20598 – Finance with Big Data

Week 4 Lecture: Empirical Asset Pricing
($\rightarrow$ Data-driven finance)

Clément Mazet-Sonilhac
clement.mazetsonilhac@unibocconi.it

Department of Finance, Bocconi University

Outline

Last week : CAPM & APT

FF3M

Making Factors ?

Factor Zoo

Feng et al., 2020

Outline

What do we learned from the CAPM / APT ?

What do we learned from the CAPM / APT ?

- W. Sharpe (1964) : the market portfolio is the tangency portfolio, i.e. the market portfolio is mean-variance efficient in the sense of Markowitz

What do we learned from the CAPM / APT ?

- W. Sharpe (1964) : the market portfolio is the tangency portfolio, i.e. the market portfolio is mean-variance efficient in the sense of Markowitz
- The CAPM model writes:

$$\tilde{r}_i - r_f = \beta_i(\tilde{\mathbf{r}}_{\mathbf{M}} - r_f)$$

  - The market return $\tilde{\mathbf{r}}_{\mathbf{M}}$ is the only risk factor
  - β is the right measure of risk
- Implication for investors?

What do we learned from the CAPM / APT ?

- W. Sharpe (1964) : the market portfolio is the tangency portfolio, i.e. the market portfolio is mean-variance efficient in the sense of Markowitz
- The CAPM model writes:

$$\tilde{r}_i - r_f = \beta_i(\tilde{\mathbf{r}}_{\mathbf{M}} - r_f)$$

  - The market return $\tilde{\mathbf{r}}_{\mathbf{M}}$ is the only risk factor
  - β is the right measure of risk
- Implication for investors? Buy stocks with high beta and short stocks with low-beta if you want to beat the market (taking a lot of risk)

What do we learned from the CAPM / APT ?

- The APT framework extends the CAPM :

$$\tilde{r}_i - r_f = \alpha_i + \beta_{i1}\,\mathbf{F_1} + ... + \beta_{in}\,\mathbf{F_n} + \epsilon_i$$

  - More flexible but no theory : what are the factors $\mathbf{F_1}, ..., \mathbf{F_n}$?

- Empirical research is mixed about the performance of those models

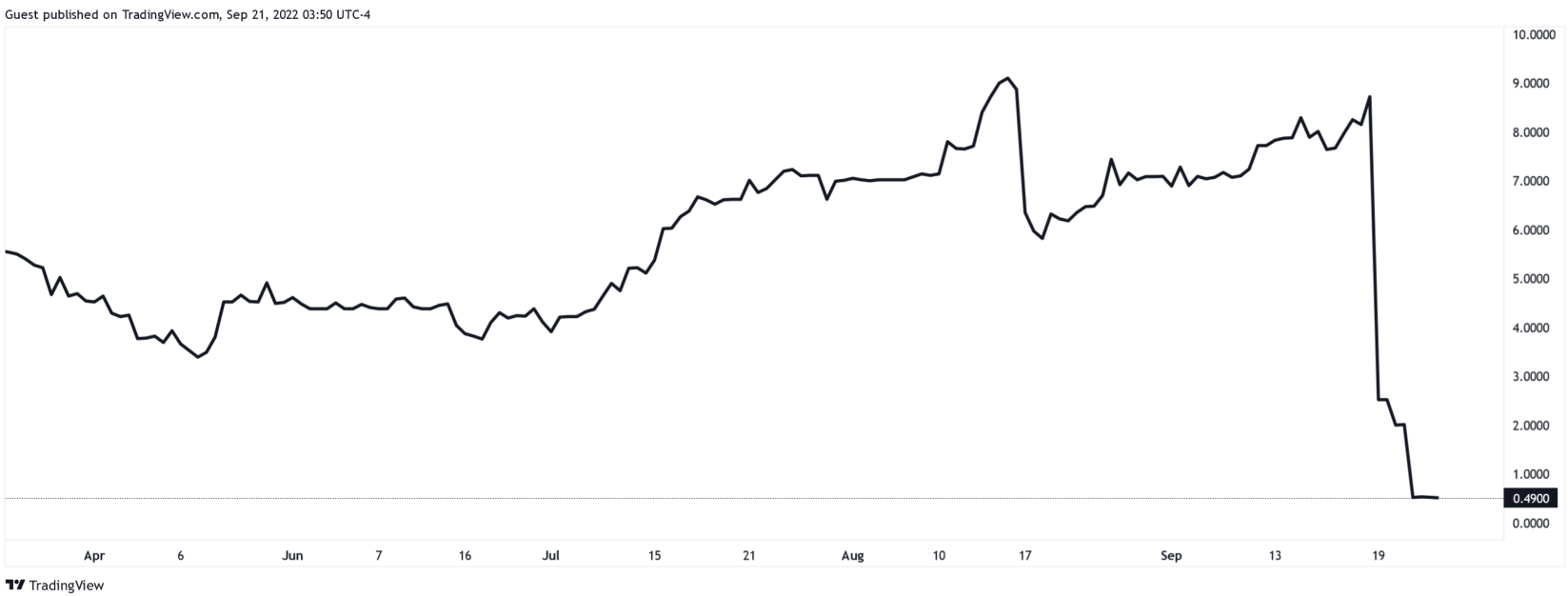
What do we learned from the CAPM / APT ?

- Virios Therapeutics : a development-stage biotechnology company focused on advancing novel antiviral therapies to treat debilitating chronic diseases

What do we learned from the CAPM / APT ?

- Virios Therapeutics : a development-stage biotechnology company focused on advancing novel antiviral therapies to treat debilitating chronic diseases

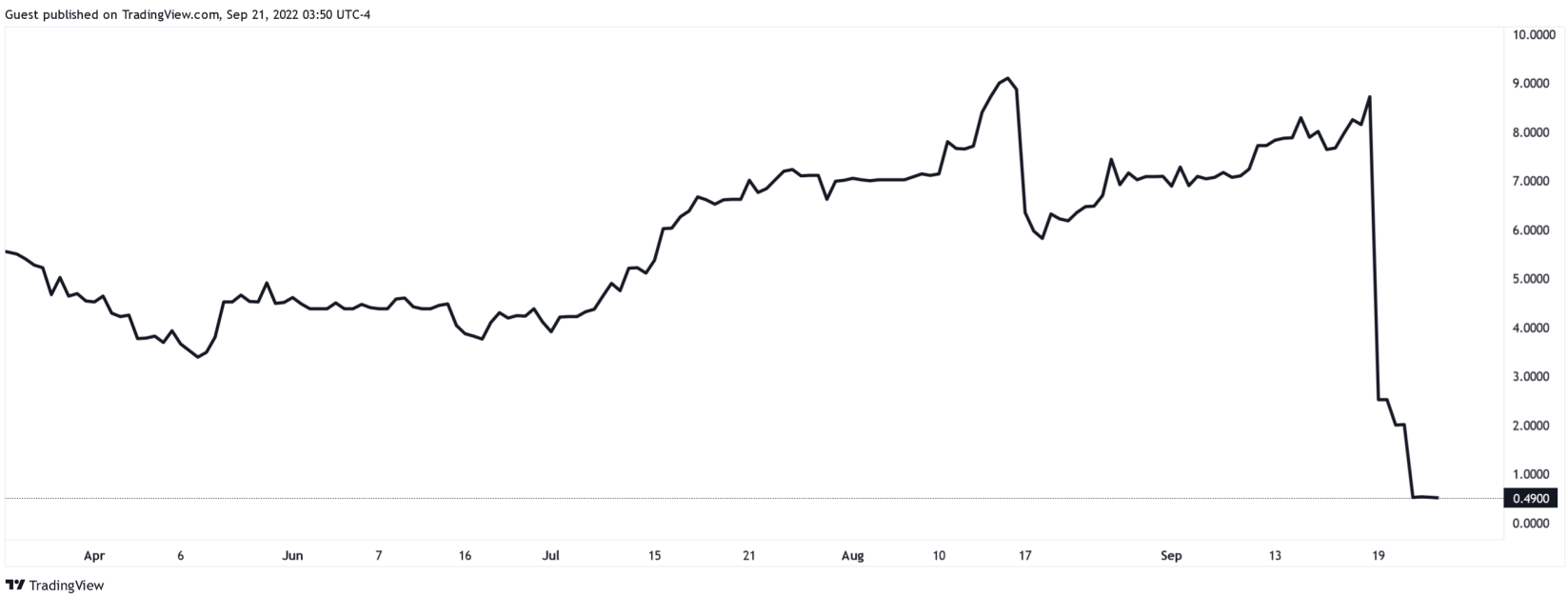
- Virios stock price slumped more than 77% exactly two years ago after the company said its new drug did not produce a significant improvement in symptoms for patients with fibromyalgia



What do we learned from the CAPM / APT ?

- Virios Therapeutics : a development-stage biotechnology company focused on advancing novel antiviral therapies to treat debilitating chronic diseases

- Virios stock price slumped more than 77% exactly two years ago after the company said its new drug did not produce a significant improvement in symptoms for patients with fibromyalgia



$\rightarrow$ Should I have this **bad** stock in my portfolio ?

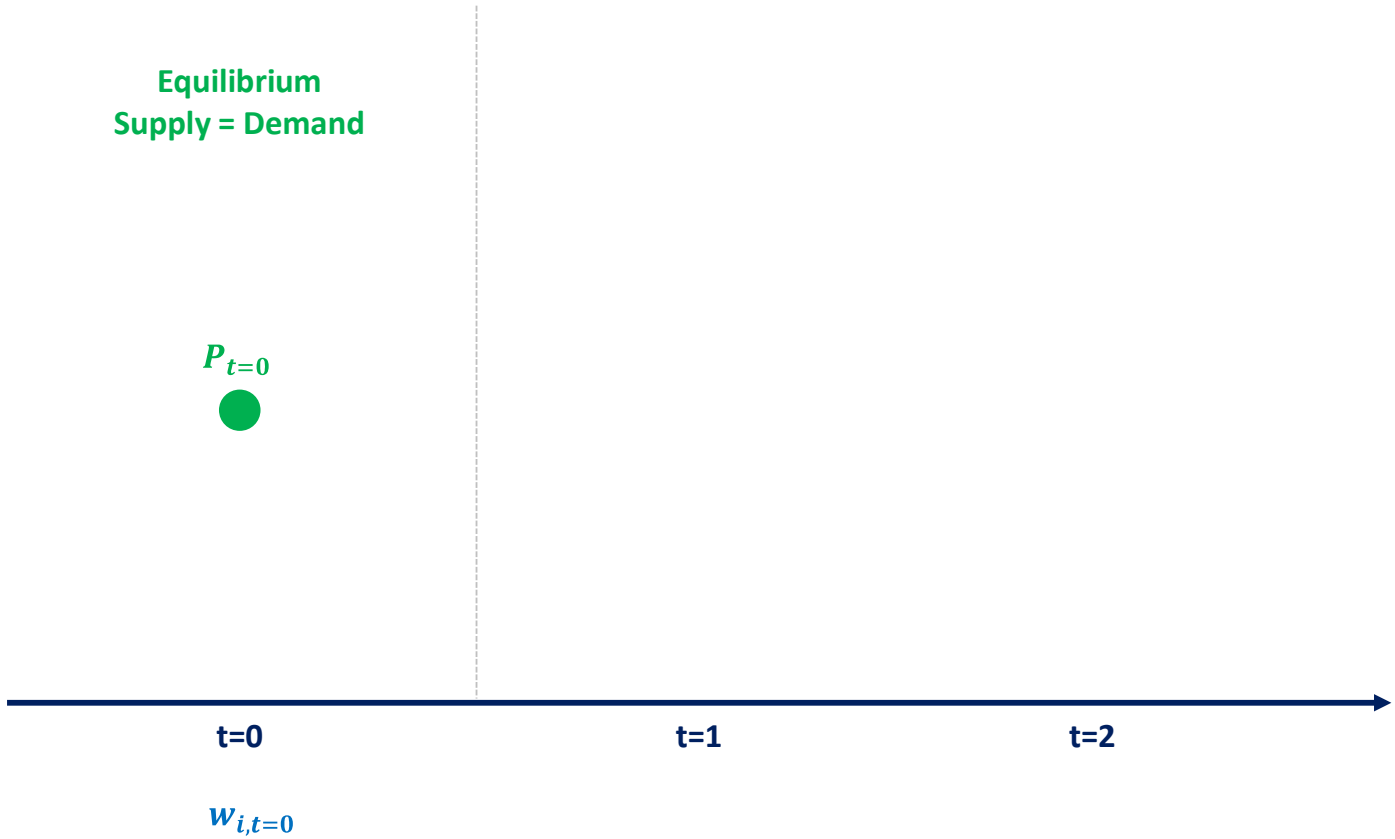
What do we learned from the CAPM / APT ?

Simple answer

- According to H. Markowitz and W. Sharpe, there is no **bad** stock
- Efficient markets: prices perfectly reflect the quality i.e., prices are quality-adjusted
- Theory tells you that the price drop was due to an idiosyncratic shock that can be diversified (= you don't care). The only thing that matters is the average return and the covariances
- After the shock is absorbed by the market, w_{virios} is really small but not equal to zero: you should still hold a tiny fraction of this stock in your portfolio.

