U.FE.A
Buyside Seminar

Risk Factors, Overfitting, and Robustness

송용재, 배동익, 이영우, 이재형

Contents

1. CAPM, APT와 실증적 연구

36기 송용재

CAPM, APT와 실증적 연구
: 여러 가정들

- **완전자본시장 가정**

    1) 소수의 투자자가 가격에 영향 불가. 모든 투자자는 Price taker.

    2) 자산을 무한히 작게 쪼개서 거래 가능

    3) 거래비용 0

    4) 세금 없음

    5) 모든 투자자가 동일한 정보를 알고 있음

    6) 공매도 제한 없음

- **동질적 기대 가정**

CAPM, APT와 실증적 연구
: Theory of Portfolio Selection – Harry Markowitz

- **가정**

  모든 투자자는 위험회피이고 합리적, 완전자본시장, 동질적
  기대, 자산수익률의 확률분포는 Time-invariant한 정규분포.

- **결과**

  포트폴리오에 속한 각 자산의 **투자 비율(w)에 따라서, 기대수익과 위험이 도출.**
  최적화된 포트폴리오에서는 둘 사이의 trade-off (하나가 좋아지면 다른 게 나빠짐)가 존재하므로, efficient portfolio(효율적 포트폴리오)는 (예산선 내에서) 위험을 늘리지 않고는 기대수익도 늘릴 수 없는 포트폴리오를 의미하며, 이는 w-space(투자 비율을 각 축으로 하는 공간)에서 line으로 나타남.

- 포트폴리오의 기대수익

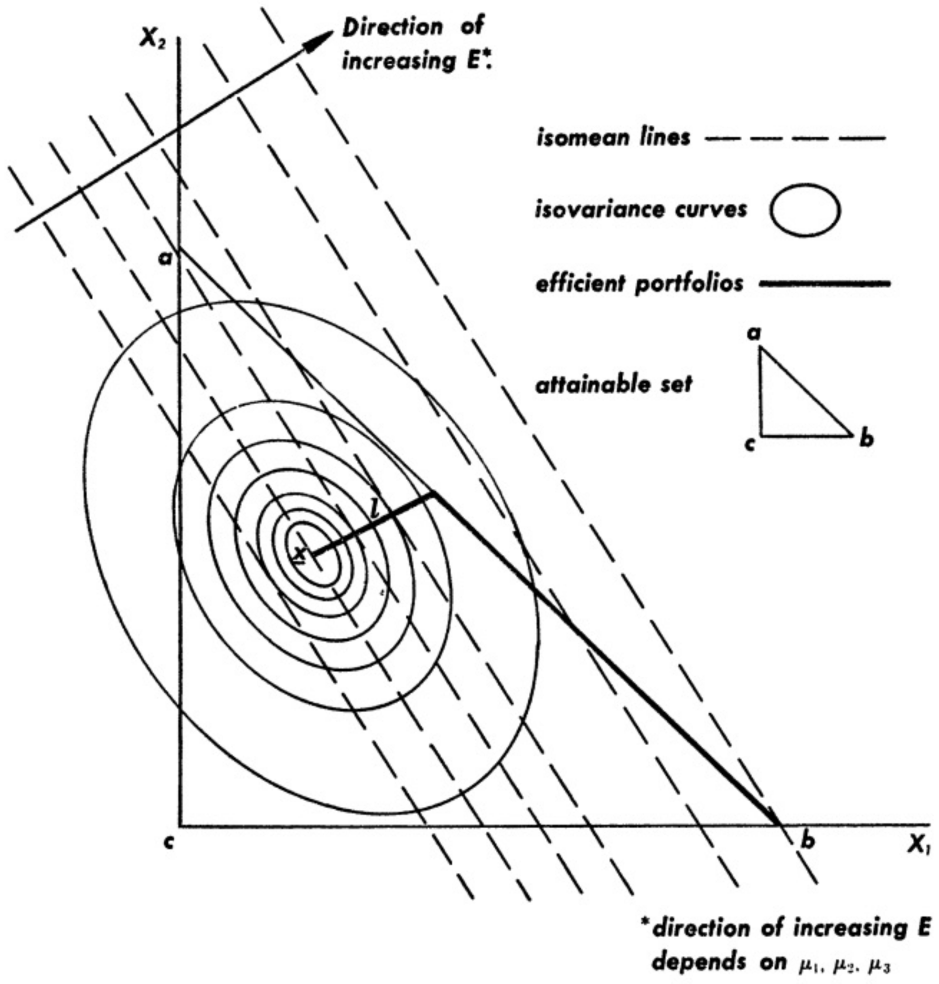
$$\mathrm{E}(R_p) = \sum_i w_i \, \mathrm{E}(R_i)$$

- 포트폴리오의 위험(분산)

$$\sigma_p^2 = \sum_i \sum_j w_i w_j \sigma_i \sigma_j \rho_{ij}$$

CAPM, APT와 실증적 연구

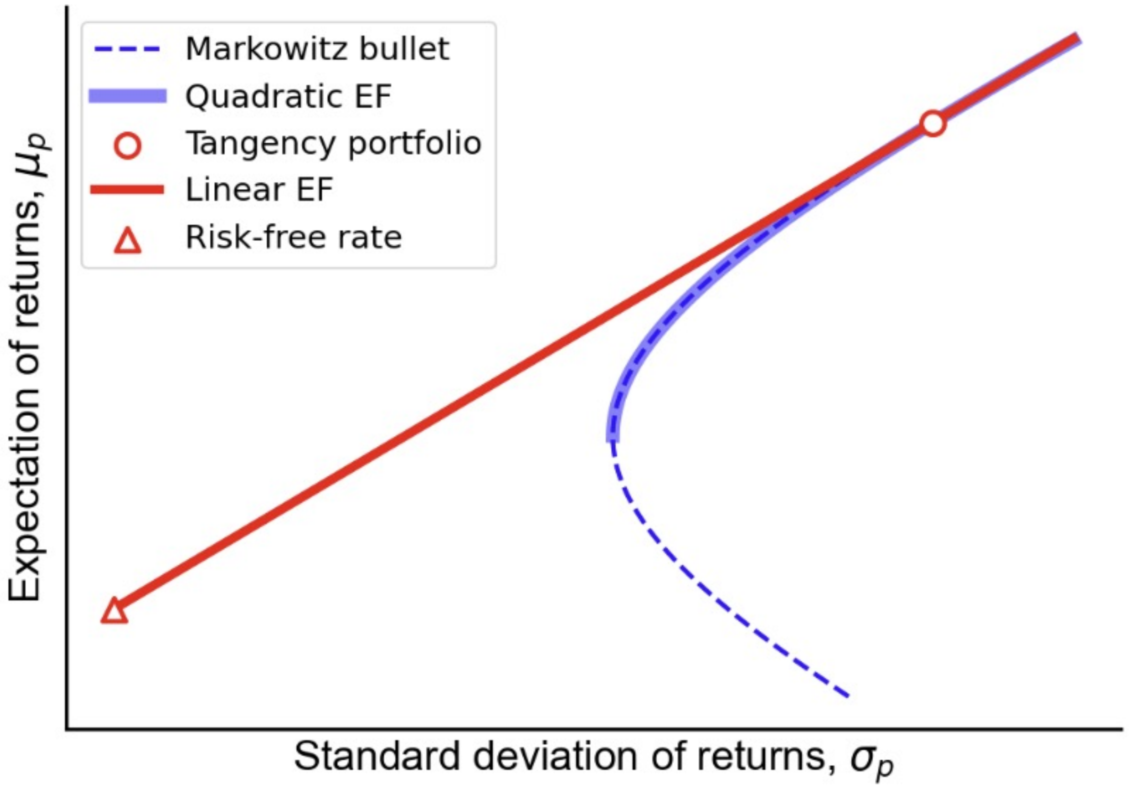
: Theory of Portfolio Selection – Harry Markowitz



CAPM, APT와 실증적 연구
: Mean-Variance Analysis

(Mean) - (Standard deviation) Space를 생각하자. (임의의 static 자산을 plot 가능)

- **efficient frontier:** 시장의 모든 위험자산을 고려한 efficient portfolio의 집합
- **capital allocation line(CAL):** 한 개 또는 몇 개의 포트폴리오와 무위험자산을 섞었을때 생기는 직선 -> **왜 직선인가?**
- **capital market line(CML):** 가장 효율적인 CAL로, 무위험 자산인 point에서 efficient frontier에 그은 접선 -> **가장 효율적?**



