Data Driven Methods in Finance: basics II

Fall 2023: IEOR 4576

Naftali Cohen

Question/s of the week

Question: We roll a 6-sided die n times. What is the probability that all faces have appeared?

Answer:

Alpha

What is alpha?

- Alpha (α) represents the **excess return** of a portfolio over the return of a reference instrument.
- α is a measure of the risk-adjusted excess return, which is the portfolio's performance after accounting for its risk relative to the reference instrument.
- Increasing the α therefore means increasing the portfolio's return **without increasing its direct exposure risk** to the reference instrument.

Alpha

Three types of alpha, even though sometimes people just mean outperformance.

- Benchmark Alpha: The return of the portfolio not associated with its risk with respect to the underlying benchmark. Given the portfolio return r_P and the benchmark return r_B, we can estimate the following equation: $r_P = \alpha + \beta r_B + \epsilon$

Alpha

Benchmark Alpha: $r_P = \alpha + \beta r_B + \epsilon$

A few remarks:

- βr_B is the expected or consensus return, which is the part of the portfolio's return related to the benchmark
- $\alpha + \epsilon$ is the residual return
- α is the expected value of the residual return
- ϵ is the deviation of the residual return from its mean (it is assumed that ϵ averages zero).
- The residual return is all that matters to the quantitative portfolio manager
- If the benchmark return is positive, it is easy enough to generate higher returns simply by increasing the portfolio's exposure to the benchmark, but the portfolio manager has not added value – **just added more risk**.
- The residual return is the part that represents an increase in return independent of increased direct benchmark exposure.

Alpha

Three types of alpha, even though sometimes people just mean outperformance.

- Benchmark Alpha: $r_P = \alpha + \beta r_B + \epsilon$
- CAPM Alpha: The return of the portfolio not associated with its risk with respect to the underlying market portfolio. Given the portfolio return r_P and the market return r_M, we can estimate the following equation: $r_P = \alpha + \beta r_M + \epsilon$

Alpha

CAPM Alpha: $r_P = \alpha + \beta r_M + \epsilon$

A few remarks:

- In this equation, βr_M is the expected or consensus return, which is the part of the portfolio's return related to the market.
- The remaining $\alpha + \epsilon$ is the residual return over the market.
- If the benchmark is S&P 500, then CAPM's α and Benchmark α are the same.
- According to CAPM theory, the CAPM α should equal zero. When it is significantly positive, the portfolio manager is providing excess risk-adjusted returns.

Alpha

Three types of alpha, even though sometimes people just mean outperformance.

- Benchmark Alpha: $r_P = \alpha + \beta r_B + \epsilon$
- CAPM Alpha: $r_P = \alpha + \beta r_M + \epsilon$
- Multi-Factor Alpha: The return of the portfolio not associated with its risk with respect to the underlying multi-factors. Given the portfolio return r_P and a series of factor returns $f_1, \dots, f_K$, we can estimate the following equation: $r_P = \alpha + \beta_1 f_1 + \dots + \beta_K f_K + \epsilon$

Alpha

Multi-Factor Alpha: $r_P = \alpha + \beta_1 f_1 + ... + \beta_K f_K + \epsilon$

A few remarks:

- $\beta_1 f_1 + ... + \beta_K f_K$ is the expected or consensus return from a multifactor model of stock returns
- The remaining $\alpha + \epsilon$ is the residual return, as before
- α is created using a model with many factors that influence stock returns
- α is a measurement of risk-adjusted excess return given multiple explanatory variables.

Alpha

Benchmark Alpha: $r_P = \alpha + \beta r_B + \epsilon$

CAPM Alpha: $r_P = \alpha + \beta r_M + \epsilon$

Multi-Factor Alpha: $r_P = \alpha + \beta_1 f_1 + \ldots + \beta_K f_K + \epsilon$

Notice*:*

- The multifactor α of a stock or a portfolio is the same as the CAPM α if the market return is the only factor in the model.
- The benchmark α of a stock or a portfolio is the same as the CAPM α if the market portfolio is the benchmark.
- The multifactor α and the benchmark α of a stock or portfolio are the same if the market return is the only factor in the model and is also the benchmark.

Alpha

Statement: if the benchmark is <u>inefficient</u> it's alpha <u>against the market</u> is negative.

Consider the following models:

- $r_P = \alpha' + \beta' r_B + \epsilon'$,
- $r_P = \alpha'' + \beta'' r_M + \epsilon''$, and
- $r_B = \alpha + \beta r_M + \epsilon$.

Substitution yields

$$r_P = \alpha' + \beta'\alpha + \beta'\beta r_M + \beta'\epsilon + \epsilon'$$

and it follows that

$$\alpha'' = \alpha' + \beta'\alpha.$$

Thus, for a positive β', if $\alpha < 0$ *it follows that*

$$\alpha' > \alpha'' \text{ or } \alpha^B > \alpha^M$$

Using an inefficient benchmark makes the portfolio's performance look better than it would if it were measured against the entire market.

Expected vs. Realized

There is usually a difference between the value we expect it to reach in the future and the value that it ultimately realizes.

- Ex-ante α is the expected α
- Ex-post α is the realized α.
- After the portfolio has been managed, take portfolio returns and benchmark returns and run a regression: $r_P = \alpha + \beta r_B + \epsilon$.

A good manager can realistically hope for is that the ex-post and ex-ante α's will be highly correlated.