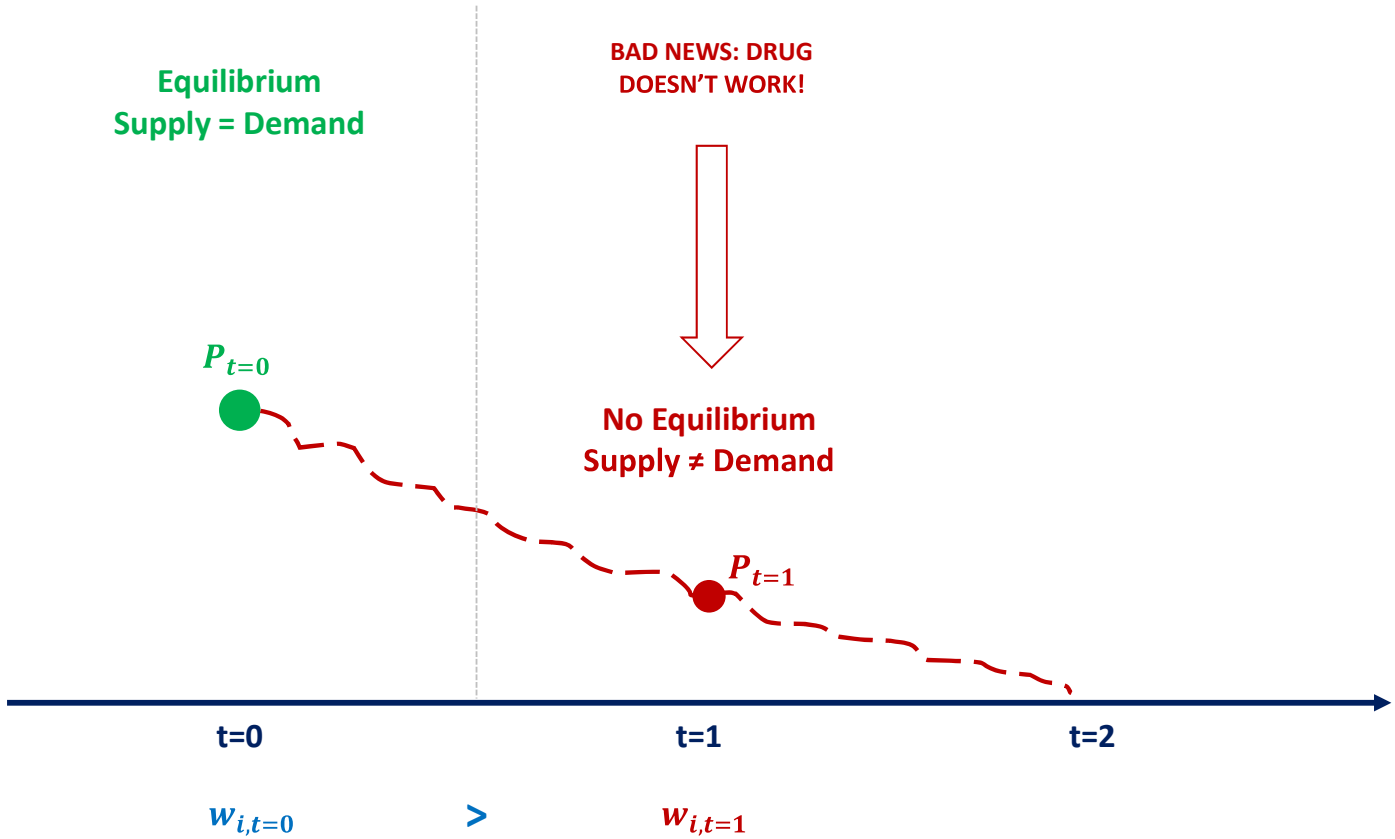
What do we learned from the CAPM / APT ?

More sophisticated answer



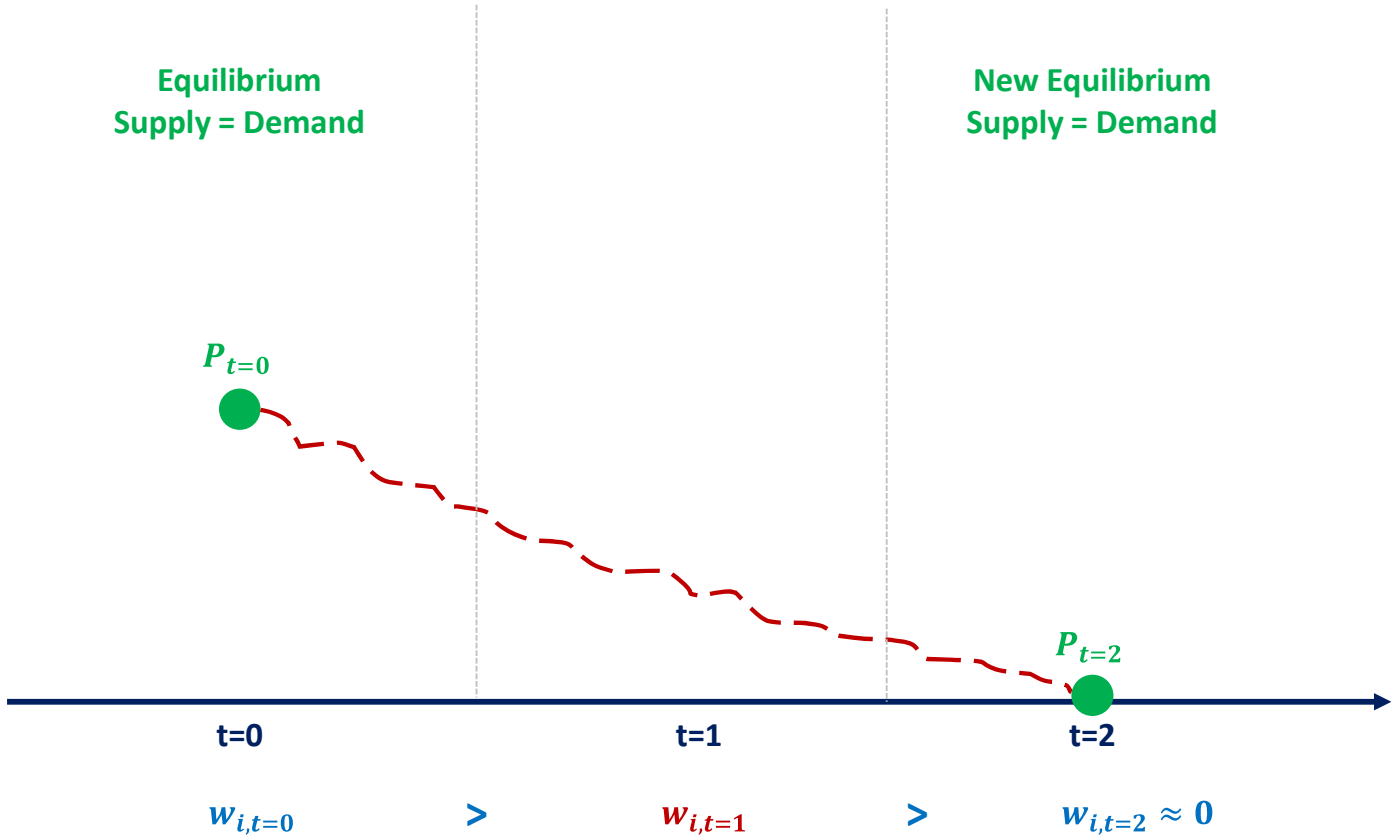
What do we learned from the CAPM / APT ?

More sophisticated answer



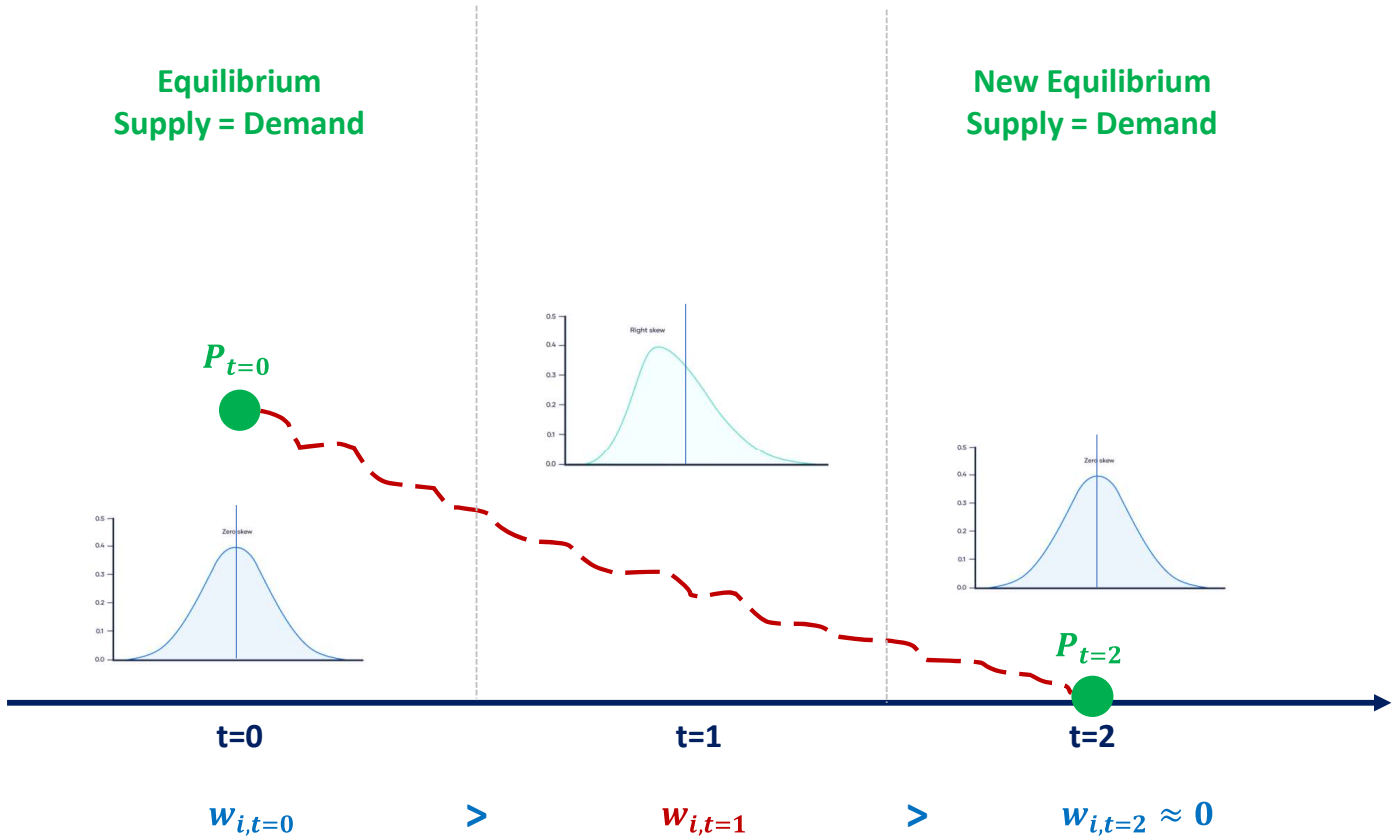
What do we learned from the CAPM / APT ?

More sophisticated answer



What do we learned from the CAPM / APT ?

More sophisticated answer



What do we learned from the CAPM / APT ?

The rationale of $\alpha_i = 0$

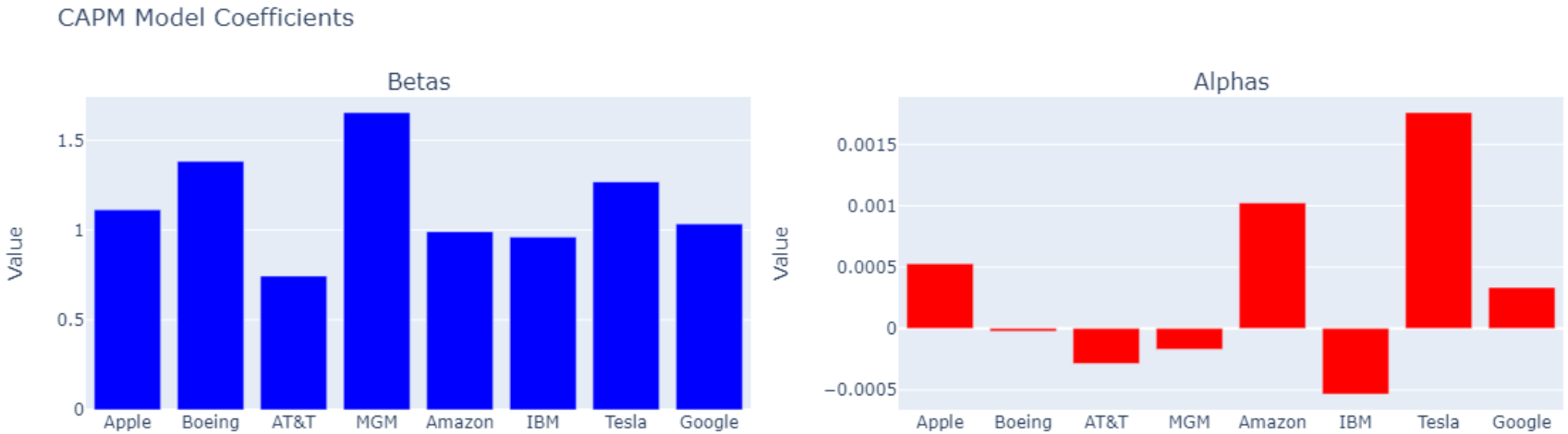
- Bringing the CAPM theory to the data gives:

$$\tilde{r}_i - r_f = \alpha_i + \beta_i(\tilde{r}_M - r_f) + \epsilon_i$$

- According to the CAPM, alpha should be 0 for all assets
- Alpha measures an asset's return in excess of its risk-adjust award

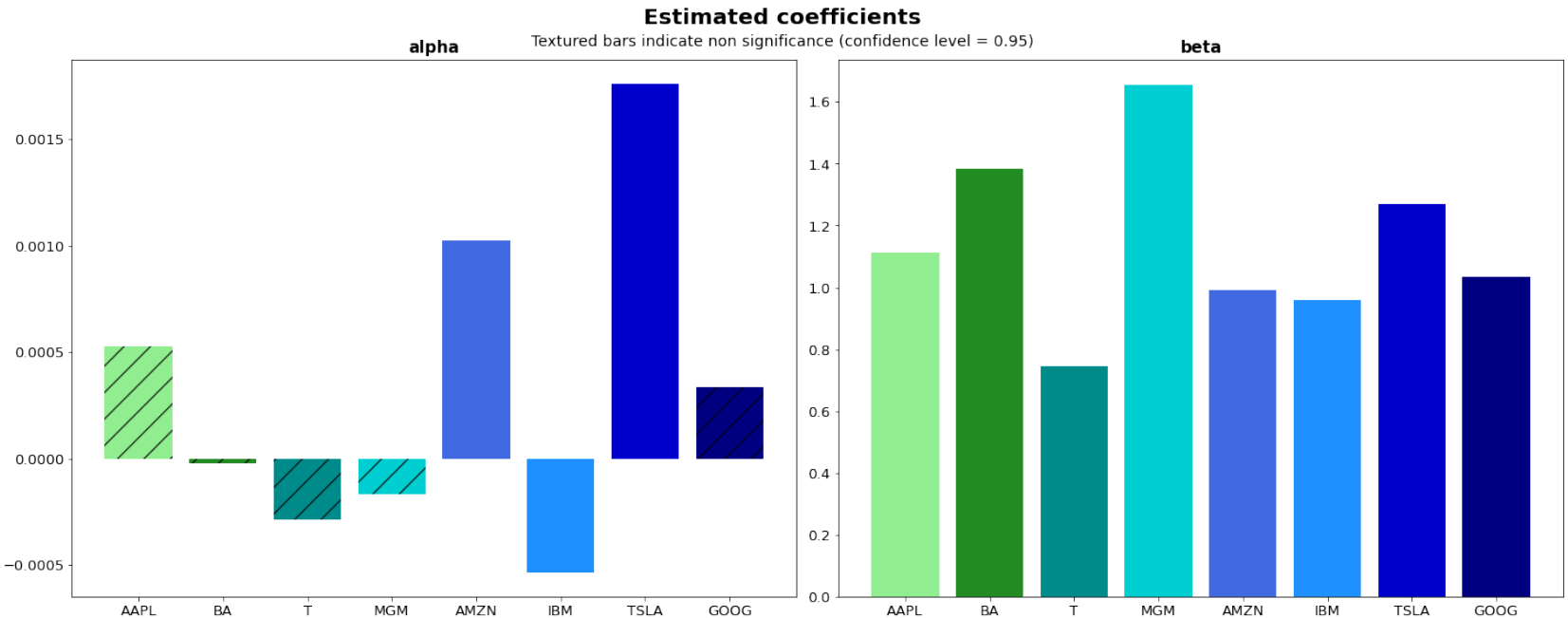
What do we learned from applying the CAPM ?