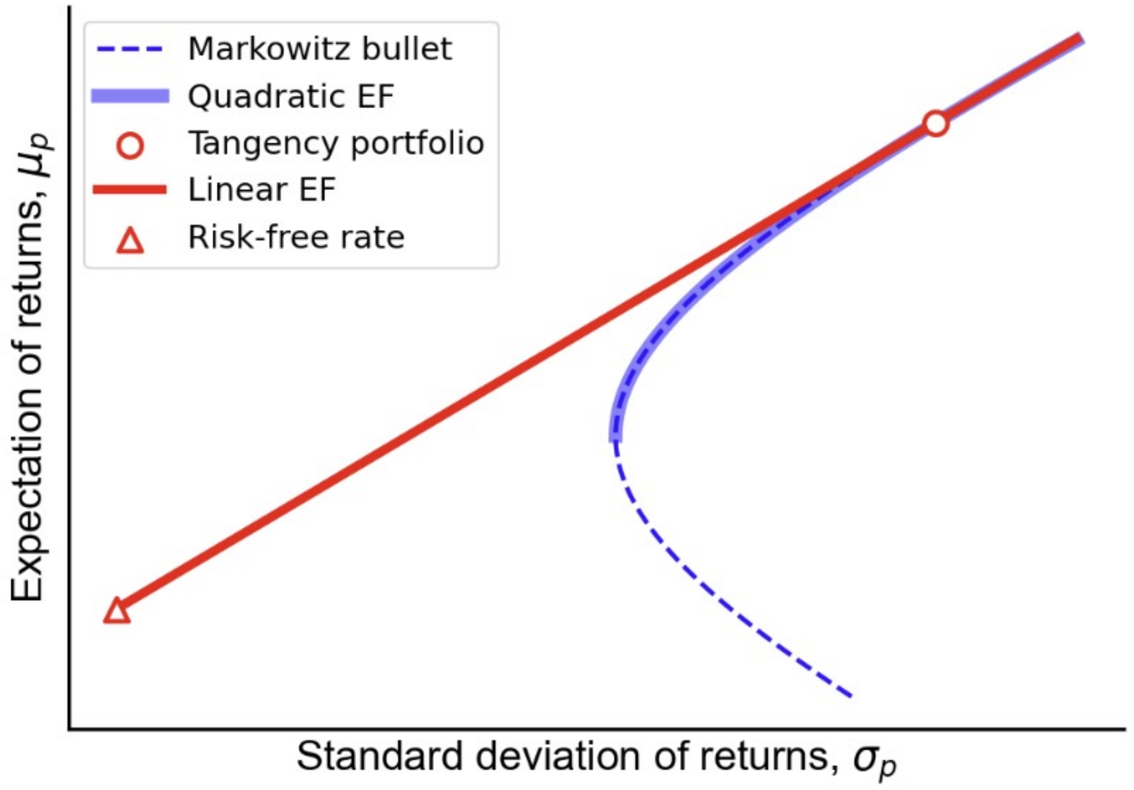
CAPM, APT와 실증적 연구
: Sharpe Ratio(샤프 지수)

샤프 지수는 **(위험에 따른 기대수익)**을 **(위험)**으로 나눈 **값**으로 **단위 위험당 기대수익**을 나타낸다고 볼 수 있다.

$$\text{sharpe} = \frac{\mu_p - r_f}{\sigma_p}.$$

CAL의 기울기가 **샤프 지수**이고, **CML은 샤프 지수가 가장 큰 CAL**임을 알 수 있다.

그리고 그 **접점의 최적화된 포트폴리오를 market portfolio(M)**라고 한다.



CAPM, APT와 실증적 연구

: Derive CAPM

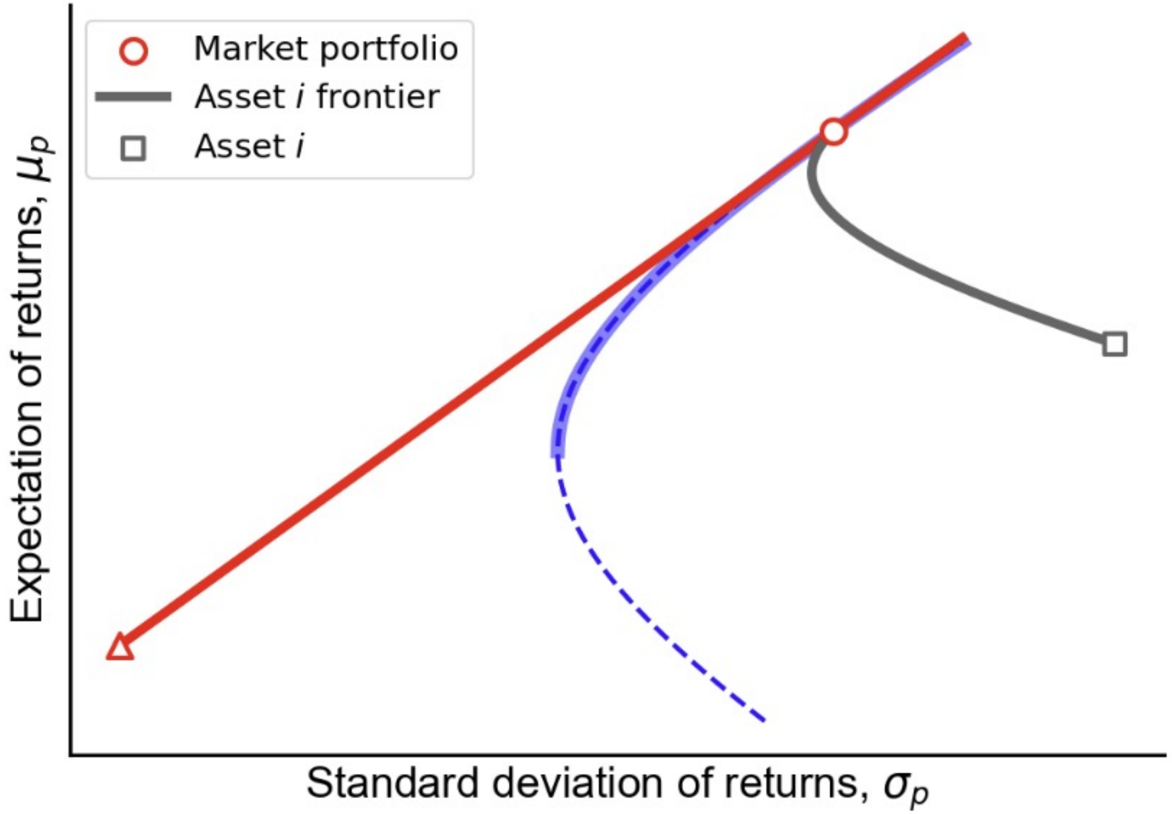
(Asset i frontier's slope at M)

= (Sharpe Ratio) ← Why?

$$\frac{\mathbb{E}[R_M] - r_f}{\sigma_M} = \frac{\sigma_M(\mathbb{E}[R_i] - \mathbb{E}[R_M])}{\sigma_{i,M} - \sigma_M^2}$$

정리하면,

$$\frac{\sigma_{i,M}}{\sigma_M^2}(\mathbb{E}[R_M] - r_f) = \mathbb{E}[R_i] - r_f$$



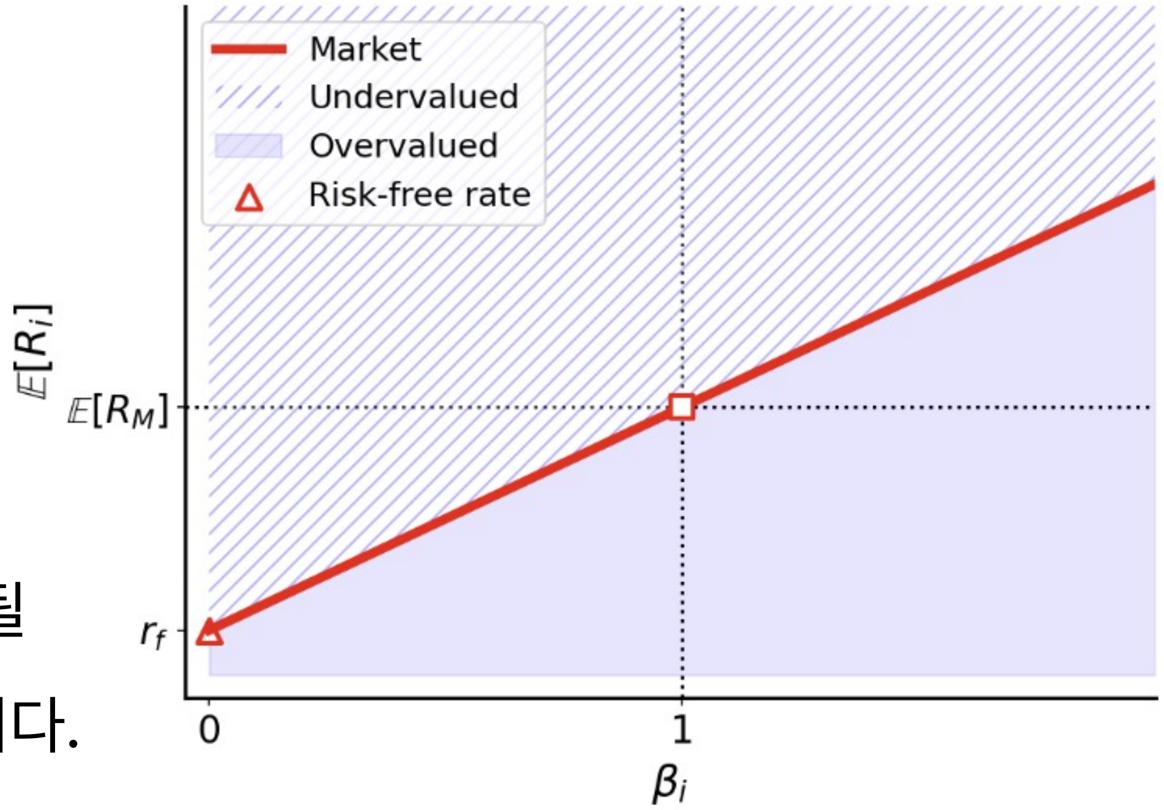
CAPM, APT와 실증적 연구

: Security Market Line(증권시장선, SML)

$$\beta_i = \frac{\sigma_{M,i}}{\sigma_M^2} \quad \beta_p = \frac{\sigma_{M,p}}{\sigma_M^2} = \sum_{i=1}^{n} \alpha_i \beta_i$$

증권시장선이란, **베타에 따른 기대수익률 곡선**으로 어떤 상품의 베타 계수를 안다면 이로부터 **기대수익률을 추정**할 수 있고, 예측된 미래 주가에 대해 주식이 지금 **과대평가나 과소평가되었는지를 판단**할 수 있다.

ex) 어떤 시장에서 주식 A의 SML로 계산한 기대수익률이 100% 라면, 내년에 2000원이 될 것으로 예상된 A의 현재 적정 가격은 1000원이다.