Expected vs. Realized

The information ratio adjusts α for the portfolio's residual risk. It measures the expected excess return over the benchmark per unit of excess risk over the benchmark.

$$\text{Expected } IR = \alpha_B / \omega, \qquad \omega = S(\epsilon)$$

Expected vs. Realized

An active portfolio manager wants:

- High Alpha
- High Information Ratio

The Rules of the game

Setting boundaries to what's achievable:

1) **Markets are mostly efficient**
    - it is not possible to make profits without taking risk
2) **Pure arbitrage opportunities do not exist**
    - Risk-free opportunities do not exist. But, still, markets are not perfectly efficient

3) **QR creates statistical arbitrage opportunities**
    - There is room to make profits by taking relatively small amounts of additional risk.
    - Statistical arbitrage opportunities exist because stock prices do not always properly reflect all the available information.
    - QR provides statistical methods for identifying information that the market has overlooked.

QR = Quant Research

The Rules of the game

4) **QR combines information in an efficient way**

- Researchers must combine all the available information into an efficient model in order to identify the pieces that are key to earning higher returns.

Decision making standards:

5) **QR should be based on sound economic theories**.

- There should be reason to believe that stock prices do not yet, **but eventually will**, reflect the information contained in each factor or in the group of factors as combined in the model.
- The model is not meaningful without a strong theoretical underpinning.

QR = Quant Research

The Rules of the game

6) **QR should reflect persistent, robust, and stable patterns**.
   - Likewise, the relationship between the factors and stock returns needs to be one that holds up over time.
   - The model should only make use of data patterns that are persistent and stable.
   - Parameter stability is essential to the generation of precise estimates and reliable forecasts with the model.
7) **Deviations of a portfolio from the benchmark are justified only if the uncertainty is small enough**.
   - Deviations from the benchmark are not always necessary.
   - A deviation from the benchmark portfolio is justified only if the portfolio's estimation error is small enough.
   - Quantitative managers always should take parameter uncertainty into account before steering the portfolio away from the benchmark onto another path dictated by the model

Market efficiency

Efficient Market: A market where all information is already reflected in stock prices

- In a fully efficient market, Quantitative Research (QR) offers no advantage.
- No excess returns from trading on news like CEO changes; price already accounts for this.
- Alpha generation by portfolio managers is impossible without added risk.
- Stock prices move randomly, showing no patterns for exploitation.

Market efficiency

Type of Efficiency	Definition	Practical Implication
Weak form	Prices reflect past trading information	Technical analysis won't help in stock picking
Semistrong form	Prices include all public information	Public data-based models won't yield excess returns
Strong form	Prices account for all information, public and private	Even inside or proprietary information won't give an edge

Market efficiency

Testing Market Efficiency Types

Weak-Form:

- Findings: Largely efficient, exceptions include momentum and reversals (autocorrelation).

Semi-Strong Form:

- Findings: Support wanes due to documented anomalies such as earnings surprises, January effect, and small-cap neglect.
- Quant Managers: Exploit inefficiencies using macro/fundamental data (slow information diffusion) or proprietary strategies (become nonpublic information).

Strong Form:

- Legal Constraints: Insider trading is illegal.
- Private to Public: Inside information often goes public.
- Testability: Difficult to validate, generally less substantiated.

Anomalies

Investment experts and researchers have identified anomalies in historical financial data that challenge the efficient market hypothesis.

Value and Size:

Value Effect

- Indicators: P/E, P/B ratios
- Example: A stock with low P/E may be considered undervalued

Size Effect

- Indicator: Market Cap
- Example: Small-cap stocks often outperform large-cap stocks

Anomalies

Investment experts and researchers have identified anomalies in historical financial data that challenge the efficient market hypothesis.

Calendar Effects:

January Effect

- Low market cap, poor returns last year
- Example: A low-cap stock rises in January after falling last year

Other Calendar Effects

- Days, seasons, weather
- Example: Stock prices tend to rise around Halloween

Anomalies

Investment experts and researchers have identified anomalies in historical financial data that challenge the efficient market hypothesis.

Less Followed, More Profit?

Neglected Firm Effect

- Fewer analysts following
- Example: A stock followed by 2 analysts outperforms one followed by 20

Anomalies

Investment experts and researchers have identified anomalies in historical financial data that challenge the efficient market hypothesis.

New Stocks and Index Changes

IPO Effect

- IPOs often underperform initially
- Example: A newly IPO'd stock drops in its first year

Index Change Effect

- Being added to or dropped from an index matters
- Example: A stock gets added to the S&P 500 and its price rises

Anomalies

Investment experts and researchers have identified anomalies in historical financial data that challenge the efficient market hypothesis.

Momentum and Technicals

Momentum

- Past success predicts future gains
- Example: A stock that gained 20% last year gains another 15% this year

Other Technical Effects

- Volume, RSI, Bollinger bands, etc.
- Example: Stock breaks above its moving average, signaling a buy

Anomalies

Investment experts and researchers have identified anomalies in historical financial data that challenge the efficient market hypothesis.

Analysts and Strategies

Wall Street Forecasts

- Buy ratings, earnings surprises
- Example: A stock with a 'Buy' rating outperforms the market

Insider Trading

- Insider buying and selling signals
- Example: CEO buys shares, and the stock rises

Anomalies

Investment experts and researchers have identified anomalies in historical financial data that challenge the efficient market hypothesis.

More Strategies

Overreaction

- "Loser" stocks can win long-term
- Example: A stock drops 30% on bad news but recovers over 2 years

Stock Buybacks & Splits

- Signals for positive returns
- Example: Company announces stock buyback and price rises

Anomalies

Investment experts and researchers have identified anomalies in historical financial data that challenge the efficient market hypothesis.

Accounting and Volatility

Accruals (accounting adjustments)

- Low accruals perform better
- Example: A stock with low accruals beats earnings estimates

Low Volatility

- Lower-risk stocks outperform
- Example: A low-volatility stock gains steadily over time

Anomalies

Investment experts and researchers have identified anomalies in historical financial data that challenge the efficient market hypothesis.