α and β – Group 5



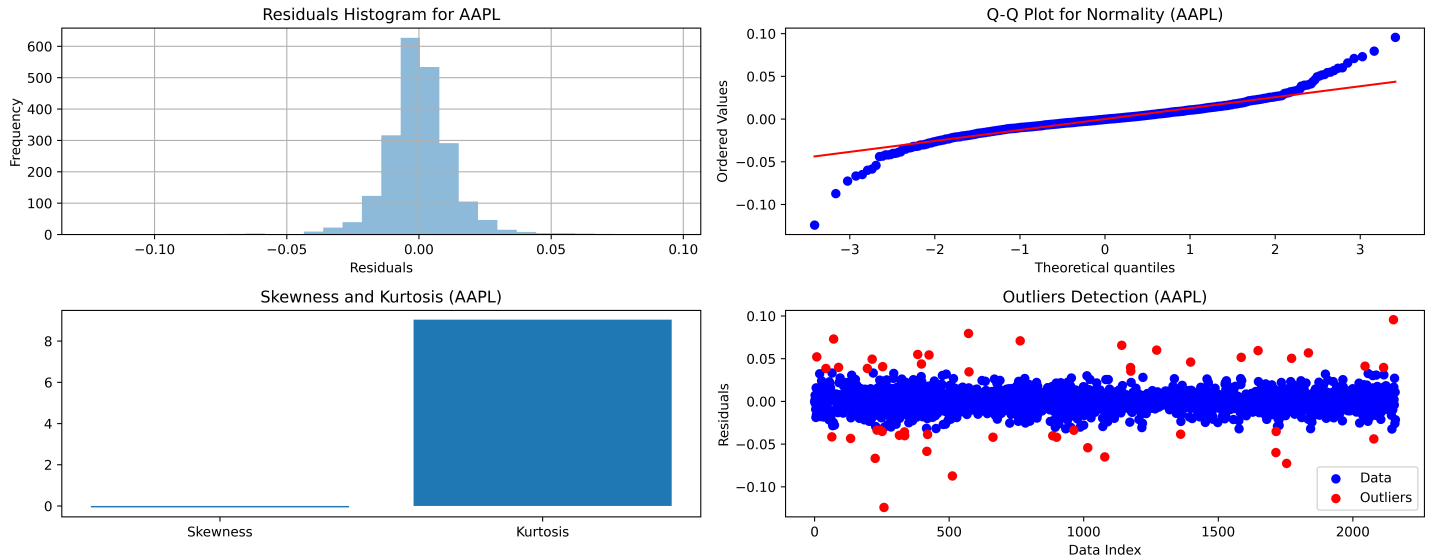
What do we learned from applying the CAPM ?

α and β – Last year : more information



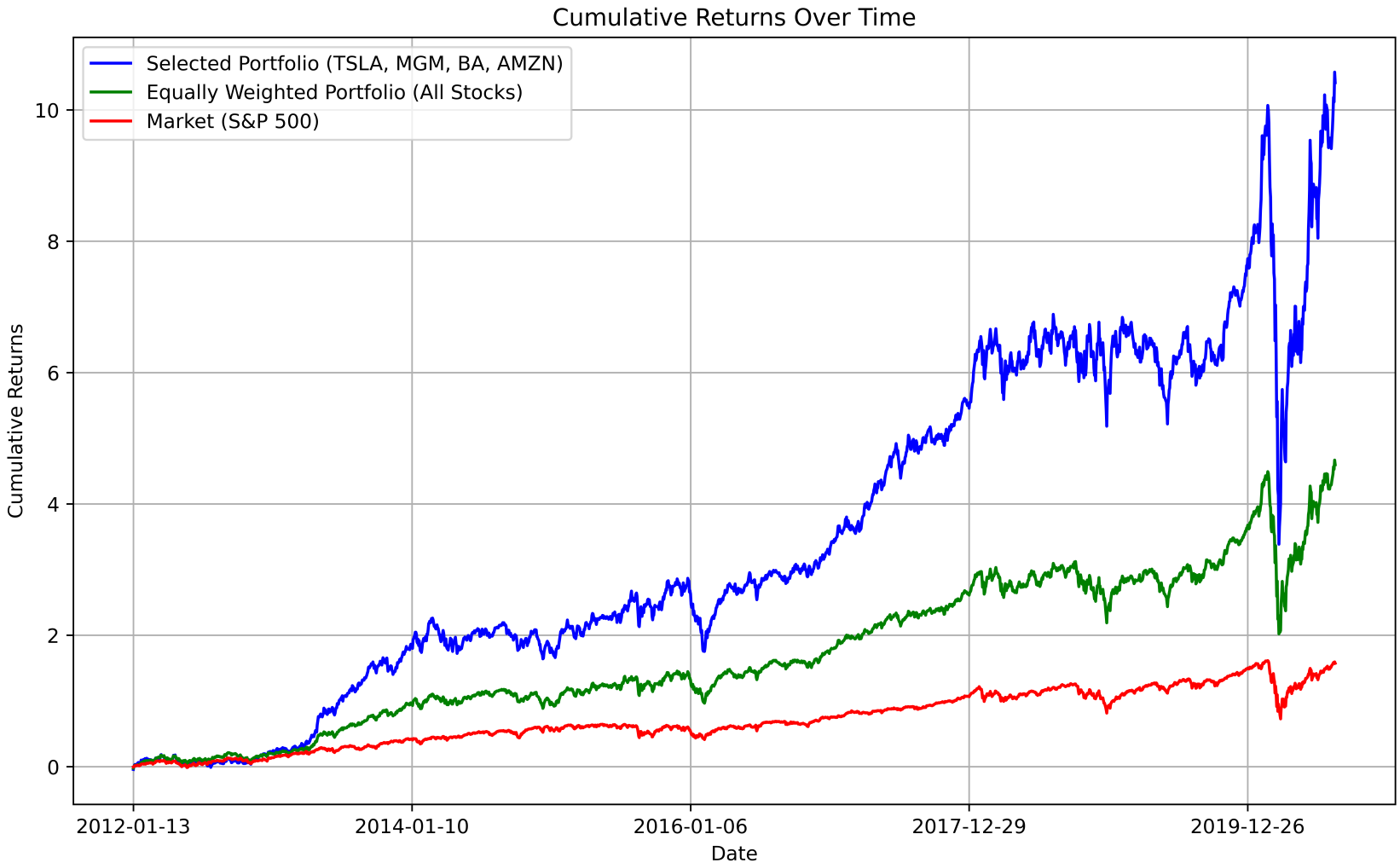
What do we learned from applying the CAPM ?

ϵ



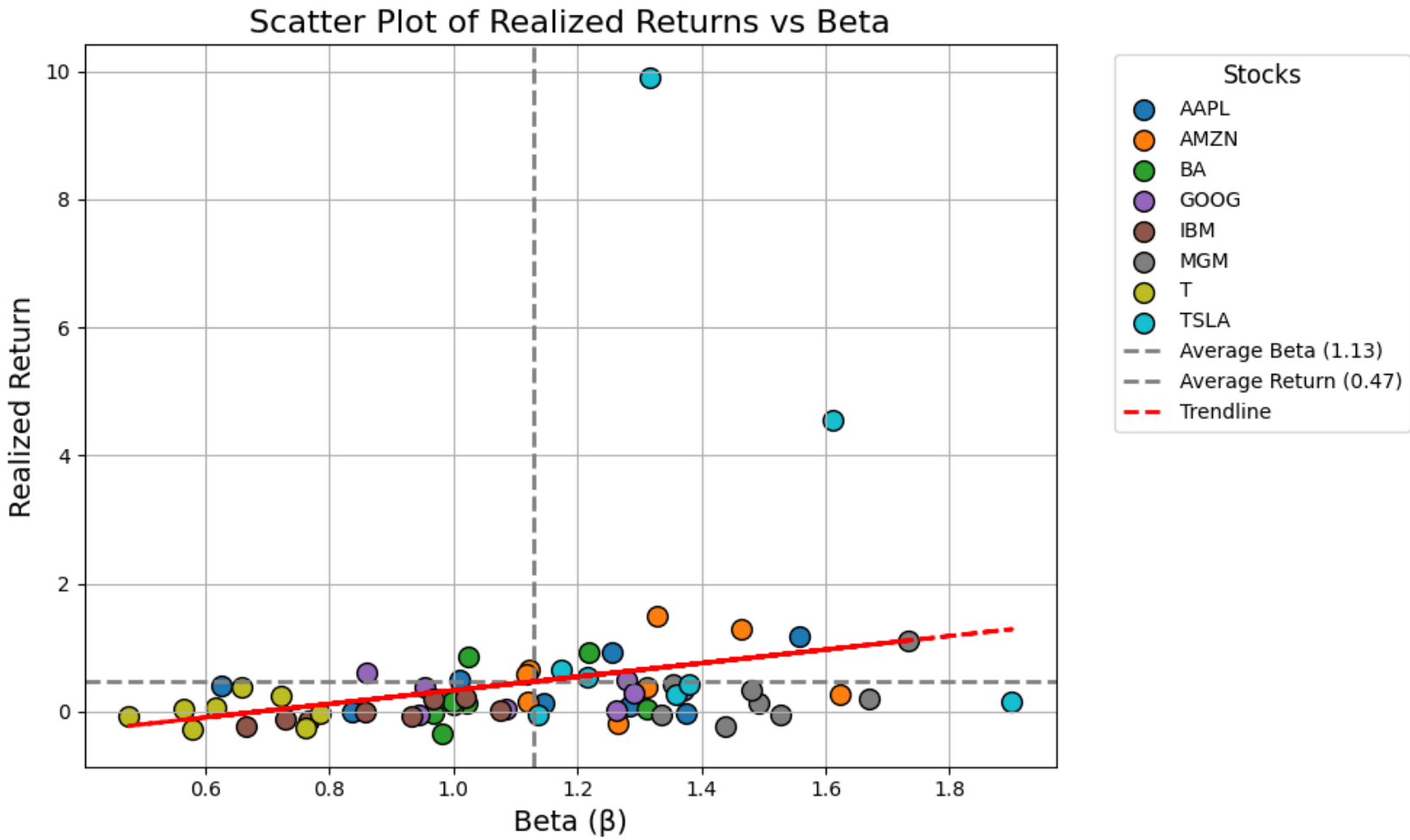
What do we learned from applying the CAPM ?