CAPM, APT와 실증적 연구

: 총위험 vs 체계적 위험(CML vs SML)

총위험 = 체계적 위험 + 비체계적 위험

$$\sigma_i^2 = \beta_i^2 \sigma_M^2 + var(\epsilon_i).$$

체계적 위험:

시장 불확실성과 연관된 부분 때문에 발생하며, 분산투자로 없앨 수 없음

비체계적 위험:

개별 주식의 불확실성과 연관되며, 시장과 무관한 랜덤위험이므로 분산투자로 없앨 수 있음

(큰 수의 법칙)

CML(총위험) VS SML(체계적 위험):

$$\overline{r} = r_f + \frac{\overline{r}_M - r_f}{\sigma_M}\sigma \qquad \overline{r}_i - r_f = \beta_i(\overline{r}_M - r_f)$$

CAPM, APT와 실증적 연구
: Asset Pricing

미래 시점에서, 증권의 기댓값을 수익률의 기댓값으로 다음과 같이 할인하면 된다. 주목할 만한 부분은 **function P(Q)가 linear**하다는 것이다.

$$P = \overline{Q}/(1+\overline{r})$$

$$P = \frac{\overline{Q}}{1 + r_f + \beta(\overline{r}_M - r_f)}$$

$$P = \frac{\overline{Q} - \frac{Cov(Q, r_M)(\overline{r}_M - r_f)}{\sigma_M^2}}{1 + r_f}.$$

$$-\frac{Cov(Q, r_M)(\overline{r}_M - r_f)}{\sigma_M^2}$$

Risk Premium처럼 생겼음.

위 term은 무위험자산에 투자했을 때의 수익과 Q에 투자했을 때의 수익의 차이를 의미한다.

If, Market과 Q가 Positively correlated, 시장수익률이 무위험수익률을 이길수록 위 값은 감소하는 것이 타당하고 실제로 그러함.

CAPM, APT와 실증적 연구
: CAPM(Capital Asset Pricing Model)’s Assumption

1. **무위험자산이 존재하고 제한 없이 이용가능**
2. 완전자본시장
3. 동질적 기대
4. 자산수익률 확률분포의 정규분포 가정
5. 위험회피성향(효용함수가 2차함수)인 합리적 투자자
6. 단일 기간 가정(확률분포는 단일 기간 동안 일정)

CAPM, APT와 실증적 연구
: Existence of Alpha

다음과 같은 선형회귀(Linear Regression)을 고려하자. 에러항은 zero-mean이다.

$$\mathbb{E}[R_i] - r_f = \alpha_i + \beta_i(\mathbb{E}[R_M] - r_f) + \varepsilon_i$$

$$\mathbb{E}[\varepsilon_i] = 0$$

CAPM이 성립한다면(CAPM의 가정이 성립한다면), 다음을 안다.

$$\mathbb{E}[\alpha_i] = 0$$

즉, 알파는 없다.(Market is efficient)

CAPM, APT와 실증적 연구

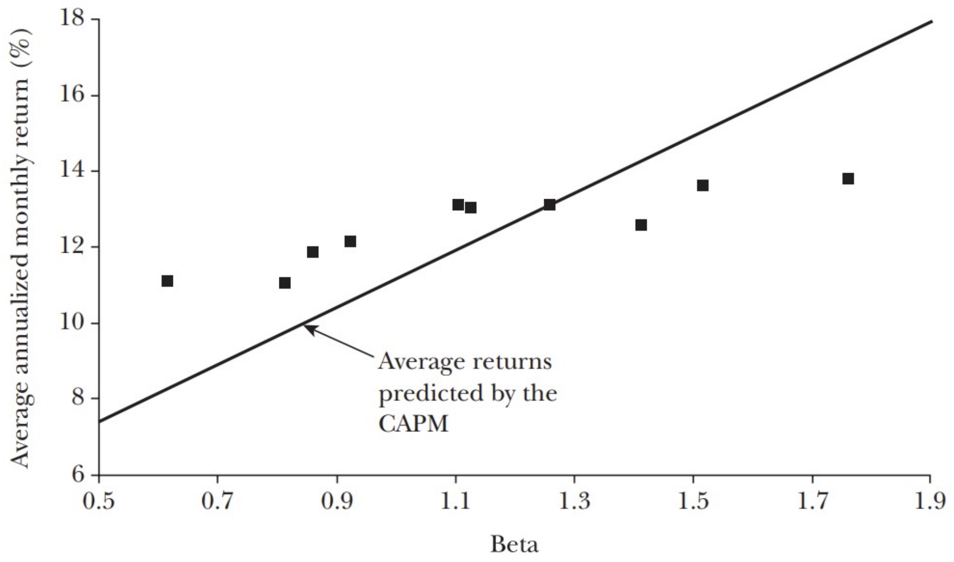
: CAPM의 실증적 연구

그럼 정말 알파가 없는가? 현실적으로 CAPM이 성립하는지 실증연구를 해 보자.

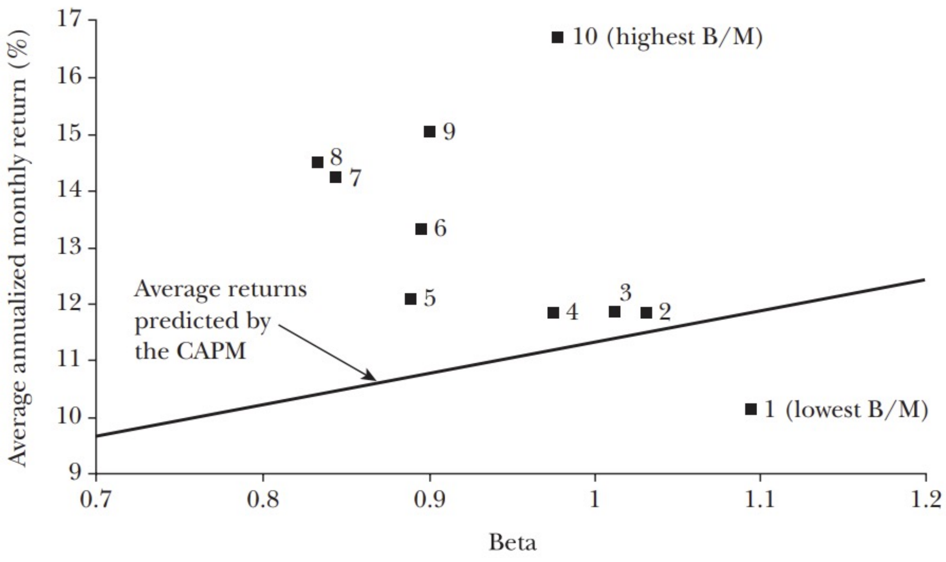
“Jensen’s Alpha”: SML에서 떨어진 정도.

$$\text{(Time-Series Regression)} \quad R_{it} - R_{ft} = \alpha_i + \beta_{iM}(R_{Mt} - R_{ft}) + \varepsilon_{it}$$

Average Annualized Monthly Return versus Beta for Value Weight Portfolios Formed on Prior Beta, 1928–2003



Average Annualized Monthly Return versus Beta for Value Weight Portfolios Formed on B/M, 1963–2003



CAPM, APT와 실증적 연구
: 롤의 비판

기존: 시장에서 **기대수익률과 베타의 선형관계** → **CAPM이 성립**한다고 생각함

기대수익률과 베타의 선형관계를 보기 위해, 연구에서는 코스피 지수 같은 어떤 **index를 Market Portfolio라고 가정**한다. 이때 **선형관계**가 의미하는 것은 **index가 efficient portfolio**라는 것이지 **CAPM의 성립여부가 아니다**. Efficient Portfolio라면 선형관계를 만족하기 때문이다.

Real Market Portfolio가 efficient Portfolio임을 보여야 CAPM의 성립여부를 증명한 것이다. **그러나 실제 시장 포트폴리오를 구성하는 것은 불가능**하므로 사실상 **CAPM의 실증 연구는 불가능**하다.

CAPM, APT와 실증적 연구
: APT(Arbitrage Pricing Theory)

가정:

1. 투자자는 위험회피 성향, 동질적 기대
2. 차익거래 없는 효율적 시장
3. 완전자본시장
4. 무한한 개수의 자산
5. 수익률은 선형 다요인 모형을 따를 것이다.

CAPM, APT와 실증적 연구

: Derive APT

자산의 수익률은 선형 다요인 모형을 따르므로,

$$r_i = \alpha_i + b_{1i}F_1 + b_{2i}F_2 + \dots + b_{Ki}F_K + \varepsilon_i$$

기대값을 취하면,

$$E(r_i) = \alpha_i + b_{1i}E(F_1) + b_{2i}E(F_2) + \dots + b_{Ki}E(F_K)$$

변변 빼주면,

$$\begin{aligned} r_i - E(r_i) = {} & (\alpha_i + b_{1i}F_1 + \dots + b_{Ki}F_K + \varepsilon_i) \\ & - (\alpha_i + b_{1i}E(F_1) + \dots + b_{Ki}E(F_K)) \end{aligned}$$

즉,

$$r_i = E(r_i) + \sum_{k=1}^{K} b_{ik}(F_k - E(F_k)) + \varepsilon_i$$

CAPM, APT와 실증적 연구
: Derive APT

다음과 같이 포트폴리오를 구성하자.

$$r_p = \sum_{i=1}^{I} x_i r_i$$

앞에서 구한 r_i를 대입하면 포트폴리오의 수익률은,

$$r_p = \sum_{i=1}^{I} x_i E(r_i) + \sum_{i=1}^{I} \sum_{k=1}^{K} x_i b_{ik}(F_k - E(F_k)) + \sum_{i=1}^{I} x_i \varepsilon_i$$

이때 위의 맨 마지막 항은 비체계적 위험 term이므로 분산투자를 하면 사라진다.