Other Factors

Low Beta

- Lower beta = better risk-adjusted returns
- Example: A low-beta stock has smaller price swings but solid returns

Liquidity

- Less liquid, more returns
- Example: A stock with low trading volume outperforms

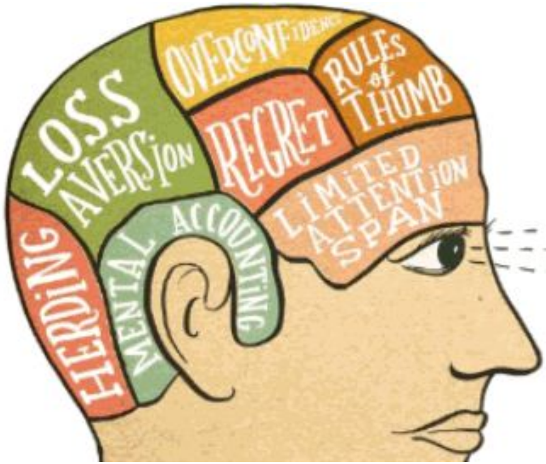
Crowding

- Less popular stocks can outperform
- Example: A stock not heavily invested in by fund managers rises

Behavioral biases

Behavioral finance presents a divergent viewpoint, proposing that psychological elements and irrational conduct play a significant role in creating anomalies within the market. These include biases like:

- **Anchoring**: The tendency to rely too heavily on the first piece of information encountered (the "anchor") when making decisions.
- **Ambiguity Aversion**: The preference for known risks over unknown risks, leading to avoidance of uncertainty.
- **Herding Mentality**: The phenomenon where individuals follow the majority, often irrationally, in decision-making processes, potentially leading to bubbles or crashes.



Behavioral biases

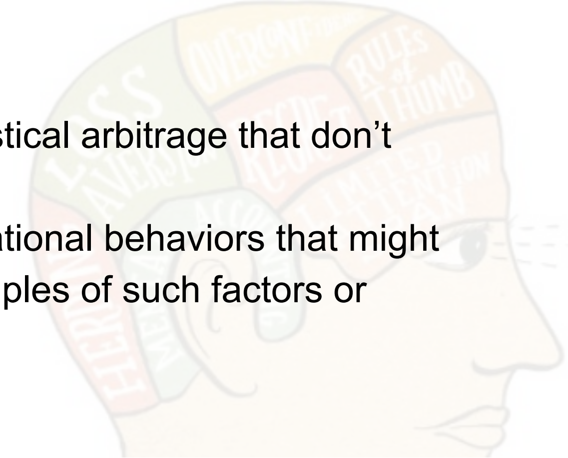
Behavioral finance presents a divergent viewpoint, proposing that psychological elements and irrational conduct play a significant role in creating anomalies within the market. These include biases like:

- **Anchoring**: The tendency to rely too heavily on the first piece of information encountered (the "anchor") when making decisions.
- **Ambiguity Aversion**: The preference for known risks over unknown risks, leading to avoidance of uncertainty.
- **Herding Mentality**: The phenomenon where individuals follow the majority, often irrationally, in decision-making processes, potentially leading to bubbles or crashes.

Class task:

These biases can result in misvalued securities, opening up possibilities for statistical arbitrage that don’t inherently come with proportional risk increases.

Reflect on these biases and contemplate additional psychological factors and irrational behaviors that might contribute to market irregularities. Can you identify three more instances or examples of such factors or behaviors?



Behavioral biases

Bias/Effect	Definition	Example
Anchoring		
Ambiguity Aversion		
Availability Bias		

Behavioral biases

Bias/Effect	Definition	Example
Anchoring	Overreliance on initial information when making decisions.	Investors may set a target price for buying a stock and fail to reevaluate, missing out on potentially good investments.
Ambiguity Aversion		
Availability Bias		

Behavioral biases

Bias/Effect	Definition	Example
Anchoring	Overreliance on initial information when making decisions.	Investors may set a target price for buying a stock and fail to reevaluate, missing out on potentially good investments.
Ambiguity Aversion	Preference for familiar over unfamiliar investments.	Investors often allocate most of their investments to domestic or familiar stocks, undermining diversification.
Availability Bias		

Behavioral biases

Bias/Effect	Definition	Example
Anchoring	Overreliance on initial information when making decisions.	Investors may set a target price for buying a stock and fail to reevaluate, missing out on potentially good investments.
Ambiguity Aversion	Preference for familiar over unfamiliar investments.	Investors often allocate most of their investments to domestic or familiar stocks, undermining diversification.
Availability Bias	Overestimating the likelihood of events based on their memorability.	A trader may wrongly assume the stock market rallies every time a war starts, based on a single past experience.

Behavioral biases

Bias/Effect	Definition	Example
Confirmation Bias		
Disposition Effect		
Endowment Effect		

Behavioral biases

Bias/Effect	**Definition**	**Example**
Confirmation Bias	Prioritizing information that confirms pre-existing beliefs.	Analysts who favor a particular stock may be more likely to upgrade it after good news, leading to overwhelmingly positive stock ratings.
Disposition Effect		
Endowment Effect		

Behavioral biases

Bias/Effect	Definition	Example
Confirmation Bias	Prioritizing information that confirms pre-existing beliefs.	Analysts who favor a particular stock may be more likely to upgrade it after good news, leading to overwhelmingly positive stock ratings.
Disposition Effect	Tendency to hold losing investments too long and sell winning ones too early.	Investors may continue to hold a poorly performing stock in the hope it will rebound, leading to greater losses.
Endowment Effect		

Behavioral biases

Bias/Effect	Definition	Example
Confirmation Bias	Prioritizing information that confirms pre-existing beliefs.	Analysts who favor a particular stock may be more likely to upgrade it after good news, leading to overwhelmingly positive stock ratings.
Disposition Effect	Tendency to hold losing investments too long and sell winning ones too early.	Investors may continue to hold a poorly performing stock in the hope it will rebound, leading to greater losses.
Endowment Effect	Overvaluing something because one owns it.	Investors may hold onto overvalued stocks they own, missing out on better opportunities elsewhere.

