피'(covfefe)에도

불구하고"라는 문장을 쓰면서 쓴 '코브피피'라는 단어와 변동성을 의미하는 '볼라틸리티'(Volatility)를 결합한 단어

→ 트럼프 대통령의 트윗이 미국 금리에 어떤 변동성을 야기하는지 분석한 지수로 2년물과 10년물 국채의 내재 변동성을 활용

→ 일일 트윗과 수익률 음(-)

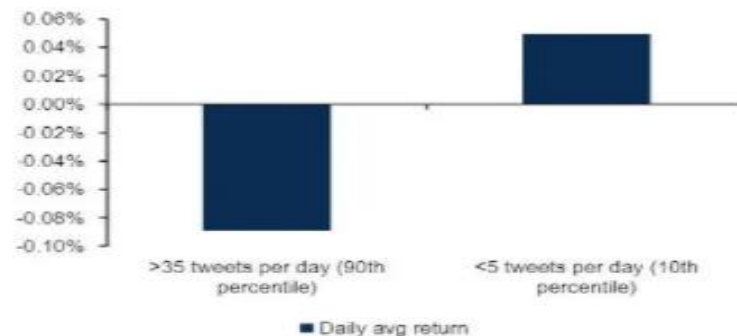
Chart 4: Trade talk and tweets by President Trump have contributed to volatility of late, from China tariffs to Fed policy to tax policy and more S&P 500 Index and number of daily Tweets by Trump



Source: BofA Merrill Lynch US Equity & Quant Strategy, Twitter, S&P

Chart 5: The market tends to fall, on average by 9bp, on days Trumps is more active on Twitter

Average daily returns when Trump writes more than 35 Tweets a day (90<sup>th</sup> percentile) versus less than five Tweets a day (10<sup>th</sup> percentile), since 2016



Source: BofA Merrill Lynch US Equity & Quant Strategy, Twitter, S&P

# 시작자가 배워야 할 투자 개념

1. 재무 계획부터 시작
2. 저축을 최우선
3. 복리의 힘 이해
4. 위험과 불확실성 구분 이해
5. 분산투자 및 자산배분 이해
6. 고전적인 투자 전략 이해
7. 아는 것에만 먼저 투자하라
8. 대출을 먼저 상환하라.