CAPM, APT와 실증적 연구
: Derive APT

다음과 같이 포트폴리오를 구성하자.

$$r_p = \sum_{i=1}^{I} x_i r_i$$

앞에서 구한 r_i를 대입하면 포트폴리오의 수익률은,

$$r_p = \sum_{i=1}^{I} x_i E(r_i) + \sum_{i=1}^{I} \sum_{k=1}^{K} x_i b_{ik}(F_k - E(F_k)) + \sum_{i=1}^{I} x_i \varepsilon_i$$

이때 위의 맨 마지막 항은 비체계적 위험 **term이므로 분산투자를 하면 사라진다.**

CAPM, APT와 실증적 연구

: Derive APT

다음으로, **APT의 균형 조건은 차익거래가 없다는 것**이다. 즉, 다음과 같이 **net investment가 zero**이고 **체계적 위험이 zero(beta's linearity!)**이면,

$$\sum_i x_i = 0 \qquad \sum_i x_i b_{ik} = 0 \text{ for all k}$$

기대수익률도 zero이어야 한다.

$$\sum_i x_i E(r_i) = 0$$

그러면, 자산 개수를 I라고 할 때, I차원 공간에서 k+1개의 독립인 벡터들의 set W={(1,...,1), (b_{i 1}),... (b_{i k})}이 있고, **W의 orthogonal complement의 임의의 원소 x_i에 대해 E(r_i)라는 벡터가 직교하는 상황**이므로, **E(r_i)**는 W의 orthogonal complement의 orthogonal complement, 즉 **W의 원소이다.**

CAPM, APT와 실증적 연구

: Derive APT

따라서 **E(r_i)는 span(W)의 원소**이므로 **basis(=W)의 선형결합**으로 나타낼 수 있다.

$$E(r_i) = \lambda_0 + \lambda_1 b_{1i} + \lambda_2 b_{2i} + \dots + \lambda_k b_{ki}$$

이제 각 계수들의 의미를 살펴보자. **b_{i k}는 asset(or portfolio) i와 k-th factor (factor set is independent)의 공분산에 비례**하므로,

1. 포트폴리오를 **무위험**으로 짜면, **b_{p k}는 0**이므로

$$r_f = \lambda_0 + \lambda_1 0 + .. + \lambda_k 0 \Rightarrow \lambda_0 = r_f$$

2. 포트폴리오를 **factor 1과 같게** 짜면, **b_{p 1}=1이고 나머지는 0**이므로

$$E(F_1) = r_f + \lambda_1 1 + \lambda_1 0 + .. + \lambda_k 0 \Rightarrow \lambda_1 = E(F_1) - r_f$$

CAPM, APT와 실증적 연구
: APT

따라서 APT의 균형 식은 다음과 같다.

$$E(r_i) = r_f + \sum_{k=1}^{K} b_{ik}(E(F_k) - r_f)$$

APT도 CAPM과 비슷하게 time series 회귀분석으로 실증적 연구를 한다. **CAPM에 비해 더 적고 현실적인 가정**을 통해 유도된 **APT는 좀 더 일반화된 모델로써 시장에 적용**될 수 있다.

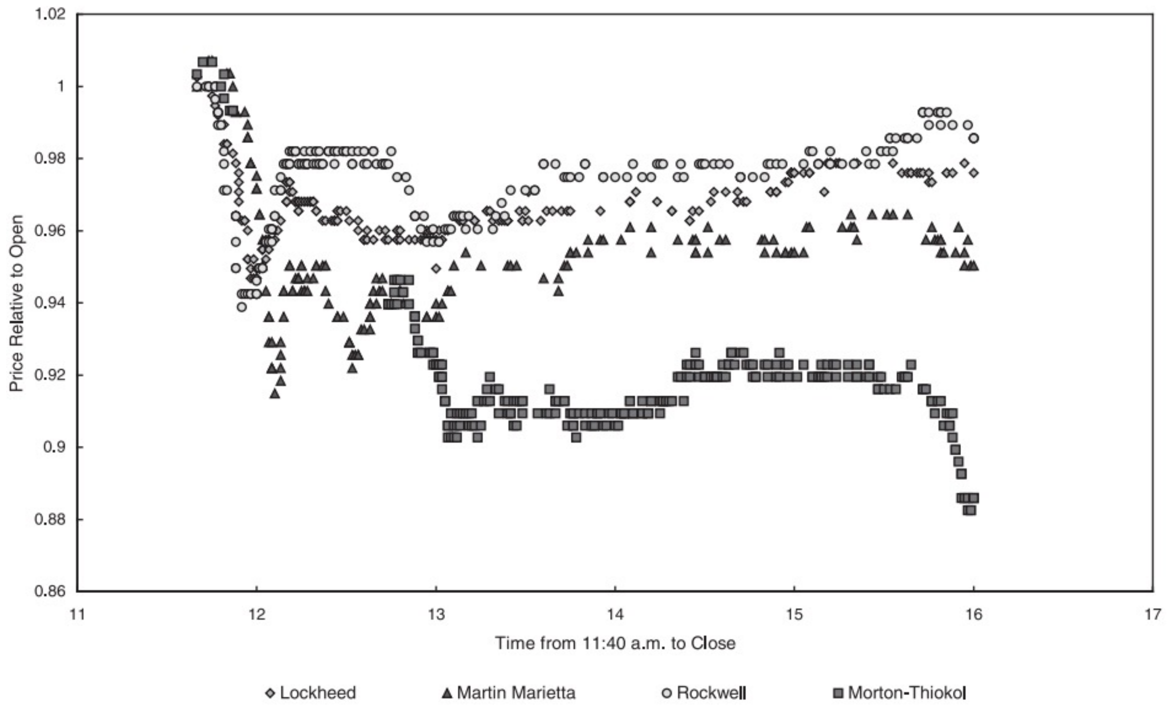
2. 효율적 시장 가설, 알파 그리고 팩터 모델

36기 배동익

효율적 시장 가설, 알파 그리고 팩터 모델

: 효율적 시장 가설

M.T. Maloney, J.H. Mulherin / Journal of Corporate Finance 9 (2003) 453–479



January 28, 1986

11:39 a.m.: Shuttle explodes

11:47 a.m.: Dow Jones News Wire: "Space Shuttle Explodes"

12:17 p.m.: Dow Jones News Wire: "Lockheed Has No Immediate Comment"

12:52 p.m.: Dow Jones News Wire: "Rockwell Intl Has No Comment"

효율적 시장 가설, 알파 그리고 팩터 모델
: 효율적 시장 가설

Eugene Fama

" 시장 가격이 **주어진 모든 정보를 fully reflect**하고 있을 때 시장은 효율적이다."

웅성웅성:

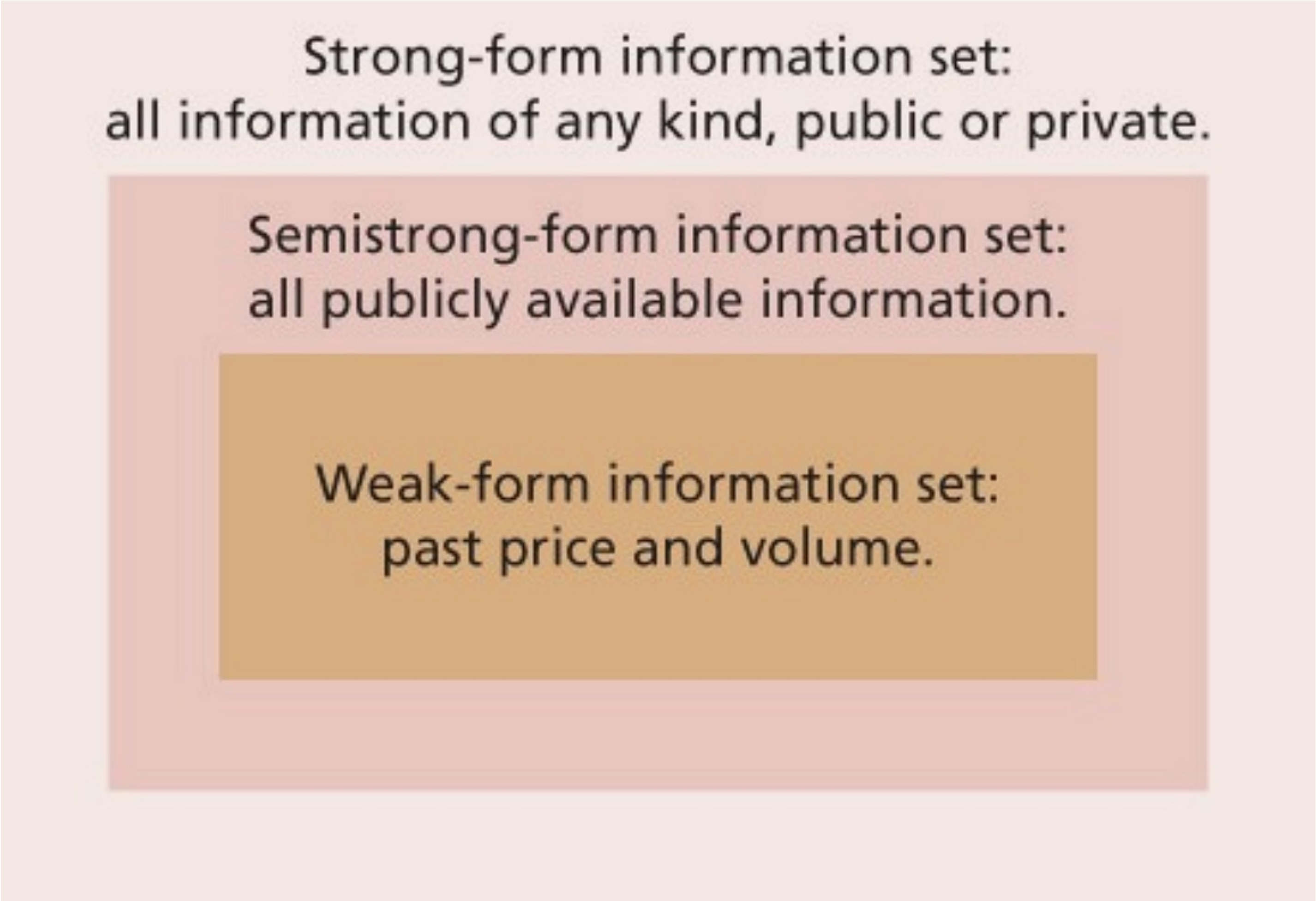
Fully reflect의 기준이 뭔데

Malkiel

"정보를 모든 시장 참여자들에게 공개하였을 때 **가격에 영향을 미치지 못할 정도**"