Behavioral biases

Bias/Effect	Definition	Example
Escalation Bias		
Herding Mentality		
Crowding		

Behavioral biases

Bias/Effect	Definition	Example
Escalation Bias	Investing more in a poor asset, amplifying potential losses.	An investor may buy more of a poorly performing stock, thinking it's now a "steal," only to compound losses.
Herding Mentality		
Crowding		

Behavioral biases

Bias/Effect	Definition	Example
Escalation Bias	Investing more in a poor asset, amplifying potential losses.	An investor may buy more of a poorly performing stock, thinking it's now a "steal," only to compound losses.
Herding Mentality	Following the investment decisions of others.	Herding can lead to asset bubbles or a lack of diversity in portfolios, increasing risk.
Crowding		

Behavioral biases

Bias/Effect	Definition	Example
Escalation Bias	Investing more in a poor asset, amplifying potential losses.	An investor may buy more of a poorly performing stock, thinking it's now a "steal," only to compound losses.
Herding Mentality	Following the investment decisions of others.	Herding can lead to asset bubbles or a lack of diversity in portfolios, increasing risk.
Crowding	A concentration of similar trades that squeezes liquidity and skews risk assessment.	A flood of investors into a specific asset can lead to a collapse in prices and a mismeasurement of investment risk.

Behavioral biases

Bias/Effect	Definition	Example
Illusion of Knowledge		
Narrow Framing		
Overconfidence		

Behavioral biases

Bias/Effect	Definition	Example
Illusion of Knowledge	Mistaking data quantity for quality.	Investors may assume that having more data means they understand a stock better, leading to ill-informed decisions.
Narrow Framing		
Overconfidence		

Behavioral biases

Bias/Effect	Definition	Example
Illusion of Knowledge	Mistaking data quantity for quality.	Investors may assume that having more data means they understand a stock better, leading to ill-informed decisions.
Narrow Framing	Evaluating investments in isolation rather than as part of a portfolio.	Investors may treat each investment as separate "mental pockets," ignoring the overall risk and return profile of their portfolio.
Overconfidence		

Behavioral biases

Bias/Effect	Definition	Example
Illusion of Knowledge	Mistaking data quantity for quality.	Investors may assume that having more data means they understand a stock better, leading to ill-informed decisions.
Narrow Framing	Evaluating investments in isolation rather than as part of a portfolio.	Investors may treat each investment as separate "mental pockets," ignoring the overall risk and return profile of their portfolio.
Overconfidence	Overestimating one's ability to predict outcomes.	Investors might trade too frequently, believing they can outsmart the market, incurring more transaction costs and reduced returns.

Behavioral biases

Bias/Effect	Definition	Example
Self-attribution Bias		
Hindsight Bias		
Probability Distortion		

Behavioral biases

Bias/Effect	Definition	Example
Self-attribution Bias	Attributing successes to oneself and failures to external factors.	Investors may blame external factors like the economy for losses while taking credit for gains.
Hindsight Bias		
Probability Distortion		

Behavioral biases

Bias/Effect	Definition	Example
Self-attribution Bias	Attributing successes to oneself and failures to external factors.	Investors may blame external factors like the economy for losses while taking credit for gains.
Hindsight Bias	Believing one knew how a situation would turn out after the fact.	After an event like a market crash, investors may claim they "knew it would happen," even if they didn't act on this supposed knowledge.
Probability Distortion		

Behavioral biases

Bias/Effect	Definition	Example
Self-attribution Bias	Attributing successes to oneself and failures to external factors.	Investors may blame external factors like the economy for losses while taking credit for gains.
Hindsight Bias	Believing one knew how a situation would turn out after the fact.	After an event like a market crash, investors may claim they "knew it would happen," even if they didn't act on this supposed knowledge.
Probability Distortion	Misjudging the likelihood of events based on subjective beliefs.	Investors may overestimate the probability of bad outcomes, like during the COVID-19 pandemic.

Behavioral biases

Bias/Effect	Definition	Example
Recency Effect		
Regret Aversion		
Anticipated Regret		

Behavioral biases

Bias/Effect	Definition	Example
Recency Effect	Being overly influenced by recent events.	Investors may avoid markets after a recent downturn, only to re-enter when conditions improve.
Regret Aversion		
Anticipated Regret		

Behavioral biases

Bias/Effect	Definition	Example
Recency Effect	Being overly influenced by recent events.	Investors may avoid markets after a recent downturn, only to re-enter when conditions improve.
Regret Aversion	Avoiding decisions that could lead to regret, even if they are the better option.	An investor may buy an overvalued stock out of fear of missing out, even if it's not the best decision.
Anticipated Regret		

Behavioral biases

Bias/Effect	Definition	Example
Recency Effect	Being overly influenced by recent events.	Investors may avoid markets after a recent downturn, only to re-enter when conditions improve.
Regret Aversion	Avoiding decisions that could lead to regret, even if they are the better option.	An investor may buy an overvalued stock out of fear of missing out, even if it's not the best decision.
Anticipated Regret	Acting to minimize future regret.	An investor may buy popular stocks like Tesla or Amazon to avoid regretting not having bought them later.