CAPM Performance



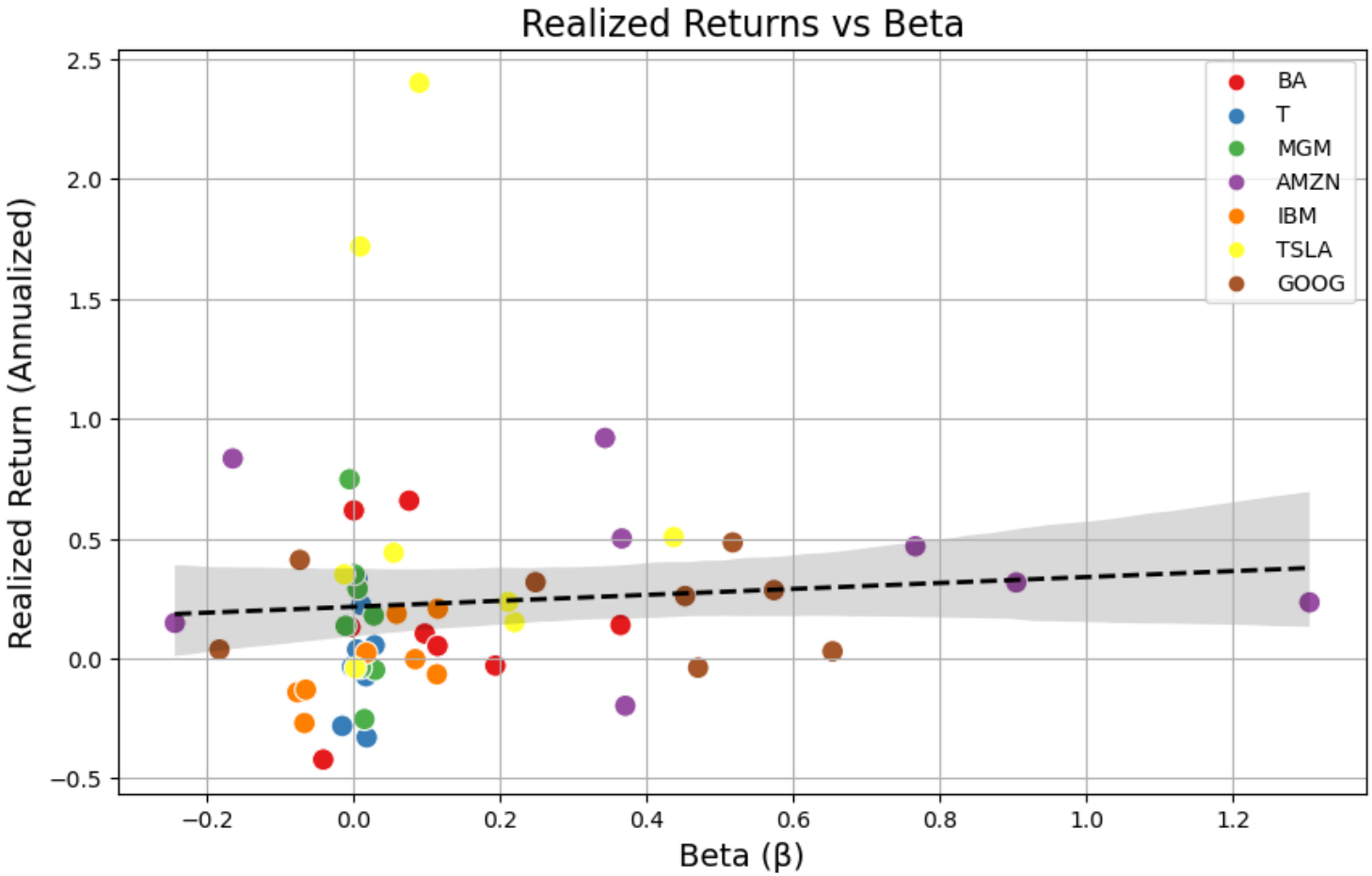
What do we learned from applying the CAPM ?

CAPM Performance – Group 8



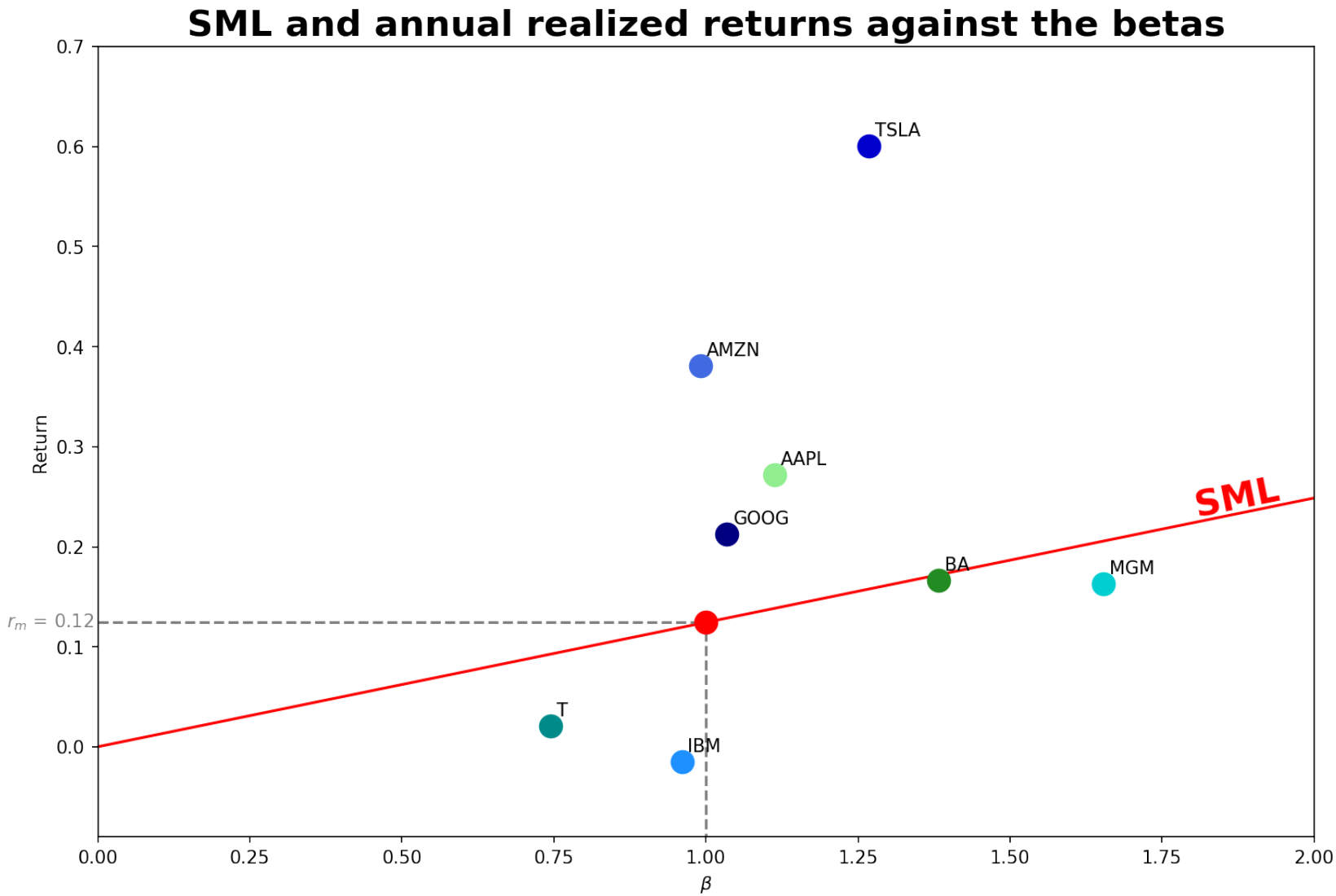
What do we learned from applying the CAPM ?

CAPM Performance – Group 6



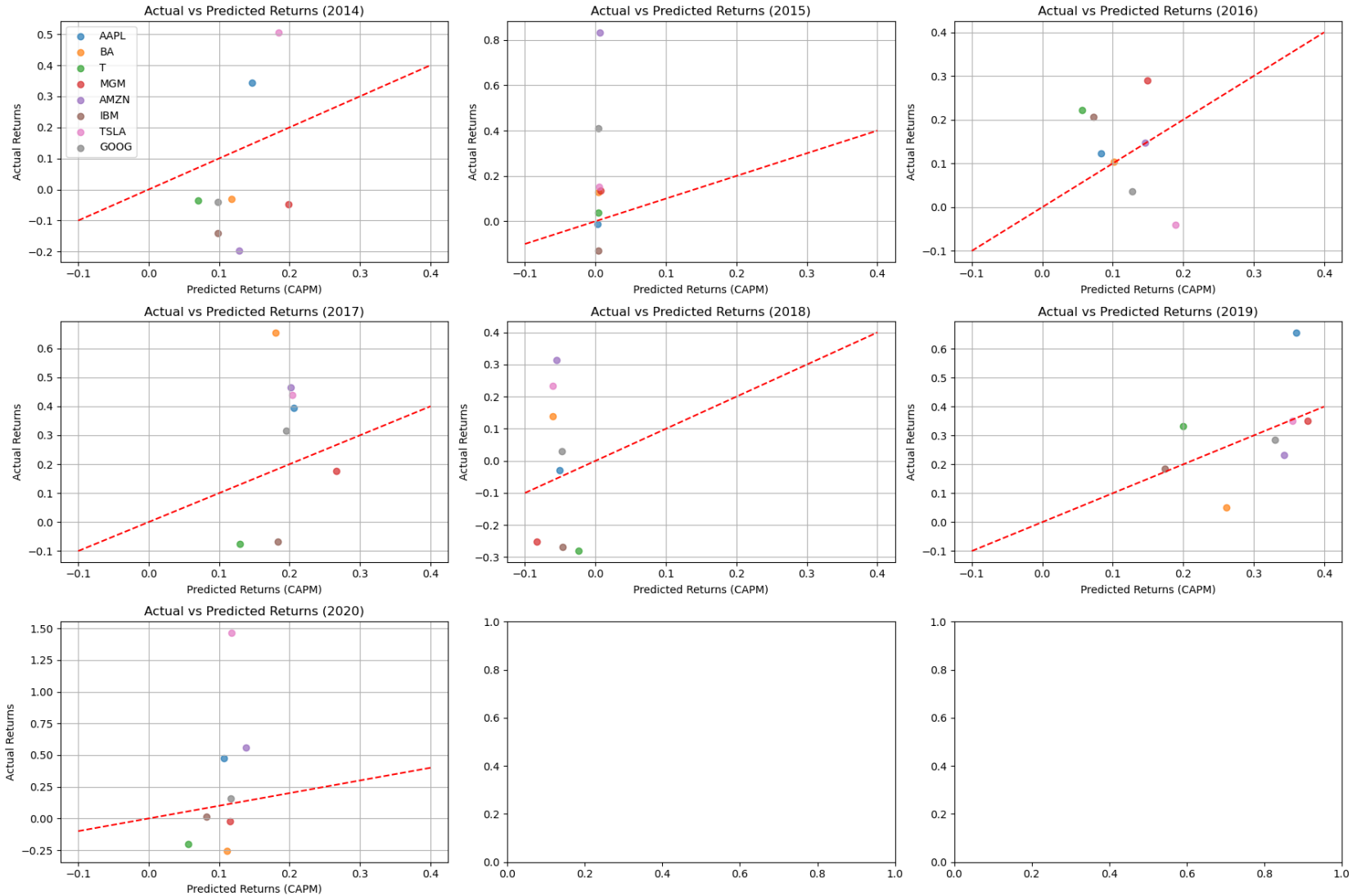
What do we learned from applying the CAPM ?

CAPM Performance - Last year : Beauty Prize



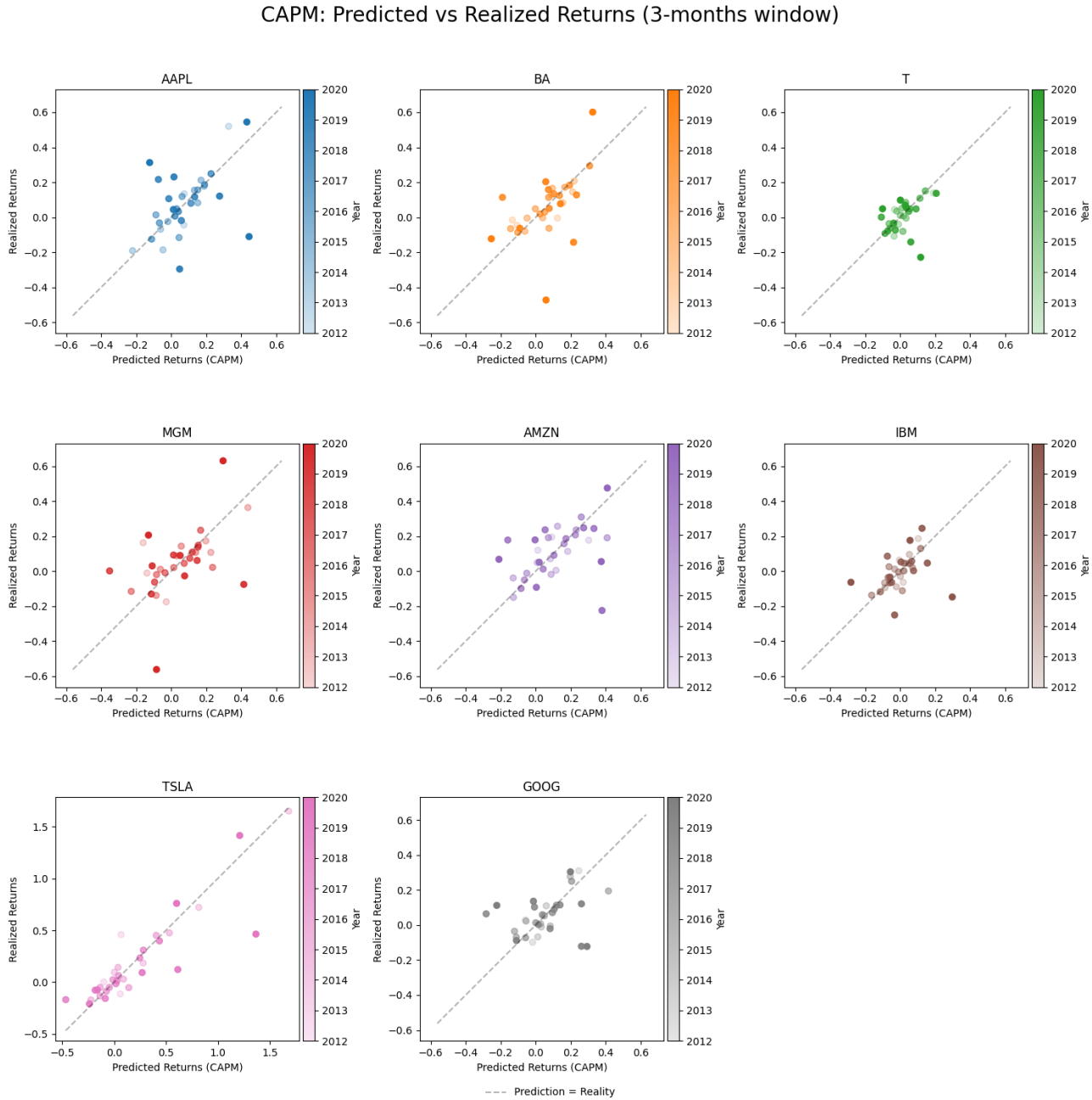
What do we learned from applying the CAPM ?