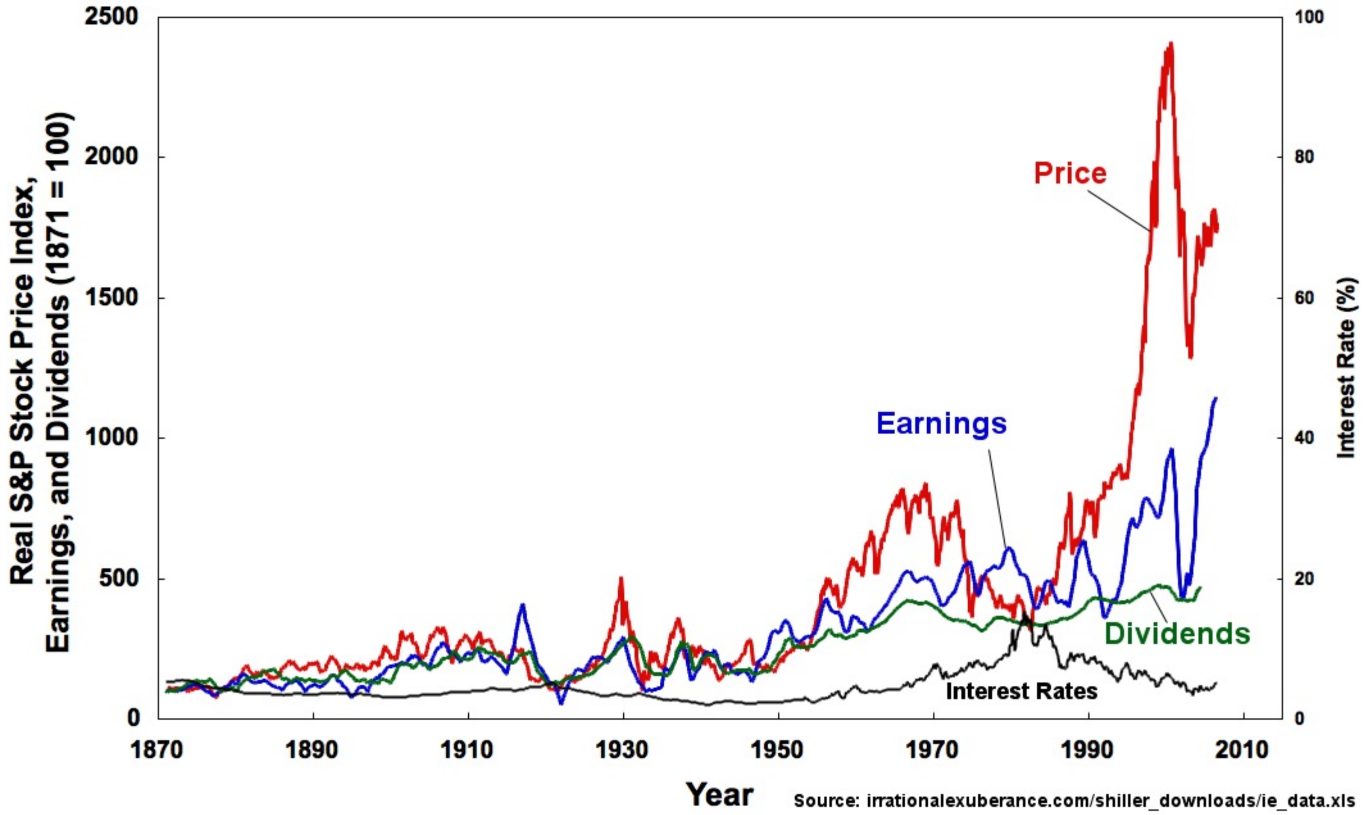
효율적 시장 가설, 알파 그리고 팩터 모델

: 효율적 시장 가설의 형태



효율적 시장 가설, 알파 그리고 팩터 모델

: 효율적 시장 가설에 대한 비판



효율적 시장 가설, 알파 그리고 팩터 모델

: 효율적 시장 가설에 대한 비판

- **Excess Predictability**

  실제로 투자자들이 fundamental 정보를 활용하여 투자를 하는 것이 맞아?

  No, **투자자들은 때때로 이상한 이유로 투자**를 한다 ex) GameStop

  따라서 **실제 주가의 등락폭은 earning과 dividend의 변화보다 큰 폭**으로 움직인다

- **Limits to arbitrage**

  공매도가 생각보다 risky하고 expensive, short squeeze

- **Systematic Noise Trading**

  Arbitrage is risky + borrowing is limited + investors focus on short-term performance

효율적 시장 가설, 알파 그리고 팩터 모델

: 효율적 시장 가설에 대한 비판

- confirmation bias,
- extrapolation bias,
- trend following ,
- anchoring bias ,
- overconfidence ,
- mental accounting ,
- availability cascades,
- loss aversion ,
- overreaction to news ,
- Other examples: Price Momentum & reversal, etc.

이러한 **행동 bias가 존재**하기 때문에

시장참여자들은 이성적이지 않다

따라서 **비이성적 상황을 가정한 모델**을 통해

주식 가격을 설명하는 것이 타당하다

효율적 시장 가설, 알파 그리고 팩터 모델

: EMH(Efficient Market Hypothesis) vs 행동재무학

EMH

- 가격에 정보가 다 반영되어 있다.
- 투자자들은 평균적으로 Rational 하다.
- 액티브한 펀드매니저도 **알파를 얻을 수 없다.**

행동재무학

- Market anomalies가 존재한다.
- 비이성적 참여자가 존재한다.
- **알파를 얻을 수 있다.**

효율적 시장 가설, 알파 그리고 팩터 모델

: 알파란 무엇인가?

- **젠센**

$$r_i - r_f = 0.00359 + 0.0108\beta_i \quad (R^2 = 0.98)$$

$$(t = 6.52) \qquad (t = 20.77)$$

- **World Quant**

데이터를 기반으로 주가를 예측할 수 있는 수학적 모델

두 알파는 서로 다른 것일까?

효율적 시장 가설, 알파 그리고 팩터 모델

: 팩터 모델

- **APT**

$$r_i - r_f = \alpha_i + \beta_1 * RP_1 + \beta_2 * RP_2 + \cdots + \beta_n * RP_n$$

- **FF 3요인 모형**

$$r_i - r_f = \alpha_i + \beta_{mkt} * (r_m - r_f) + \beta_s * SMB + \beta_v * HML$$

효율적 시장 가설, 알파 그리고 팩터 모델

: 알파의 분해

- 무위험 초과 성과
- 리스크 프리미엄 수취
- 스마트 베타

효율적 시장 가설, 알파 그리고 팩터 모델
: 적응적 시장 가설, AMH

Andre W. Lo:

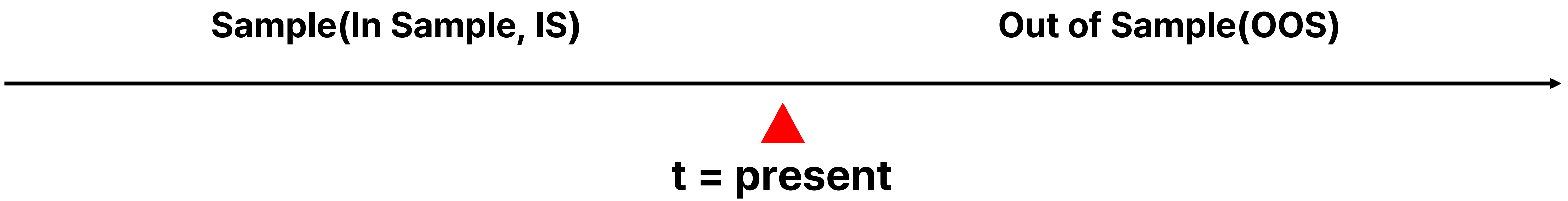
“시장은 계속해서 환경에 따라 변화를 거듭해나가기 때문에 **항상 효율적이거나 항상 비효율적이라고 할 수 없다.**”

따라서, 팩터의 성과는 상황가변적이다.