Behavioral biases

Bias/Effect	Definition	Example
Representativeness		
Saliency		

Behavioral biases

Bias/Effect	Definition	Example
Representativeness	Using heuristics or rules of thumb to make decisions.	An investor may invest heavily in a stock due to one positive earnings announcement, without considering market expectations.
Saliency		

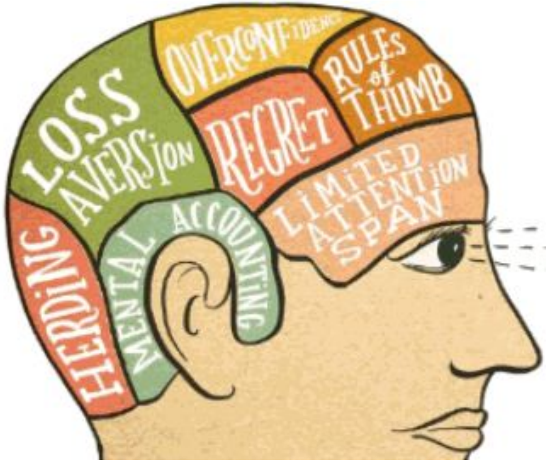
Behavioral biases

Bias/Effect	Definition	Example
Representativeness	Using heuristics or rules of thumb to make decisions.	An investor may invest heavily in a stock due to one positive earnings announcement, without considering market expectations.
Saliency	Focusing on attention-grabbing factors at the expense of less noticeable ones.	Investors may be drawn to high-profile stocks, neglecting less glamorous but potentially more profitable options.

Behavioral biases: implications

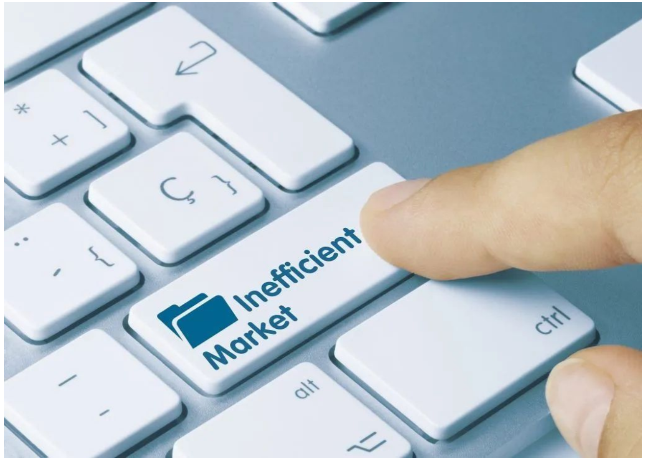
Implications for investment strategy:

- Behavioral biases in human decision-making can cause price deviations from "fair value."
- Arbitrage opportunities may not fully correct these market biases.
- Utilizing portfolio management techniques that target these biases could yield above-average risk-adjusted returns.



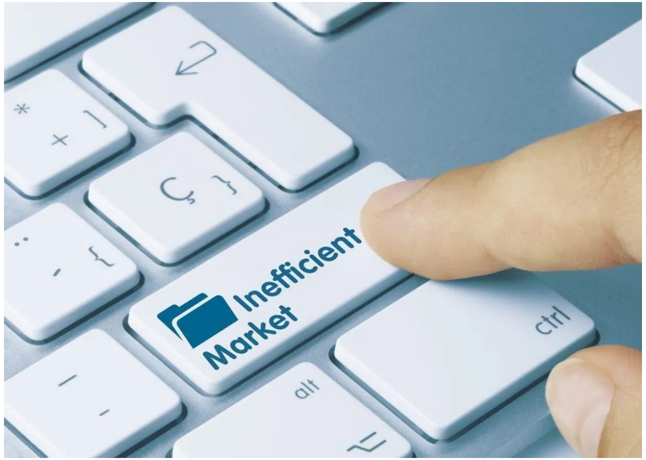
Reasons for Market Inefficiencies

- **High Cost of Information**: Not all participants are willing or able to invest in acquiring valuable information.



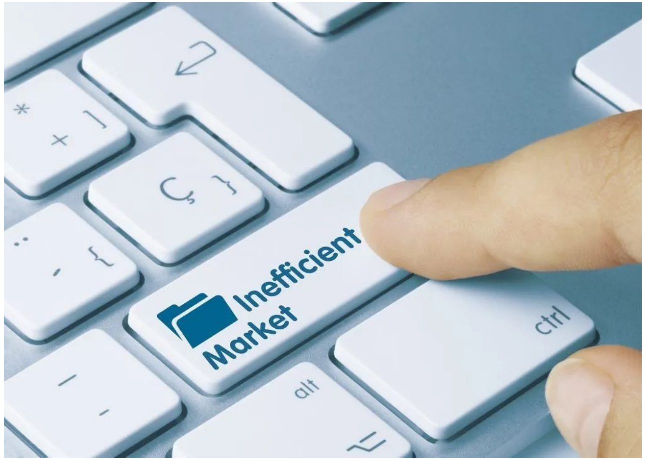
Reasons for Market Inefficiencies

- **High Cost of Information**: Not all participants are willing or able to invest in acquiring valuable information.
- **Delayed Information Diffusion**: Even publicly available information doesn't disseminate instantly across the market.



Reasons for Market Inefficiencies

- **High Cost of Information**: Not all participants are willing or able to invest in acquiring valuable information.
- **Delayed Information Diffusion:** Even publicly available information doesn't disseminate instantly across the market.
- **Limited Information Processing Capability**: Not every investor has the skill to analyze and interpret large volumes of, particularly quantitative, information.