CAPM Performance – Group 1



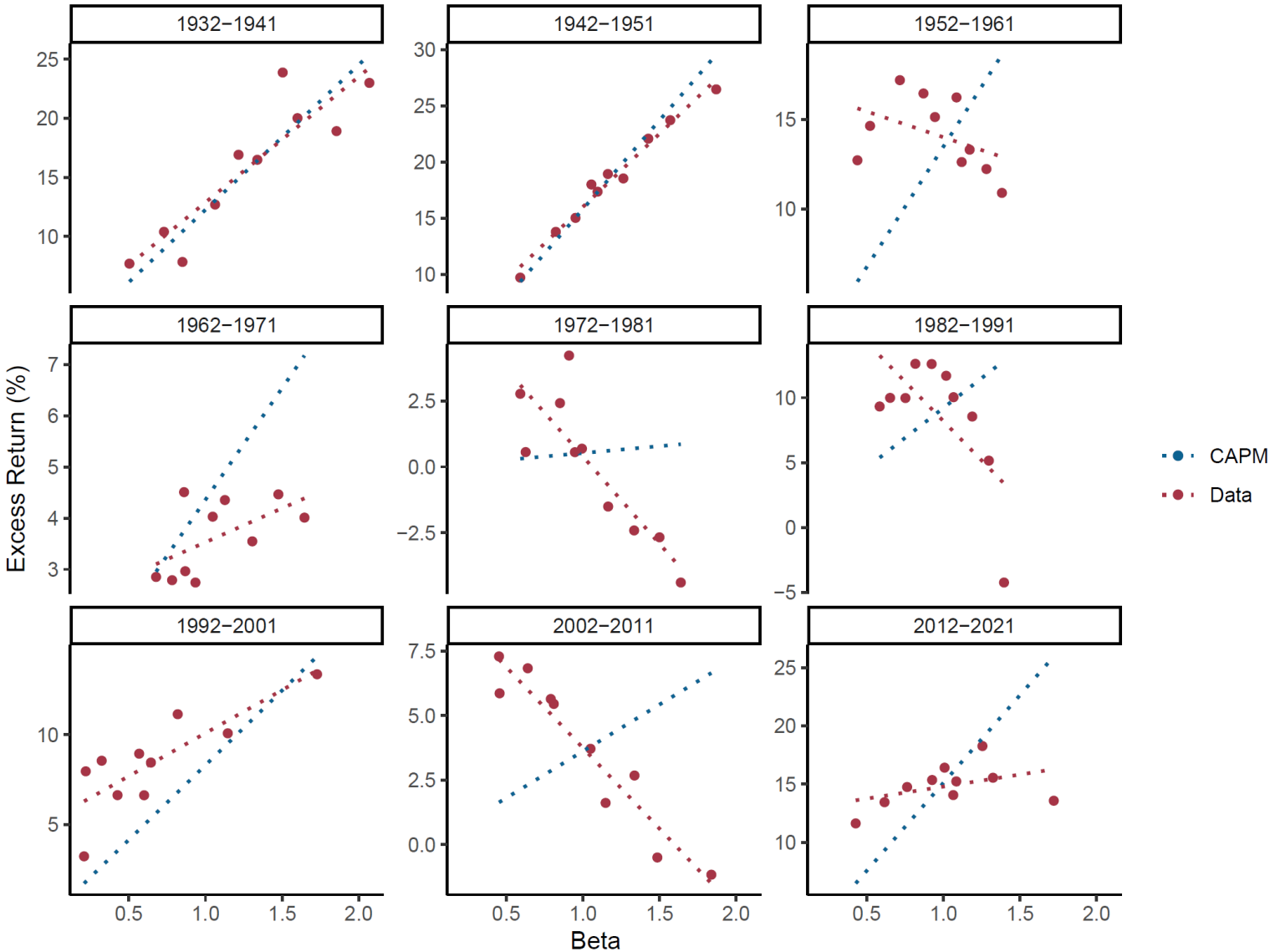
What do we learned from applying the CAPM ?

CAPM Performance – Group 11



What do we learned from applying the CAPM ?

Empirical evidence (Kogan & Wang, 2021)



Lucas critique (1976)

> This essay has been devoted to an exposition and elaboration of a single syllogism: given that the structure of an econometric model consists of optimal decision rules of economic agents, and that optimal decision rules vary systematically with changes in the structure of series relevant to the decision maker, it follows that any change in policy will systematically alter the structure of econometric models.

Portfolio Theory

APT

Bloomberg = data!
1981

IBM-Microsoft
1980

Factor « zoo »

Big data = too much data!
Machine Learning

You!

Pre-1950

1952

1964

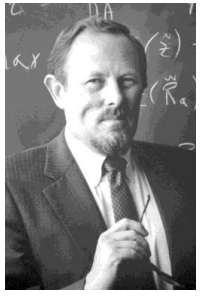
1976

1993

2011

2022

The « Chaos »



CAPM

3-Factor Model

Internet = more data!
2000

Today: From Normative to Data-Driven Finance

Mirete-Ferrer et al. (2022)

As Cochrane (2011) exposed, "In the beginning, there was chaos; practitioners thought one only needed to be clever to earn high returns. Then came the CAPM. Every clever strategy to deliver high average returns ended up delivering high market betas as well. Then anomalies erupted, and there was chaos again. The "value effect" was the most prominent anomaly". In the last 10 years, the huge amount of anomalies found has driven the academic professionals to call the phenomenon as a "factor zoo".

Fama-French 3-Factor Model

Fama–French three-factor model

Genesis

- The CAPM is an elegant theory but :
  - α is not equal to zero
  - The market return $\tilde{\mathbf{r}}_{\mathrm{M}}$ is not a great predictor alone

- Empirical contradictions to the CAPM model :
  - firm size matters (Banz, 1980)
  - leverage predicts expected return (Bhandari, 1988)
  - also : Market-to-Book ratio (Stattman, 1980 ; Rosenberg et al., 1985) and Earnings-to-Price ratio (Basu, 1983)

$\Rightarrow$ Call for a (new) **unified theory**

Fama–French three-factor model (≈ 25K citations)

THE JOURNAL OF FINANCE • VOL XLVII, NO 2 • JUNE 1992

The Cross-Section of Expected Stock Returns

EUGENE F. FAMA and KENNETH R. FRENCH*

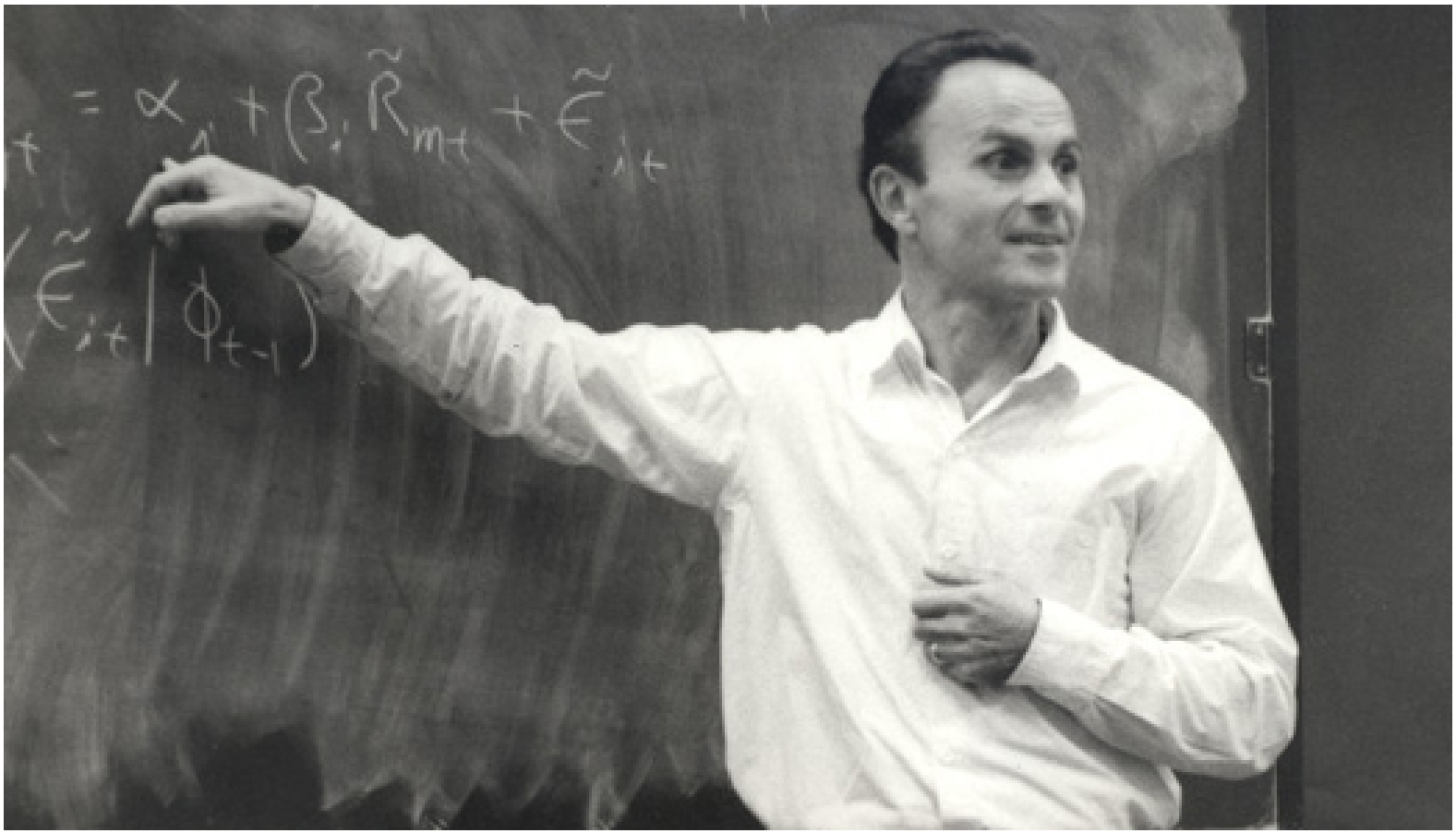
ABSTRACT

Two easily measured variables, size and book-to-market equity, combine to capture the cross-sectional variation in average stock returns associated with market β, size, leverage, book-to-market equity, and earnings-price ratios. Moreover, when the tests allow for variation in β that is unrelated to size, the relation between market β and average return is flat, even when β is the only explanatory variable.

THE ASSET-PRICING MODEL OF Sharpe (1964), Lintner (1965), and Black (1972) has long shaped the way academics and practitioners think about average returns and risk. The central prediction of the model is that the market portfolio of invested wealth is mean-variance efficient in the sense of Markowitz (1959). The efficiency of the market portfolio implies that (a) expected returns on securities are a positive linear function of their market βs (the slope in the regression of a security's return on the market's return), and (b) market βs suffice to describe the cross-section of expected returns.

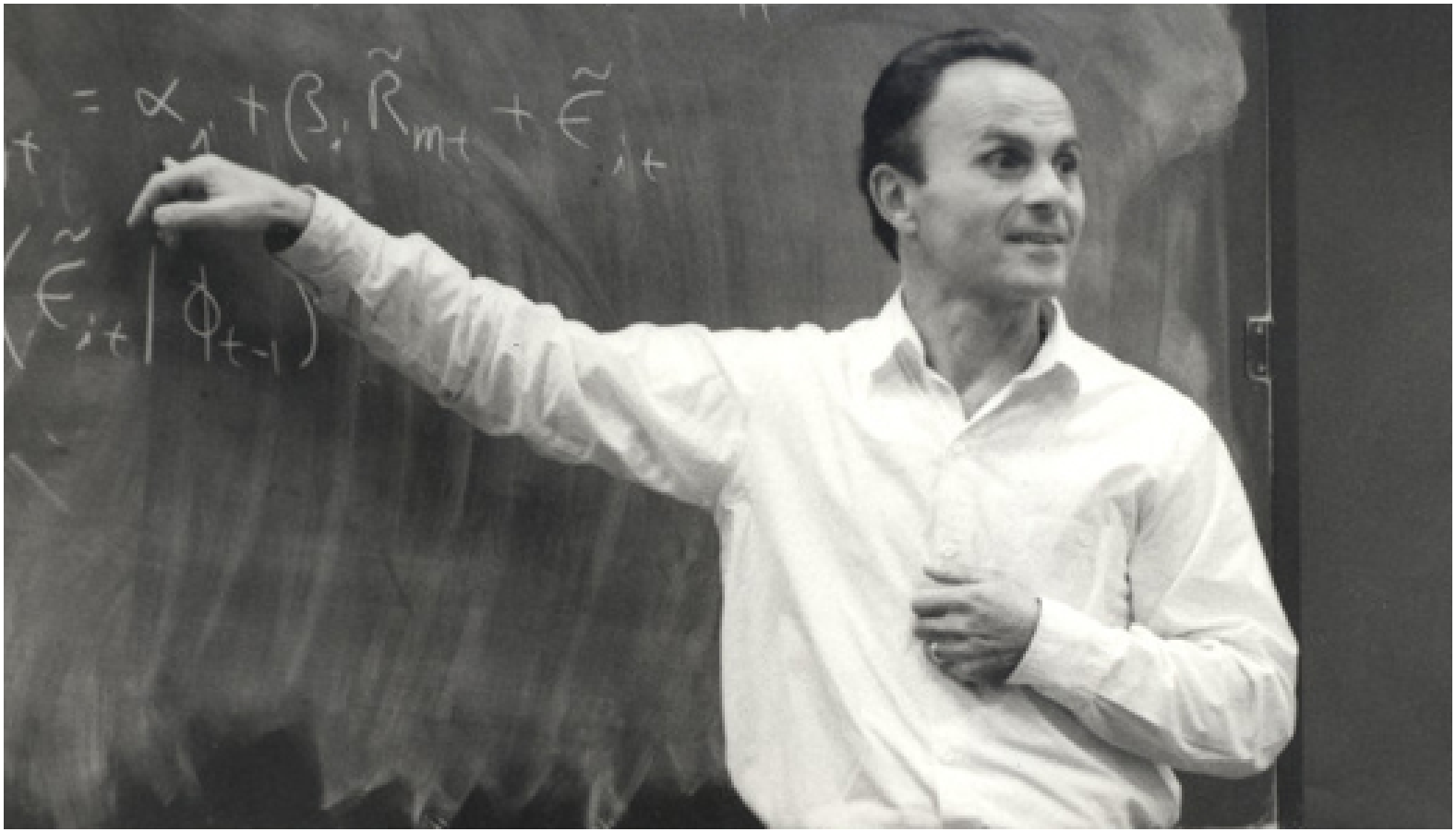
Fama–French three-factor model ($\approx$ 25K citations)