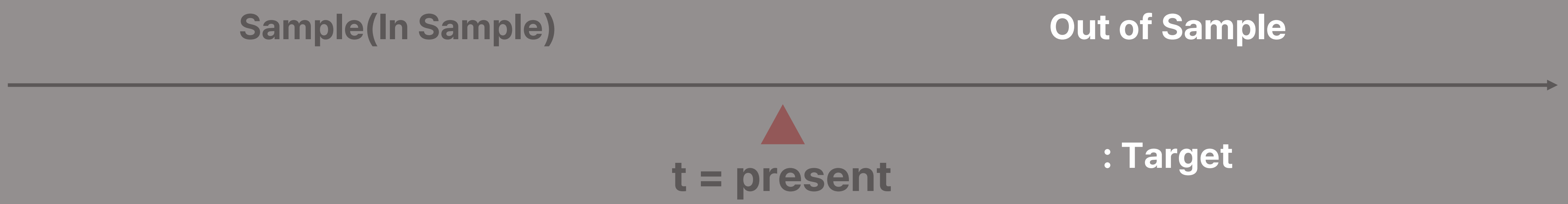
3. Backtest Overfitting

36기 이영우

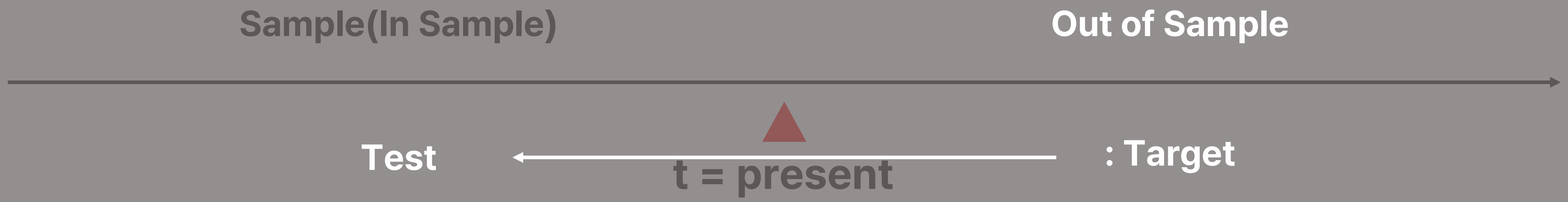
Backtest Overfitting



Backtest Overfitting



Backtest Overfitting



Backtest Overfitting

Backtest Overfitting

Backtesting

A historical simulation of how an algorithmic strategy would have **performed in the past.**
(David. H., 2015)

그러나, Backtesting은 과적합 되었을 가능성이 존재함

Backtest Overfitting

과적합(Overfitting)

The Situation when a **model targets particular observations** rather than a general structure(David. H., 2014)

We say that a strategy selection process **overfits** if the expected **performance of the Strategies selected IS is less than the median performance rank OOS of all strategies** (David. H., 2015)

Backtest Overfitting
: 과적합을 측정하는 방법

PBO
: Probability of Backtest Overfitting(David. H., 2015)

: 최적 In-sample로 선택된 모델 구성이 N개 모델 구성의 중앙값 OOS보다 성능이 낮을 확률

Backtest Overfitting
: 과적합을 측정하는 방법

PBO
: Probability of Backtest Overfitting(David. H., 2015)

: IS에서 좋았다면, OS에서도 좋아야지.

Backtest Overfitting
: 과적합을 측정하는 방법

CSCV
: combinatorically symmetric cross-validation

Backtest Overfitting

: 과적합을 측정하는 방법

CSCV

: combinatorically symmetric cross-validation

Backtest Overfitting
: 과적합을 측정하는 방법; Example

Number of Trial

	M_11	M_12	...	...	...	M_1N
	M_21	...			...	...
	...		...	...		
Time						
	...		...	...		
	...	...			...	
	M_T1	...	...	...	...	M_TN

Strategy: S_a, M_ij: S_a's performance at Trial i, time j(ex; sharpe ratio)

Backtest Overfitting

: 과적합을 측정하는 방법; Example

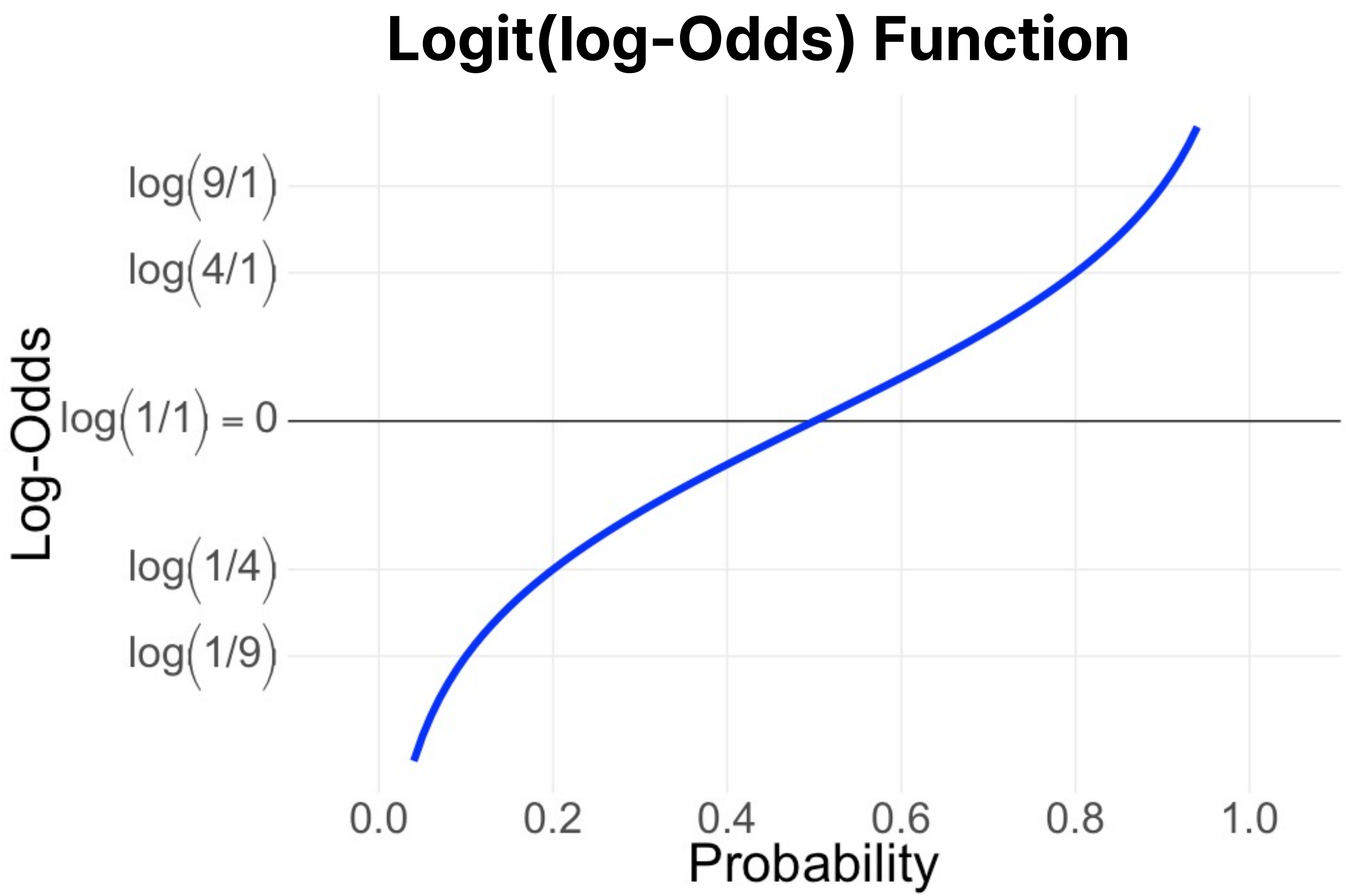
Number of Trial

	M_11	M_12	…	…	…	M_1N
	M_21	…			…	…
	…		…	…		
Time						
	…		…	…		
	…	…			…	
	M_T1	…	…	…	…	M_TN

Strategy: S_a, M_ij: S_a's performance at Trial i, time j(ex; sharpe ratio)

Backtest Overfitting

: 과적합을 측정하는 방법; Example



Backtest Overfitting

: 과적합을 측정하는 방법; Example

PBO = Sum{(logit<0.5; 전략의 상대적 순위 Q2 이하) / (전체 시도 횟수)}

Backtest Overfitting
: 과적합을 예방하는 방법

1. **표본 외(OOS) 검증**

   특정 투자 상황에서 성과가 좋은 전략은 다른 상황에서도 성과가 좋다.

2. **표본 검증에서 샤프 비율 요구사항 증가**

   표본 검증에서의 높은 샤프 비율은 표본 외 상황에서 샤프 비율이 높을 확률을 높인다.

3. **교차 검증(Cross-Validation)**

   지역 간 교차 검증(미국-중국), 표본 내 교차 검증(CSCV)

4. **Min BLT 적용**

   과적합이 발생하지 않는 Backtest 구간의 길이, Backtest 시행 횟수에 따른 Backtest 구간의 길이 증가(David. H., 2014)

Backtest Overfitting

: Backtesting의 적절한 사용

Backtesting is **not a research tool**. It provides us with very little insight into the reason why a particular strategy would have money.(Prado., 2018)

Backtesting은 투자 전략의 유효성을 검증하는 도구이지, 개발 수단이 되어서는 안 된다.

4. Robustness

36기 이재형

Robustness

: 강건성이 필요한 이유

- 알파 연구의 목적은 시장을 예측하고 그것을 넘어서는 것
- 수익률 + 낮은 위험노출 + 결과 예측력
- 강력한 알파?