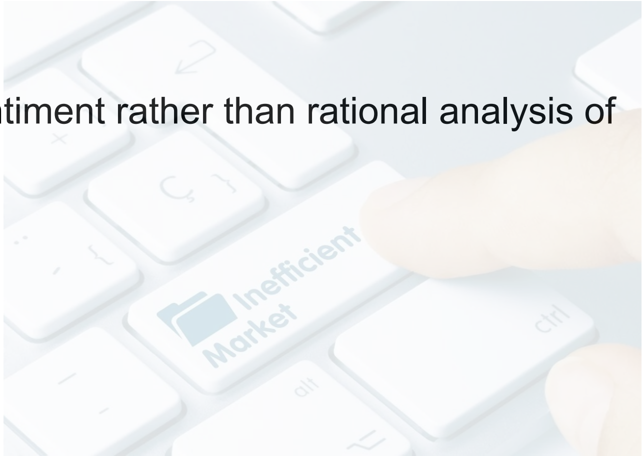


Reasons for Market Inefficiencies

- **High Cost of Information**: Not all participants are willing or able to invest in acquiring valuable information.
- **Delayed Information Diffusion**: Even publicly available information doesn't disseminate instantly across the market.
- **Limited Information Processing Capability**: Not every investor has the skill to analyze and interpret large volumes of, particularly quantitative, information.
- **Emergence of 'Private' Information**: Some individuals may filter public information in a way that effectively turns it into non-public, or "private," information.

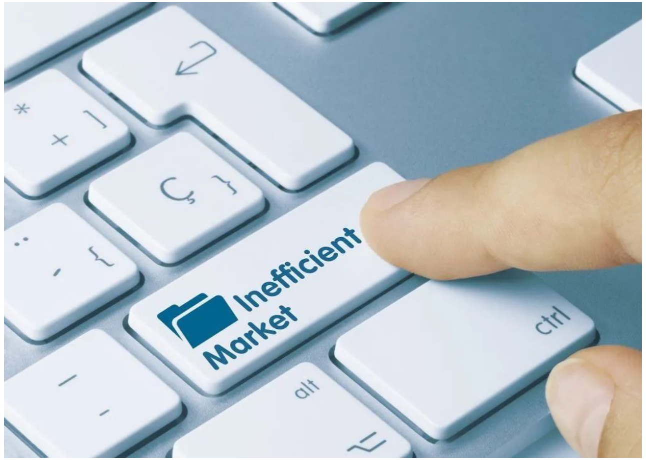
Reasons for Market Inefficiencies

- **High Cost of Information**: Not all participants are willing or able to invest in acquiring valuable information.
- **Delayed Information Diffusion**: Even publicly available information doesn't disseminate instantly across the market.
- **Limited Information Processing Capability**: Not every investor has the skill to analyze and interpret large volumes of, particularly quantitative, information.
- **Emergence of 'Private' Information**: Some individuals may filter public information in a way that effectively turns it into non-public, or "private," information.
- **Emotional Investing**: Certain investors make decisions based on sentiment rather than rational analysis of available data.



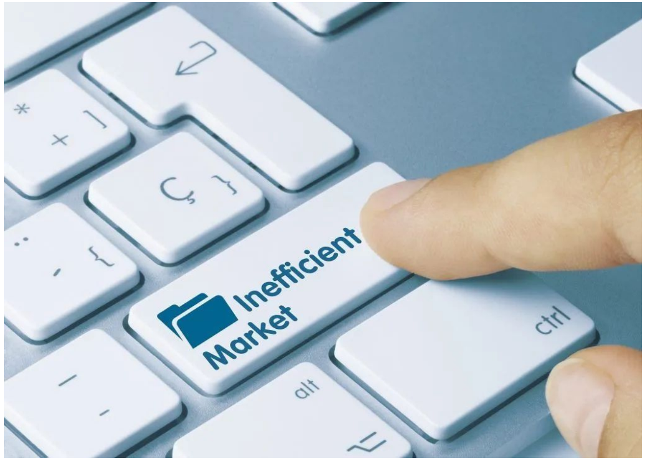
Reasons for Market Inefficiencies

- **Counterproductive Exploits**: Efforts to capitalize on perceived irrationality in the market can sometimes exacerbate inefficiencies.



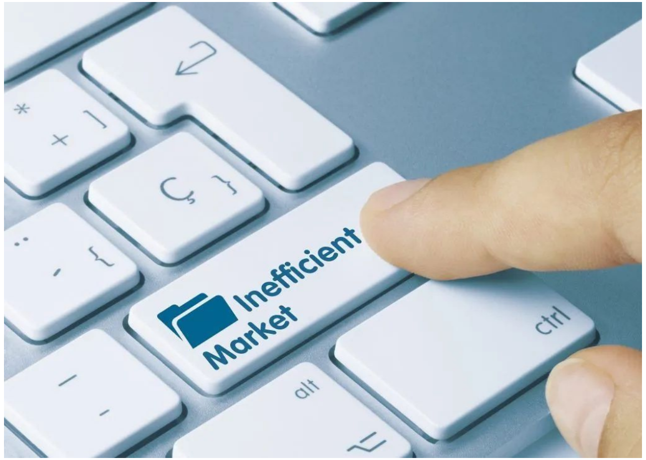
Reasons for Market Inefficiencies

- **Counterproductive Exploits**: Efforts to capitalize on perceived irrationality in the market can sometimes exacerbate inefficiencies.
- **Changing Economic Conditions**: Technological shifts and other economic changes occur constantly, and market participants require time to adapt.