E. Fama received the Nobel Prize in 2013 for this contribution



Fama–French three-factor model ($\approx$ 25K citations)

E. Fama received the Nobel Prize in 2013 for this contribution

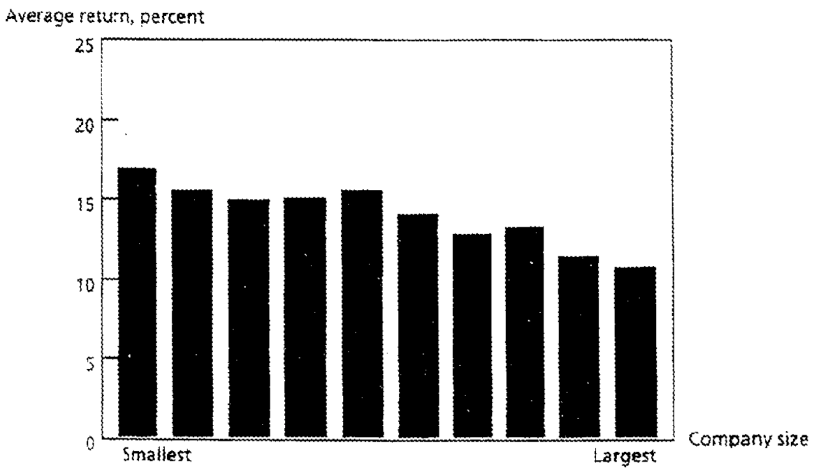


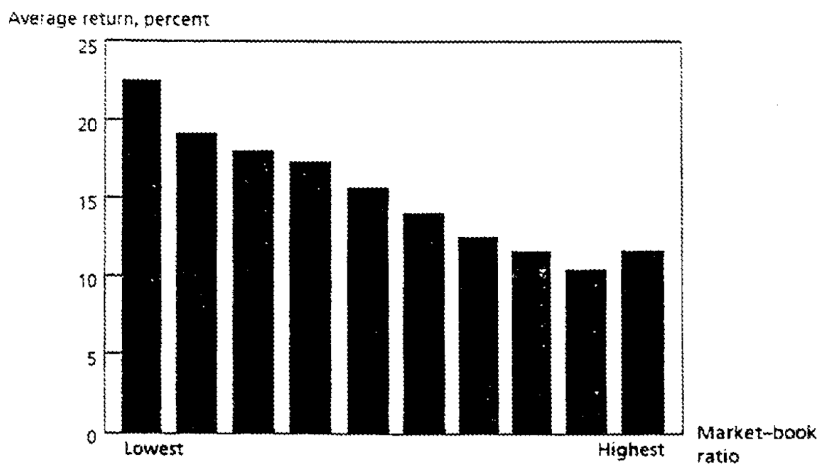
+ this guy is crazy rich

Fama–French three-factor model

Motivation

- Main idea: Firms' characteristics seem to explain (predict) returns





- Two particularly robust facts since mid-1960s :
  - Size : small stocks outperformed large stocks
  - Value : stocks with low ratios of Market-to-Book value (or high B/M) outperformed stocks with high ratios (low B/M)

Fama–French three-factor model

Ratio of Book-to-Market value

- Definition :

$$\text{Book-to-Market} = \frac{\text{Total Book Value}}{\text{Market Capitalization}}$$

- The book value is the amount that would be left if the company liquidated all of its assets and repaid all of its liabilities
- The market value is the current stock price of all outstanding shares (i.e. the price that the market believes the company is worth)
- What does it mean if this ratio is high ($>>1$) ?

Fama–French three-factor model

Ratio of Book-to-Market value

- Definition :

$$\text{Book-to-Market} = \frac{\text{Total Book Value}}{\text{Market Capitalization}}$$

- The book value is the amount that would be left if the company liquidated all of its assets and repaid all of its liabilities
- The market value is the current stock price of all outstanding shares (i.e. the price that the market believes the company is worth)
- What does it mean if this ratio is high ($>>1$) ?
  - high B/M $\rightarrow$ company is undervalued (value stock)
  - low B/M $\rightarrow$ investors expect high future cash-flows (growth stock)

Fama–French three-factor model

The Economic Debate

- Small stocks and high B/M stocks earn positive abnormal returns according to the CAPM. What does it mean ?

Fama–French three-factor model

The Economic Debate

- Small stocks and high B/M stocks earn positive abnormal returns according to the CAPM. What does it mean ?

- Two possible conclusions :
  - Small stocks and high B/M stocks are mispriced (the market is wrong)
  - The CAPM doesn't measure risk properly $\rightarrow$ missing factors

Fama–French three-factor model

The Economic Debate

- Fama-French's interpretation: (hidden) risk that could be modeled by other factors

Fama–French three-factor model

The Economic Debate

- Fama-French's interpretation: (hidden) risk that could be modeled by other factors
- Construct a 3-factors asset-pricing model, which makes small and high B/M stocks look riskier
- Add factors SMB and HML and show that small stocks load heavily on the SMB factor and value stocks load heavily on the HML factor
- Economic intuition ?

Fama–French three-factor model

The Economic Debate

- Fama-French's interpretation: (hidden) risk that could be modeled by other factors
- Construct a 3-factors asset-pricing model, which makes small and high B/M stocks look riskier
- Add factors SMB and HML and show that small stocks load heavily on the SMB factor and value stocks load heavily on the HML factor
- Economic intuition ? Stocks with different economic characteristics have different betas on systematic economic shocks

Fama–French three-factor model

- In practice, they add 2 factors to the CAPM Model:
  - Market : $r_M - r_f$ (CAPM)
  - **SMB** (Small-Minus-Big) : Size
  - **HML** (High-Minus-Low) : Value

Fama–French three-factor model

- In practice, they add 2 factors to the CAPM Model:
  - Market : $r_M - r_f$ (CAPM)
  - **SMB** (Small-Minus-Big) : Size
  - **HML** (High-Minus-Low) : Value

Fama-French 3-Factor Model :

$$\tilde{r}_i - r_f = \alpha_i + \beta_i \left(\tilde{\mathbf{r}}_{\mathbf{M}} - r_f\right) + \gamma_i \, \mathbf{SMB}_t + \delta_i \, \mathbf{HML}_t + \epsilon_i$$

$\Rightarrow$ Key question : How **SMB** and **HML** are constructed ?

How to Make Factors ?

How to Make Factors ?

1. Collect data on excess returns r_t^i

How to Make Factors ?

1. Collect data on excess returns r_t^i
2. Collect data on some characteristics (or signal) s_t^i for firm i at time t
   - Make sure that the set of securities is free of selection bias
   - Make sure that s_t^i was known before the beginning of period t (no look-ahead bias)
   - Clean the data (without introducing biases)

How to Make Factors ?

1. Collect data on excess returns r_t^i

2. Collect data on some characteristics (or signal) s_t^i for firm i at time t

   - Make sure that the set of securities is free of selection bias
   - Make sure that s_t^i was known before the beginning of period t (no look-ahead bias)
   - Clean the data (without introducing biases)

3. Compute the difference of return for stocks with high signals vs. low signals :

$$\tilde{f}_{t+1} = \sum_i w_t^{i,H} r_{t+1}^i - \sum_i w_t^{i,L} r_{t+1}^i$$

How to Make Factors ?

1. Collect data on excess returns r_t^i
2. Collect data on some characteristics (or signal) s_t^i for firm i at time t
   - Make sure that the set of securities is free of selection bias
   - Make sure that s_t^i was known before the beginning of period t (no look-ahead bias)
   - Clean the data (without introducing biases)
3. Compute the difference of return for stocks with high signals vs. low signals :

$$\tilde{f}_{t+1} = \sum_i w_t^{i,H} r_{t+1}^i - \sum_i w_t^{i,L} r_{t+1}^i$$

4. define weights of the *high* $w_t^{i,H}$ and *low* $w_t^{i,L}$

How to Make Factors ?

Example with Fama-French (1992)

- Construction of $\mathbf{SMB}$ and $\mathbf{HML}$:
  1. Split stocks in 2 according to size
  2. Split stocks in 3 according to Book-to-Market
  3. Obtain 2×3 portfolios

	Median ME	
70^{th} percentile of B/M−	Small Value	Big Value
	Small Neutral	Big Neutral
30^{th} percentile of B/M−	Small Growth	Big Growth

How to Make Factors ?

Example with Fama-French (1992)

- Construction of $\mathbf{SMB}$ and $\mathbf{HML}$:
  1. Split stocks in 2 according to size
  2. Split stocks in 3 according to Book-to-Market
  3. Obtain 2×3 portfolios

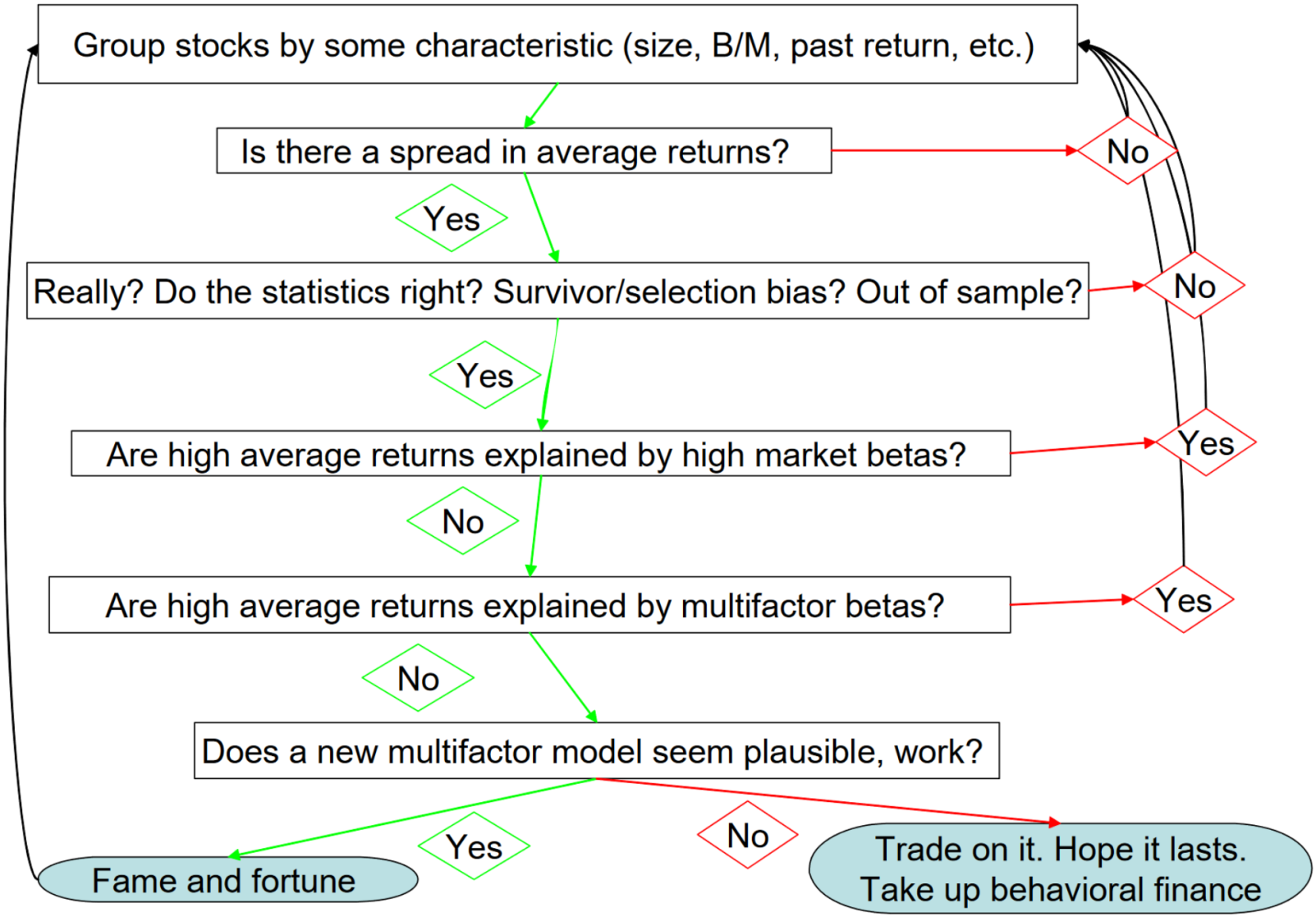
	Median ME	
70^{th} percentile of B/M−	Small Value	Big Value
	Small Neutral	Big Neutral
30^{th} percentile of B/M−	Small Growth	Big Growth

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

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

How to Make Factors?

Flowchart



More Factors ?

Fama-French (2015)

- Fama-French (1992) is an incomplete model for expected returns because its 3 factors miss much of the variation in average returns related to :
  - profitability
  - investment

- According to you, what did they do ?

More Factors ?

Fama-French (2015)

- Fama-French (1992) is an incomplete model for expected returns because its 3 factors miss much of the variation in average returns related to :
  - profitability
  - investment

- According to you, what did they do ?

→ The authors offered an extended version of their model :

$$\tilde{r}_i - r_f = \alpha_i + \beta_i \left(\tilde{\mathbf{r}}_{\mathbf{M}} - r_f\right) + \gamma_i \mathbf{SMB}_t + \delta_i \mathbf{HML}_t + \omega_i \mathbf{RMW}_t + \kappa_i \mathbf{CMA}_t + \epsilon_i$$

  - $\mathbf{RMW}$ (robust-minus-weak) : profitability
  - $\mathbf{CMA}$ (conservative-minus-aggressive) : investment
    (i) conservative $\equiv$ low investment

- No attempt to make HML, RMW and CMA uncorrelated !