  투자 유니버스가 바뀌어도 그대로인 불변성

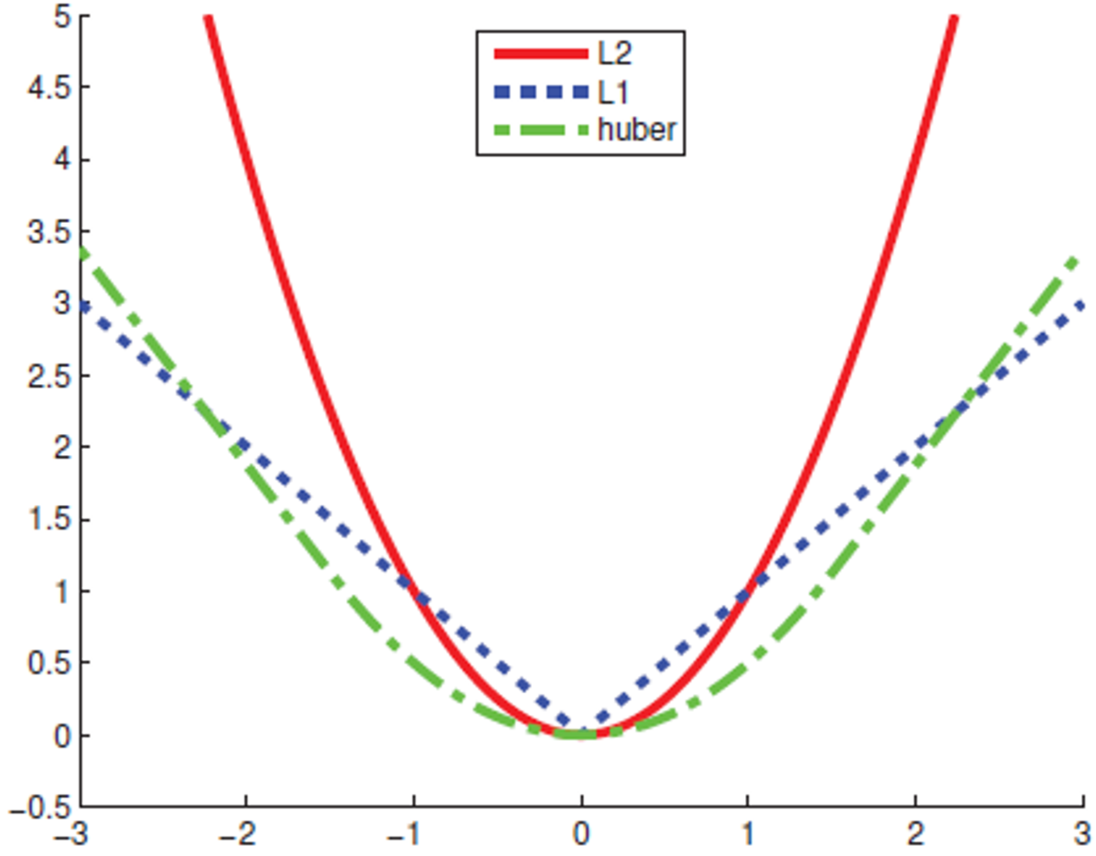
  극단 값에도 견고성

$$L_H(r,\delta) = \begin{cases} r^2/2 & \text{if } |r| \le \delta \\ \delta|r| - \delta^2/2 & \text{if } |r| > \delta \end{cases}$$

https://colab.research.google.com/drive/1Q_6dVCbD14AA3fWUK-0hN1lewRiG1cUU#scrollTo=wYfVM-H6PJq2

Robustness

: Huber Regression



$$L_H(r,\delta) = \begin{cases} r^2/2 & \text{if } |r| \le \delta \\ \delta|r| - \delta^2/2 & \text{if } |r| > \delta \end{cases}$$

https://colab.research.google.com/drive/1Q_6dVCbD14AA3fWUK-0hN1lewRiG1cUU#scrollTo=wYfVM-H6PJq2

Robustness

: Long-Short Portfolio

Stocks in the Alpha vector	Alpha value on 20-Mar	Subtract Average	Centred around 0	Absolute value	Normalized weights on 20-Mar	Absolute value	Allocate $ 20 Mn capital	Suppose Returns on 21-Mar are	$Mn profits on 21-Mar would be
Stock 1 (F)	0.00	(0.50)	-0.5	0.50	-0.22	0.22	-4.4	2%	-0.09
Stock 2 (C)	0.14	(0.50)	-0.4	0.36	-0.16	0.16	-3.1	2%	-0.06
Stock 3 (B)	0.29	(0.50)	-0.2	0.21	-0.09	0.09	-1.9	-2%	0.04
Stock 4 (A)	0.43	(0.50)	-0.1	0.07	-0.03	0.03	-0.6	-2%	0.01
Stock 5 (G)	0.57	(0.50)	0.1	0.07	0.03	0.03	0.6	1%	0.01
Stock 6 (D)	0.71	(0.50)	0.2	0.21	0.09	0.09	1.9	0%	0.00
Stock 7 (H)	0.86	(0.50)	0.4	0.36	0.16	0.16	3.1	1%	0.03
Stock 8 (E)	1.00	(0.50)	0.5	0.50	0.22	0.22	4.4	2%	0.09
Sum			- 1.1 + 1.1	**2.3**	-0.5 + 0.5	**1.0**	+10 M -10 M		**0.03**
Mean	0.50		0.0		0.0	Normalized	0.0		P&L

Robustness
: 강건성을 개선하기 위한 방법

1. **순서 방법(ordering method)**

   랭킹, 분위수 근사치

2. **정규분포 근사치(approximation to normal distribution)**

   Z-score

3. **제한 방법(limiting method)**

   트리밍, 윈저화

Robustness

: 순서 방법(ordering method)

- 순위 간 차이의 절대적 값이 의미 없음
- 비모수(파라미터가 없는) 데이터에서 사용

[랭킹]: example

[0,1] 구간으로 재조정시

[0.2,0.6,.0.8,2.5]에서 10이 추가되면?

랭킹: [0,0.33,0.66,1] ->[0,0.25,0.5,0.75,1]

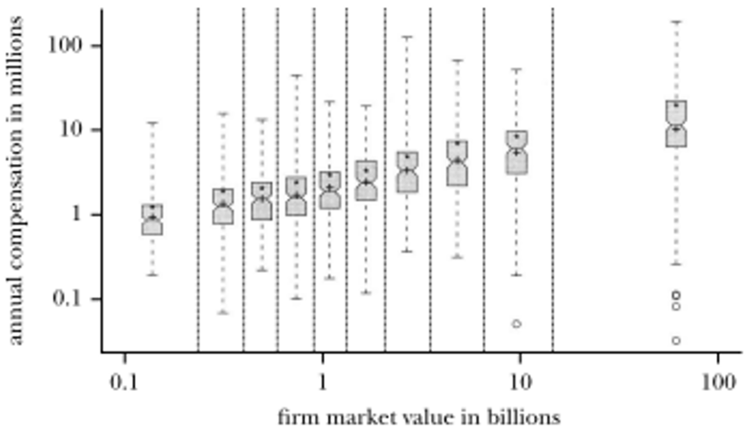
정규화(Min-max scaling): [0,0.17,0.26,1] -> [0,0.04,0.06,0.23,1]

표준화:[-0.94, -0.48, -0.26, 1.68] -> [-0.71,-0.60,-0.55,-0.09, 1.95]

Robustness
: 순서 방법(ordering method)

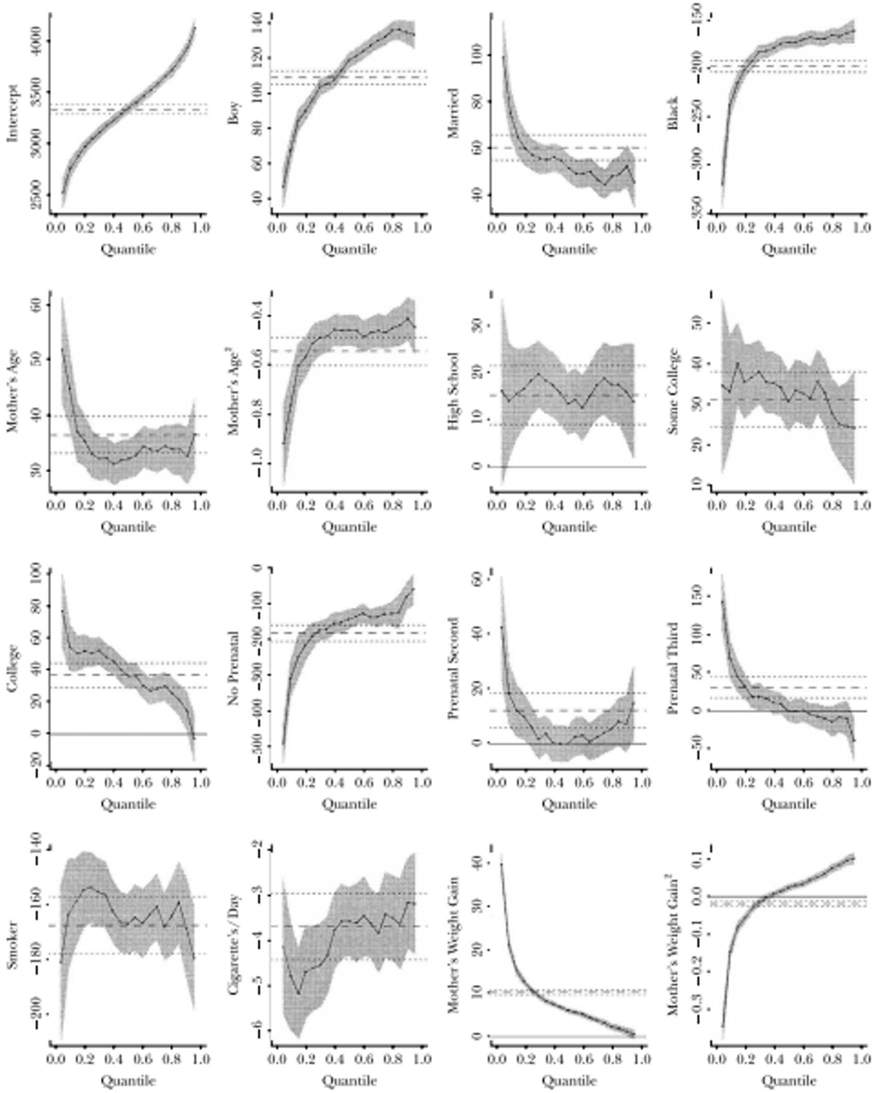
[분위수 근사치] : Least Quantile of Squares

Figure 1
Pay of Chief Executive Officers by Firm Size



Notes: The boxplots provide a summary of the distribution of CEO annual compensation for ten groupings of firms ranked by market capitalization. The light gray vertical lines demarcate the deciles of the firm size groupings. The upper and lower limits of the boxes represent the first and third quartiles of pay. The median for each group is represented by the horizontal bar in the middle of each box.
Source: Data on CEO annual compensation from EXECUCOMP in 1999.