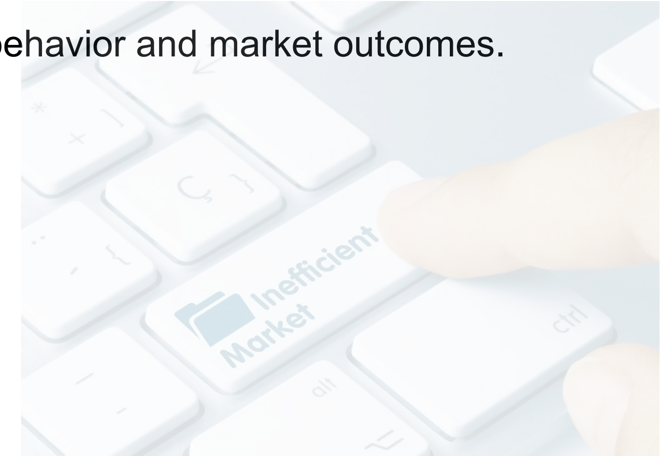
Reasons for Market Inefficiencies

- **Counterproductive Exploits**: Efforts to capitalize on perceived irrationality in the market can sometimes exacerbate inefficiencies.
- **Changing Economic Conditions**: Technological shifts and other economic changes occur constantly, and market participants require time to adapt.
- **Transaction Costs**: These introduce discrepancies between theoretical economic models and real-world market behavior.



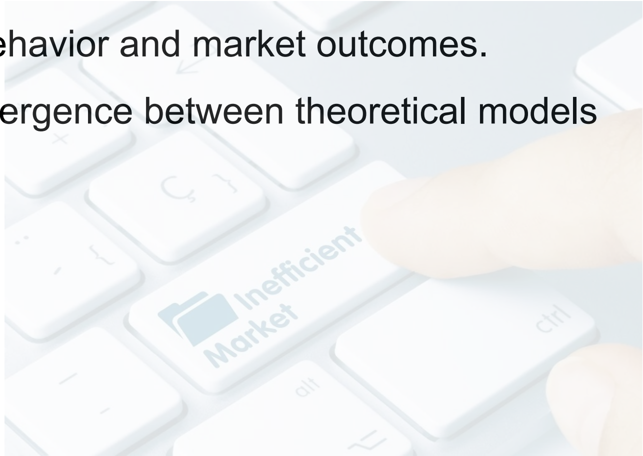
Reasons for Market Inefficiencies

- **Counterproductive Exploits**: Efforts to capitalize on perceived irrationality in the market can sometimes exacerbate inefficiencies.
- **Changing Economic Conditions**: Technological shifts and other economic changes occur constantly, and market participants require time to adapt.
- **Transaction Costs**: These introduce discrepancies between theoretical economic models and real-world market behavior.
- **Tax-Induced Distortions**: The presence of taxes can alter investor behavior and market outcomes.



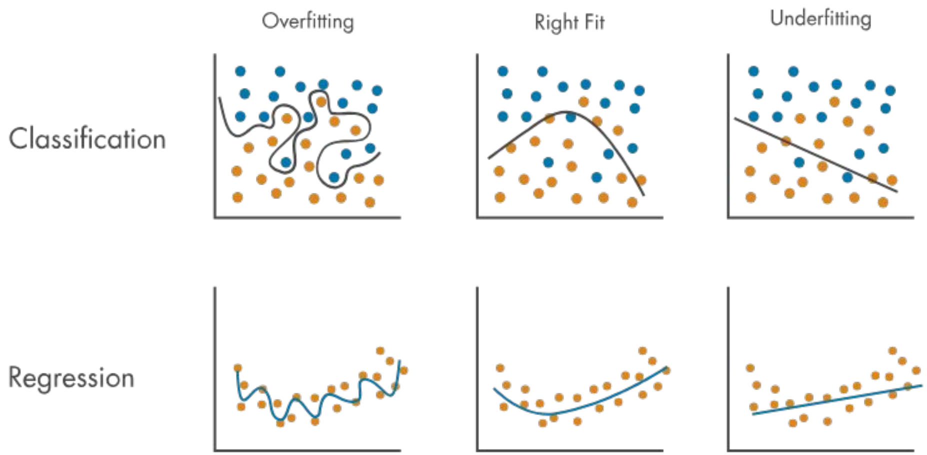
Reasons for Market Inefficiencies

- **Counterproductive Exploits**: Efforts to capitalize on perceived irrationality in the market can sometimes exacerbate inefficiencies.
- **Changing Economic Conditions**: Technological shifts and other economic changes occur constantly, and market participants require time to adapt.
- **Transaction Costs**: These introduce discrepancies between theoretical economic models and real-world market behavior.
- **Tax-Induced Distortions**: The presence of taxes can alter investor behavior and market outcomes.
- **Regulatory Impact**: Government rules and oversight can create a divergence between theoretical models and actual market dynamics.



Statistical Challenges in Analysis

- **Data Mining Risks**: The potential for overfitting models due to excessive testing on the same dataset.



Statistical Challenges in Analysis

- **Data Mining Risks**: The potential for overfitting models due to excessive testing on the same dataset.

    - **Data mining** typically involves exhaustively testing multiple strategies on historical data until a seemingly effective one is found. Often, portfolio managers may use statistical methods to assess the validity of this "successful" strategy, while disregarding the impact of previous tests. This approach can introduce significant bias, undermining the reliability of insights gained from the historical data.

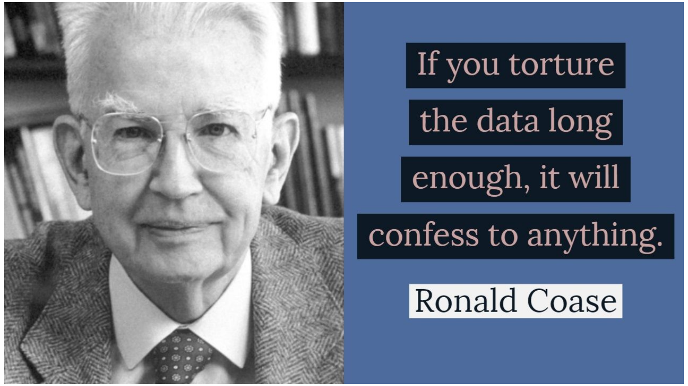
Statistical Challenges in Analysis

- **Data Mining Risks**: The potential for overfitting models due to excessive testing on the same dataset.