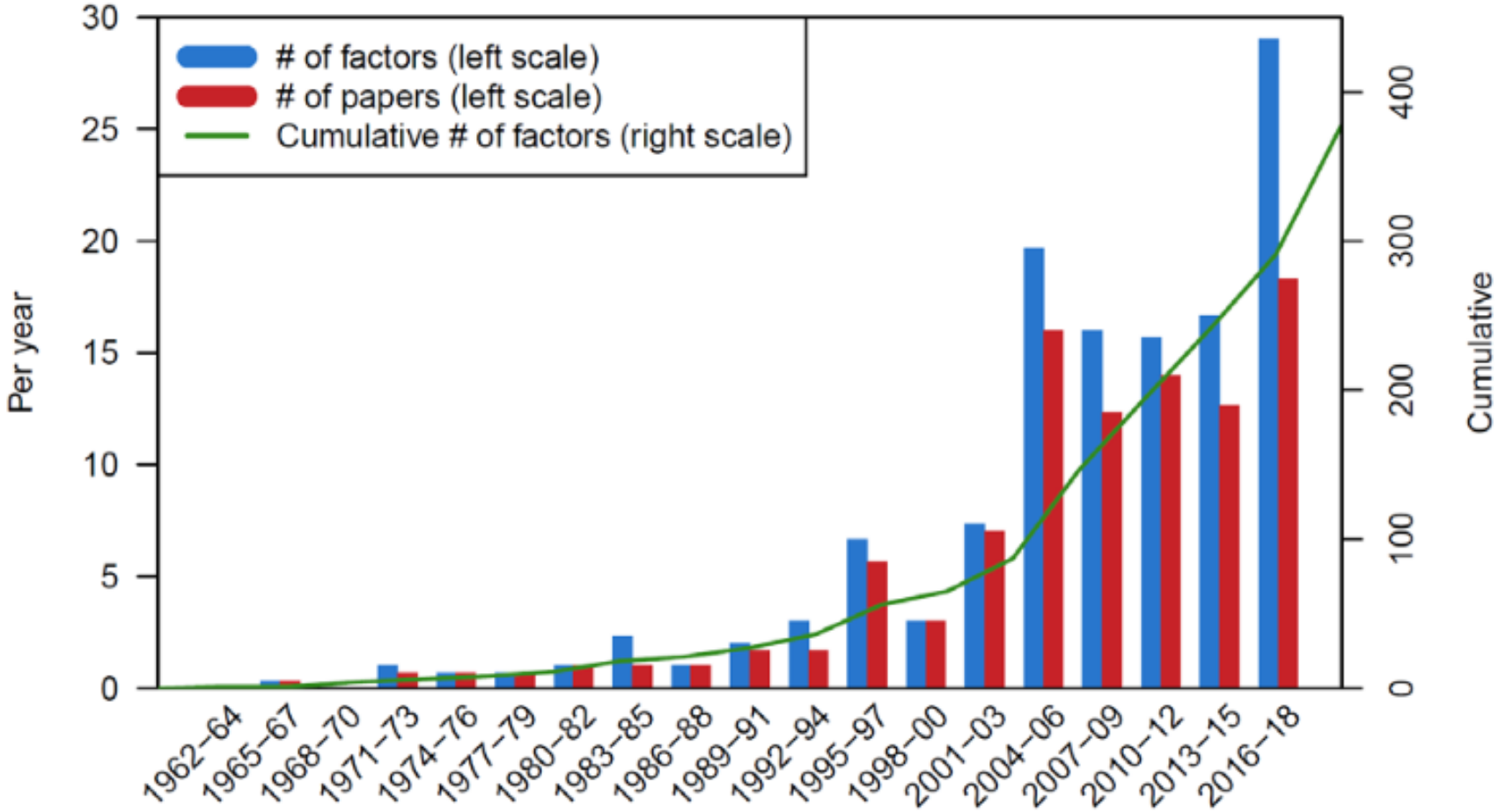
More Factors ?

- Hundreds of factors claim to predict stock returns :
  - Value and momentum
  - Accruals
  - Short-term reversal
  - Long-term reversal
  - QMJ = Quality-Minus-Junk
  - Environmental, social, governance (ESG)
  - Liquidity risk

Too Many Factors?

Harvey and Liu (2016)

Figure 1. Out of control factor production - through January 2019



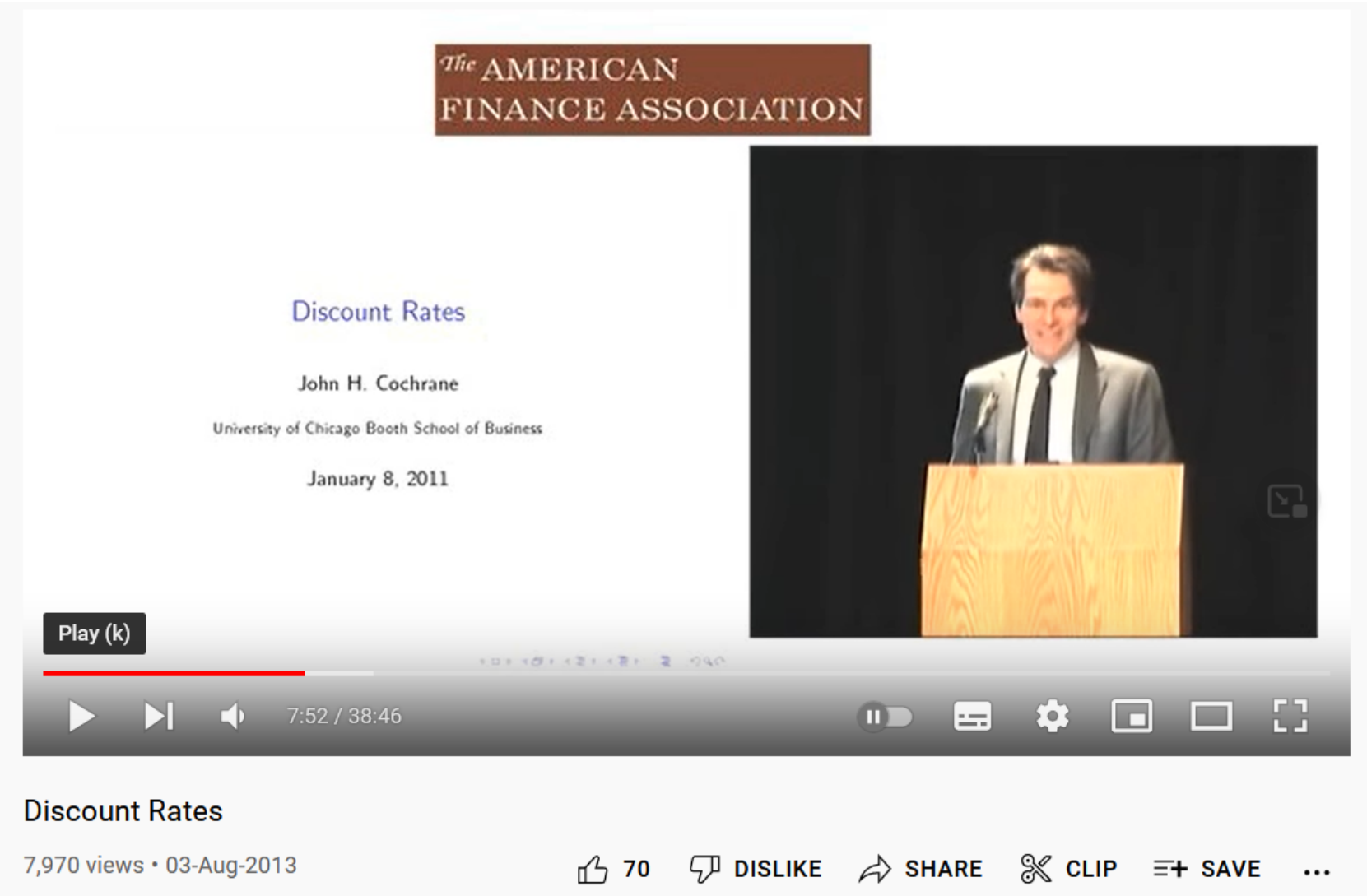
Too Many Factors?

Harvey and Liu (2016)

The rate of factor production in the academic research is out of control. We document over 400 factors published in top journals. Surely, many of them are false. We explore the incentives that lead to factor mining and explore reasons why many of the factors are simply lucky findings. The backtested results published in academic outlets are routinely cited to support commercial products. As a consequence, investors develop exaggerated expectations based on inflated backtested results and are then disappointed by the live trading experience.

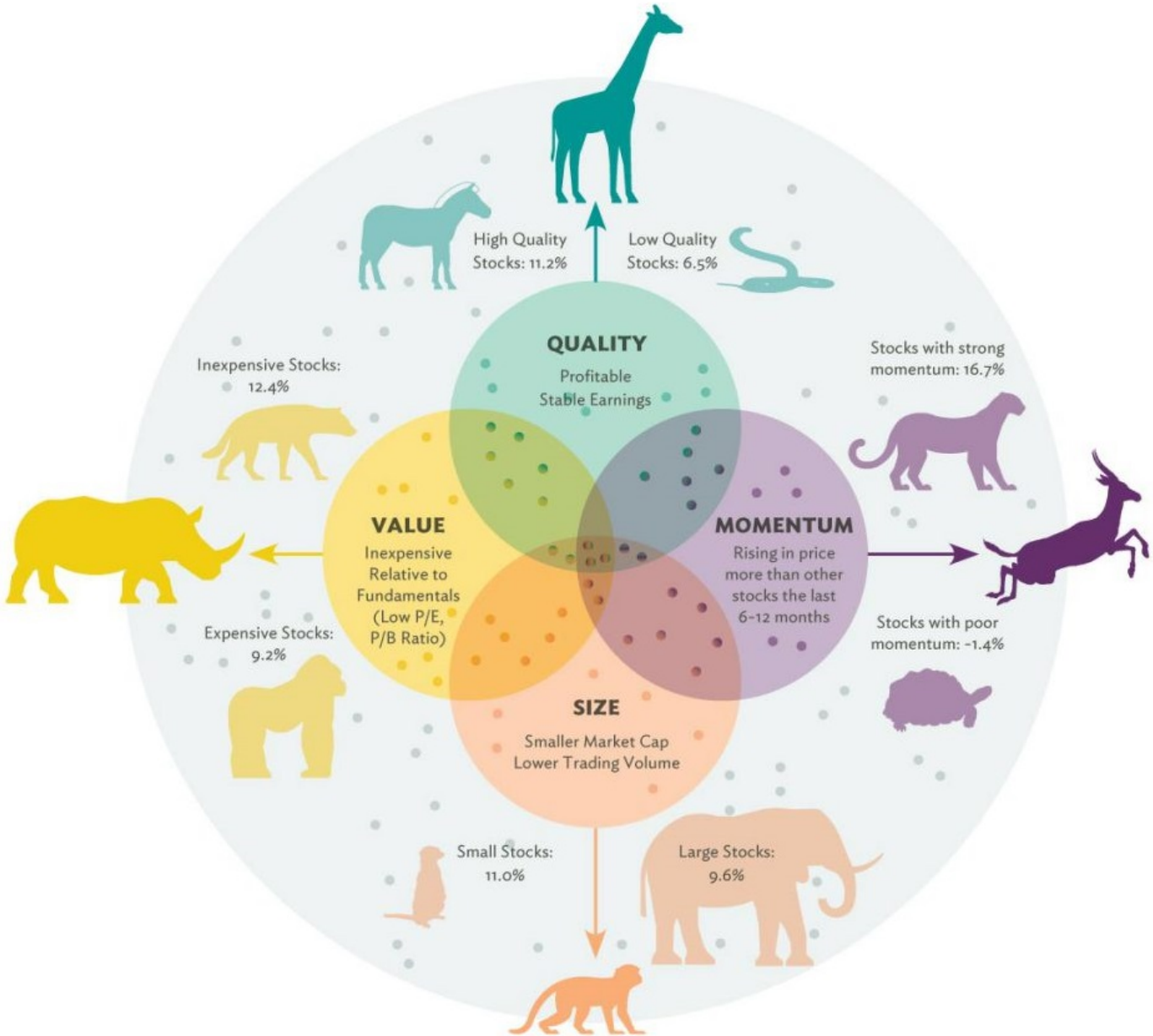

The Factor Zoo

Cochrane (2011): Factor **Zoo**



- Link to the video : [J. Cochrane AFA, 2011]

Cochrane (2011) : Factor **Zoo**



Cochrane (2011) : Factor **Zoo**

- What are the concerns ?
  - Data mining or "p-hacking"
  - Publication bias
  - Factors should be economically motivated

Cochrane (2011) : Factor **Zoo**

- What are the concerns ?
  - Data mining or "p-hacking"
  - Publication bias
  - Factors should be economically motivated
  - **Academia vs. real world**

Portfolio Theory

APT

Bloomberg = data!
1981

IBM-Microsoft
1980

Factor « zoo »

Big data = too much data!
Machine Learning

You!

Pre-1950

1952

1964

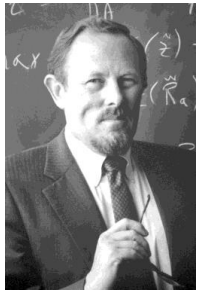
1976

1993

2011

2022

The « Chaos »



CAPM

3-Factor Model

Internet = more data!
2000

Taming the Zoo Factor [Paper Link] [Video Link]

Feng et al., 2020

- Main motivation: bring more discipline to the proliferation of factors
- Most of these factors are redundant, due to omitted variable bias from the benchmark (i.e. insufficient controls)

Taming the Zoo Factor [Paper Link] [Video Link]

Feng et al., 2020

- Main motivation: bring more discipline to the proliferation of factors
- Most of these factors are redundant, due to omitted variable bias from the benchmark (i.e. insufficient controls)
- Research question: **how to judge whether a new factor adds explanatory power for asset pricing**, relative to the hundreds of factors the literature has so far produced ?