Figure 4
Ordinary Least Squares and Quantile Regression Estimates for Birthweight Model



Robustness

: 순서 방법(ordering method)

[분위수 근사치] : Least Quantile of Squares

Linear Regression Model – Mean Squared Error

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \ldots + \beta_p x_{ip} \qquad MSE = \frac{1}{n}\sum_{i=1}^{n}(y_i - (\beta_0 + \beta_1 x_{i1} + \ldots + \beta_p x_{ip}))2$$

Quantile Regression Model Equation for the τth quantile - Median Absolute Deviation

$$Q_\tau(y_i) = \beta_0(\tau) + \beta_1(\tau)x_{i1} + +\beta_p(\tau)x_{ip} \qquad MAD = \frac{1}{n}\sum_{i=1}^{n}\rho_\tau(y_i(\beta_0(\tau) + \beta_1(\tau)x_{i1} + +\beta_p(\tau)x_{ip}))$$

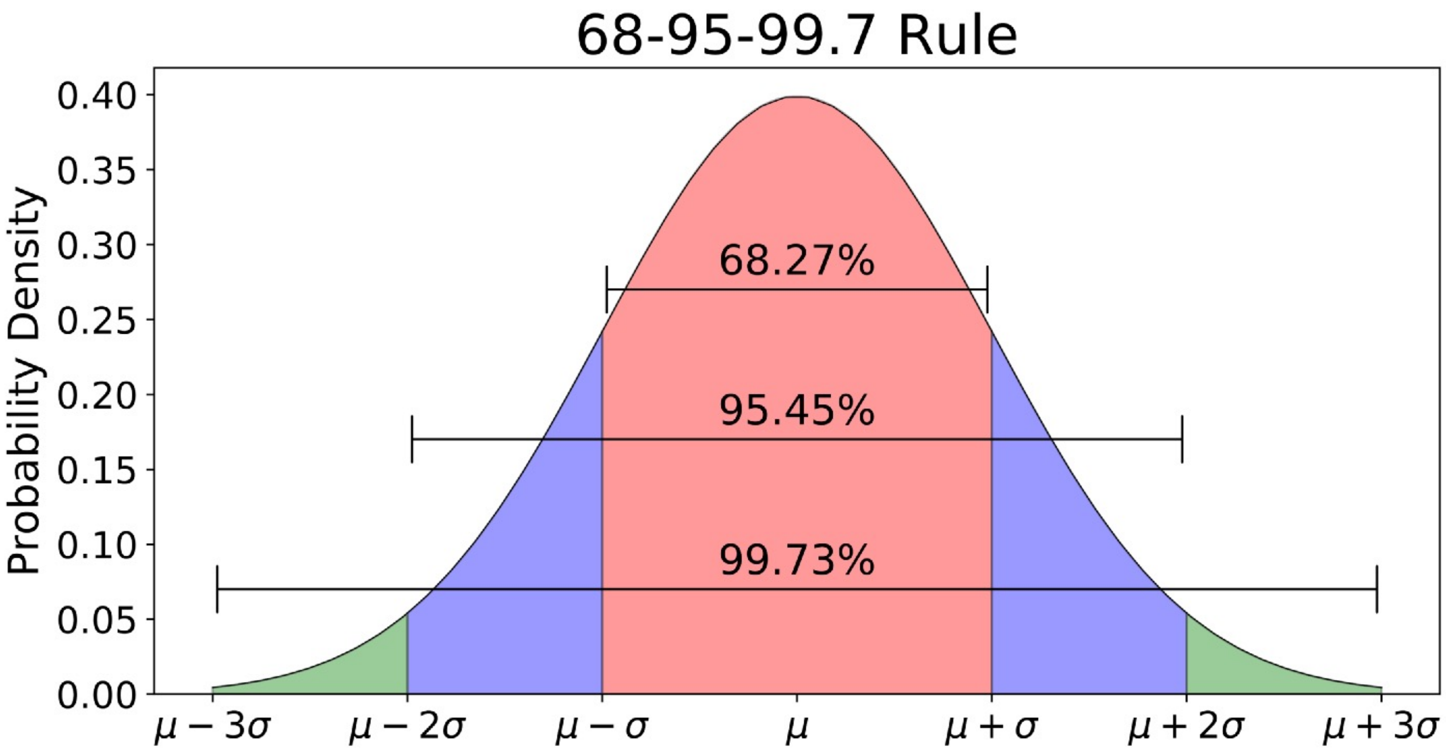
$$\rho_\tau(u) = \tau \max(u, 0) + (1 - \tau)\max(-u, 0)$$

Robustness
: 정규분포 근사치

[Z-score]

$$z = \frac{x - \mu}{\sigma}$$

$\mu =$ Mean
$\sigma =$ Standard Deviation



Robustness

: 제한방법(Limiting method)

[트리밍]: 임계치보다 높거나 낮으면 제거

[윈저화]: 극단값을 버리지 않고 컷오프 값으로 대체

[Example]

[3,5,7,10,100]

평균: 25, 중위수: 7

20% 트리밍 평균: 상하위 20% 제거 (5+7+10)/3=7.33

20% 윈저화 평균: 상하위 20%를 가장 가까운 값으로 대체

(5+5+7+10+10)/5=7.4

Robustness

: 시뮬레이션(Simulation)

브레인 플랫폼 기준으로 롱숏 전략 구축, **수익률은 정규분포**를 따르고 **100개 종목 1000일 동안 시뮬레이션**.

[Rank]: [0,1] 사이 구간으로 **순서에 따라 비율 할당** 롱숏 전략 위해 -0.5 해줌

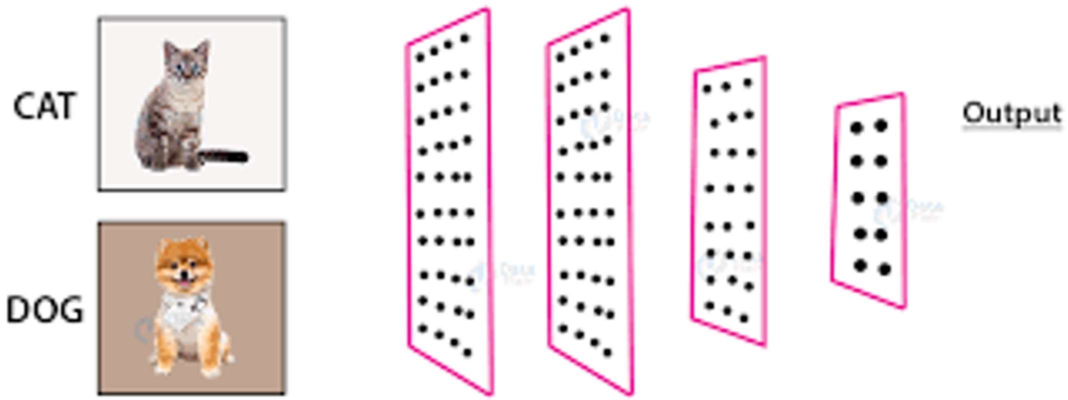
[Trim]: 정규분포 상 n*sigma **이상 값 제거**

[Winz]: 정규분포 상 n*sigma 이상 값 n*sigma 값으로 **대체**

생각해보기:
Alpha?

딥러닝의 시작: 강아지 고양이 분류(2만장 넘는 이미지)

이미 99%에 육박하는 정확성.

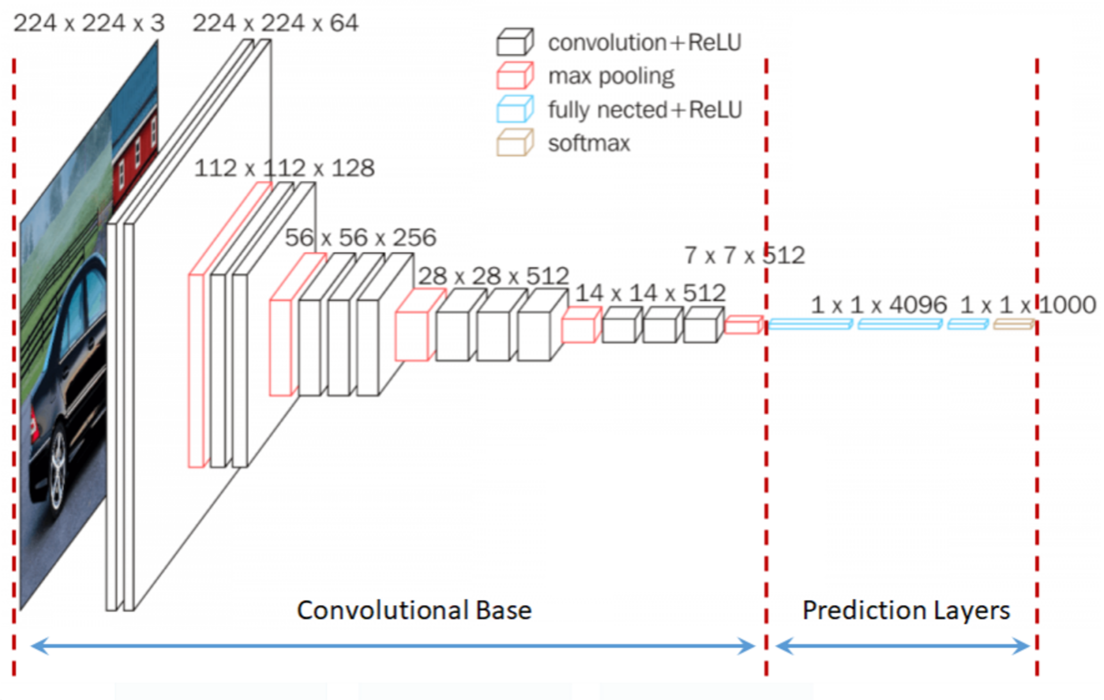


생각해보기:

Alpha?

BUY SIDE의 알파찾기(통계적 차익거래)는 근본이 없는가?

기준의 지평을 넓힌다고 생각



Any Questions?