Taming the Zoo Factor [Paper Link] [Video Link]

Feng et al., 2020

- Main motivation: bring more discipline to the proliferation of factors
- Most of these factors are redundant, due to omitted variable bias from the benchmark (i.e. insufficient controls)
- Research question: **how to judge whether a new factor adds explanatory power for asset pricing**, relative to the hundreds of factors the literature has so far produced ?
- This paper: provide a methodology for estimating and testing the marginal importance of any factor g_t in pricing the cross section of expected returns beyond what can be explained by a high-dimensional set of potential factors h_t

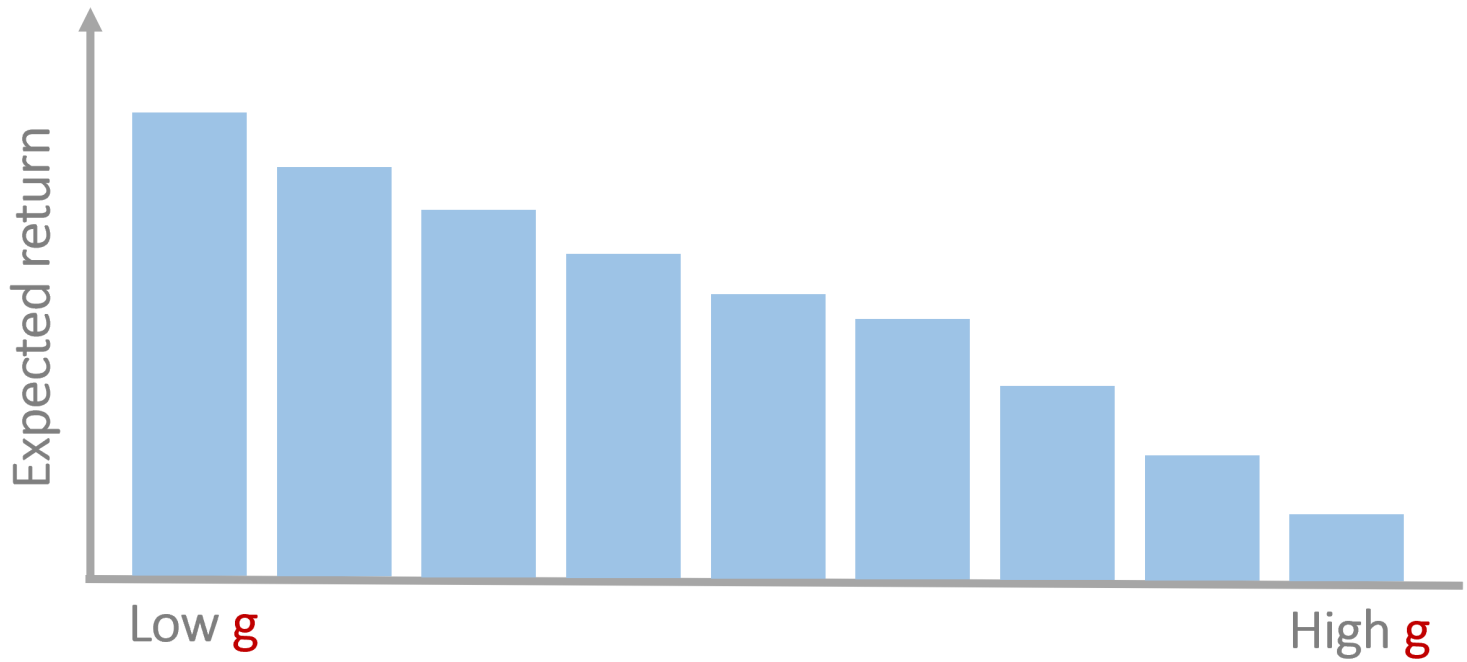
Taming the Zoo Factor

Example

Taming the Zoo Factor

Example

- You are a data scientist working for a hedge fund
- Your boss finds out that a signal g seems to matter for predicting expected returns



Taming the Zoo Factor

Example

- You want to know if $\mathbf{g}$ is correlated with other existing factors h_j
- If $\#\{j\}$ is low (e.g., the CAPM), you can test empirically with an OLS estimation of this model :

$$\tilde{r}_i - r_f = \alpha_i + \beta_i \left(\tilde{\mathbf{r}}_{\mathbf{M}} - r_f\right) + \gamma_i \, \mathbf{g} + \epsilon_i$$

- What if γ_i is significant and large ? Are you rich and famous ?

Taming the Zoo Factor

Example

- You want to know if $\mathbf{g}$ is correlated with other existing factors h_j
- If $\#\{j\}$ is low (e.g., the CAPM), you can test empirically with an OLS estimation of this model :

$$\tilde{r}_i - r_f = \alpha_i + \beta_i \left(\tilde{\mathbf{r}}_{\mathbf{M}} - r_f\right) + \gamma_i \, \mathbf{g} + \epsilon_i$$

- What if γ_i is significant and large ? Are you rich and famous ?

$\rightarrow$ Omitted variable bias

Taming the Zoo Factor

Example

- Now imagine $\#\{j\}$ is > 1, you must add other existing factors :

$$\tilde{r}_i - r_f = \alpha_i + \beta_i \left(\tilde{\mathbf{r}}_{\mathbf{M}} - r_f\right) + \gamma_i \, \mathbf{g} + \sum_j \beta_i^j h_j + \epsilon_i$$

- Issue : what if $j \approx 500$?

Taming the Zoo Factor

Example

- Now imagine $\#\{j\}$ is > 1, you must add other existing factors :

$$\tilde{r}_i - r_f = \alpha_i + \beta_i \left(\tilde{\mathbf{r}}_{\mathbf{M}} - r_f\right) + \gamma_i \, \mathbf{g} + \sum_j \beta_i^j h_j + \epsilon_i$$

- Issue : what if $j \approx 500$? $\rightarrow$ **curse of dimensionality**
  - poor estimates (in OLS, errors depends on $n - p$)
  - difficult interpretation
  - invalid inference (not robust OOS)

Taming the Zoo Factor

Example

- Now imagine $\#\{j\}$ is > 1, you must add other existing factors :

$$\tilde{r}_i - r_f = \alpha_i + \beta_i \left(\tilde{\mathbf{r}}_{\mathbf{M}} - r_f\right) + \gamma_i \, \mathbf{g} + \sum_j \beta_i^j h_j + \epsilon_i$$

- Issue : what if $j \approx 500$? $\rightarrow$ **curse of dimensionality**
  - poor estimates (in OLS, errors depends on $n - p$)
  - difficult interpretation
  - invalid inference (not robust OOS)
- As data scientists, what is your solution ?

Taming the Zoo Factor

Reducing the dimensionality

- They use machine learning / model selection methods to reduce the dimensionality of the factor zoo

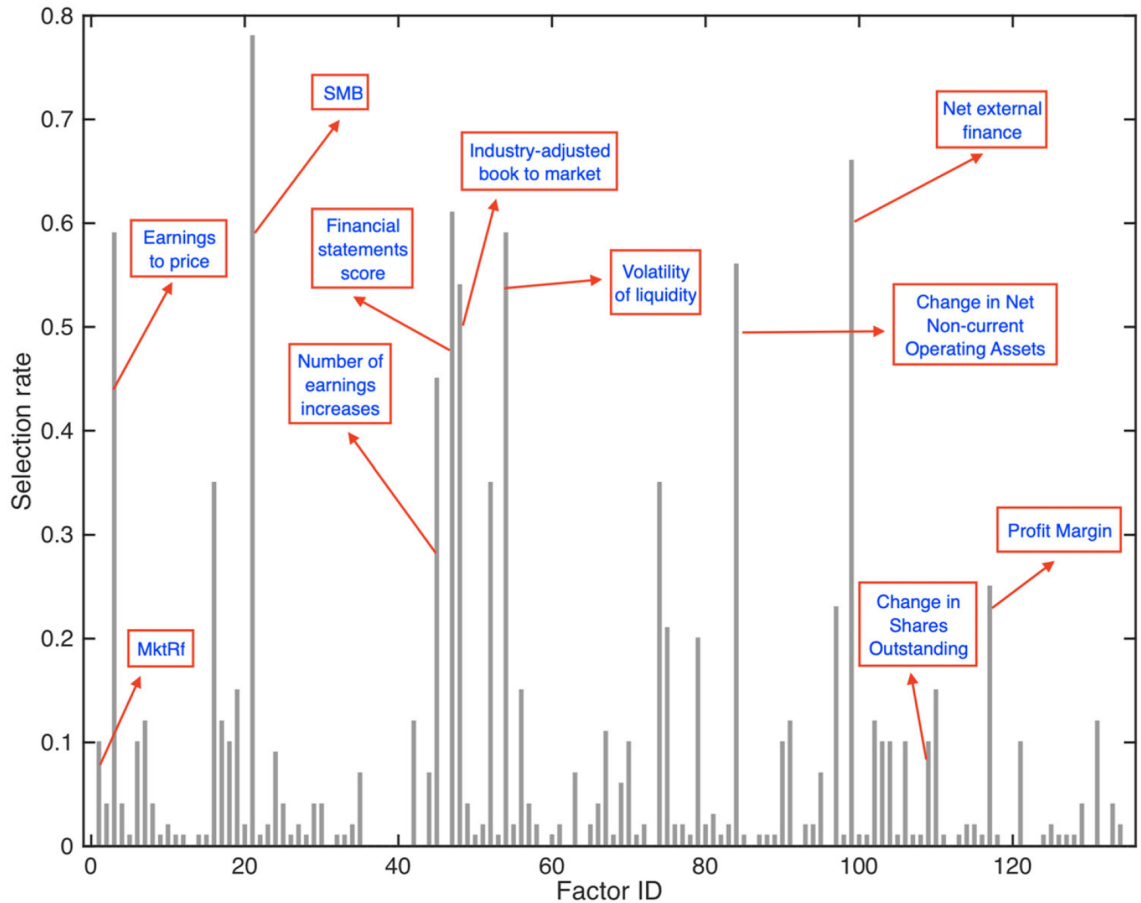
Taming the Zoo Factor

Reducing the dimensionality

- They use machine learning / model selection methods to reduce the dimensionality of the factor zoo $\rightarrow$ LASSO
- Have to take into account potential model-selection mistakes
  - LASSO may miss some true factors $\rightarrow$ omitted variable bias occurs again.
  - LASSO may miss some *non-true* factors but highly correlated with g
  - see Belloni, Chernozhukov, and Hansen (2014)

Taming the Zoo Factor : LASSO is unstable

- Randomly draw a bootstrap sub-sample, regress average returns on factors to check if a factor is selected



- Even the Market is selected only 10% of the time

Taming the Zoo Factor

The double-selection estimation procedure

- We have j candidate factors h for the benchmark group for $\mathbf{g}$.
- We want to select a sub-sample:
  1. implement LASSO to select N_1 factors h_j^1 explaining expected returns

  $\rightarrow$ dimension-reduction (*true* asset pricing model)
  $\rightarrow$ Factors which appear to contribute little to pricing assets in the cross section are excluded from the set of controls

Taming the Zoo Factor

The double-selection estimation procedure

- We have j candidate factors h for the benchmark group for $\mathbf{g}$.
- We want to select a sub-sample:
  1. implement LASSO to select N_1 factors h_j^1 explaining expected returns

  $\rightarrow$ dimension-reduction (*true* asset pricing model)
  $\rightarrow$ Factors which appear to contribute little to pricing assets in the cross section are excluded from the set of controls

  2. implement LASSO to select N_2 factors h_j^2 that are highly correlated with g

  $\rightarrow$ identify the factors most likely to cause omitted variable bias
  $\rightarrow$ variables selected from this stage are redundant or even useless

Taming the Zoo Factor

The double-selection estimation procedure

- We have j candidate factors h for the benchmark group for $\mathbf{g}$.
- We want to select a sub-sample:
  1. implement LASSO to select N_1 factors h_j^1 explaining expected returns

  $\rightarrow$ dimension-reduction (*true* asset pricing model)
  $\rightarrow$ Factors which appear to contribute little to pricing assets in the cross section are excluded from the set of controls

  2. implement LASSO to select N_2 factors h_j^2 that are highly correlated with g

  $\rightarrow$ identify the factors most likely to cause omitted variable bias
  $\rightarrow$ variables selected from this stage are redundant or even useless

- Post-selection: regress expected return on $\mathbf{g}$ + control variables selected in steps (1) and (2) – benchmark

Taming the Zoo Factor

Data

- 99 monthly factors covering from Jul. 1980 to Dec. 2016
- Also explored adding 197 non-tradable factors (squared factors + interactions with SMB)
- They use a total of 1,825 portfolios as test assets

Taming the Zoo Factor : Results

- Strongest new factors: profitability and investment
- Many others are redundant

Year	(1) # Assets	(2) # Controls	(3) New factors (IDs)											
1994	138	25	26	27										
1995	150	27	28	29	30									
1996	150	30	31	32	33									
1997	168	33	34											
1998	174	34	35	36	37	38	39	40	41	42	43	44		
1999	228	44	45	46										
2000	234	46	47	48	49	50	51							
2001	252	51	52	53	54	55	56	57	58					
2002	294	58	59	60	61									
2003	312	61	62	63	64	65	66							
2004	336	66	67	68	69	70	71	72	73	74				
2005	372	74	75	76	77	78	79	80	81	82	83	84	85	86
			87	88	89	90								
2006	456	90	91	92	93	94	95	96	97	98	99	100	101	102
2007	516	102	103	104	105	106	107	108						
2008	552	108	109	110	111	112	113	114	115	116	117	118	119	120
2009	618	120	121	122	123	124								
2010	636	124	125	126	127	128	129							
2011	666	129	130	131	132	133	134	135						
2012	702	135	136											
2013	708	136	137	138	139									
2014	720	139	140	141	142	143	144							
2015	738	144	145	146	147	148								
2016	750	148	149	150										

Taming the Zoo Factor : Results

Robustness

- Results are similar when using alternative methods (ElasticNet, PCA, etc.)

What do we learned ?

- Following Fama-French (1992), vast empirical evidence against the CAPM
- More data available + more computing power = more factors
- From 1980 onward: proliferation of factors (Factor Zoo)
- Recent developments: machine learning methods help taming the zoo factor

Discussion