  - **Data mining** typically involves exhaustively testing multiple strategies on historical data until a seemingly effective one is found. Often, portfolio managers may use statistical methods to assess the validity of this "successful" strategy, while disregarding the impact of previous tests. This approach can introduce significant bias, undermining the reliability of insights gained from the historical data.
    - **Example**: Data mining entails sifting through historical data to identify patterns that appear successful. For instance, one might analyze 100 months of stock returns using a model with 99 factors: $r_t = \alpha + \beta_1 f_{1t} + \beta_2 f_{2t} + \ldots + \beta_{99} f_{99t} + \epsilon_t$ with t = 1,...100 (equal number of unknowns and observations). In this way, it's always possible to "data mine" a seemingly effective strategy from past data.

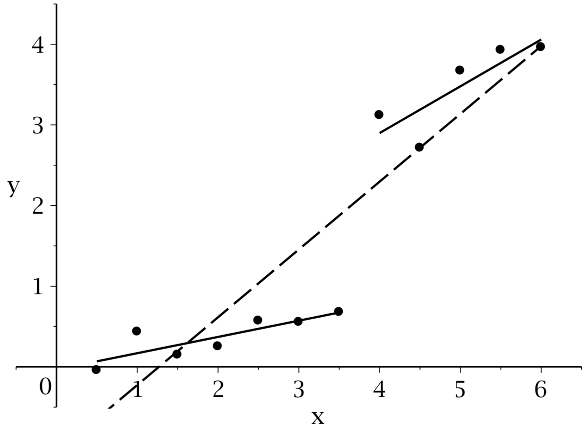
Statistical Challenges in Analysis

- **Data Mining Risks**: The potential for overfitting models due to excessive testing on the same dataset.
  - **Data snooping** is similar to data mining, but it occurs when an individual relies on strategies informed by external sources—such as articles, anecdotes, or the outcomes of others' tests—rather than conducting their own exhaustive analysis.



Statistical Challenges in Analysis

- **Data Mining Risks**: The potential for overfitting models due to excessive testing on the same dataset.
- **Parameter Stability Concerns**: The risk that model parameters may not remain consistent over different time periods.

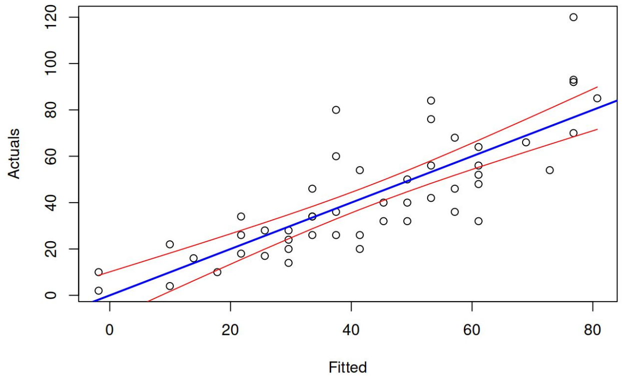


Statistical Challenges in Analysis

- **Data Mining Risks**: The potential for overfitting models due to excessive testing on the same dataset.
- **Parameter Stability Concerns**: The risk that model parameters may not remain consistent over different time periods.
  - **Example**: predicting the performance of Technology stocks based on two factors: interest rates and consumer spending: $r_t = \alpha + \beta_1 \times \text{Interest Rate} + \beta_2 \times \text{Consumer Spending} + \epsilon$.
    Suddenly, there's a significant policy change—perhaps a new tariff on technology imports. The relationship between Technology stocks, interest rates, and consumer spending may now be different than before. If you continue to use the same model with the same parameters you might find that it no longer accurately predicts stock returns. This could result in poor investment choices.

Statistical Challenges in Analysis

- **Data Mining Risks**: The potential for overfitting models due to excessive testing on the same dataset.
- **Parameter Stability Concerns**: The risk that model parameters may not remain consistent over different time periods.
- **Uncertainty in Parameter Estimation**: The likelihood that estimated parameters may not accurately represent the underlying data distribution.

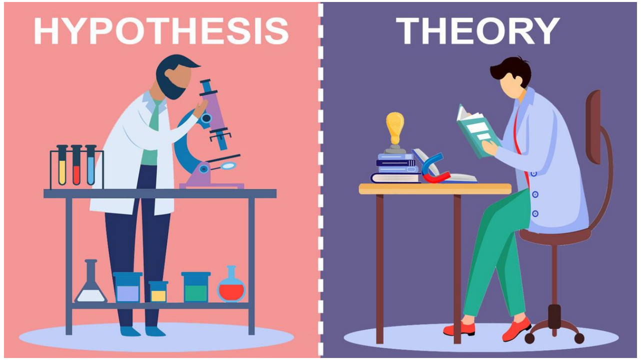


Statistical Challenges in Analysis

- **Data Mining Risks**: The potential for overfitting models due to excessive testing on the same dataset.
- **Parameter Stability Concerns**: The risk that model parameters may not remain consistent over different time periods.
- **Uncertainty in Parameter Estimation**: The likelihood that estimated parameters may not accurately represent the underlying data distribution.
  - **Example**: forecasting monthly sales for a retail store where monthly sales are a function of advertising spend and seasonal factors: $\text{Sales} = \alpha + \beta_1 \times \text{Advertising Spend} + \beta_2 \times \text{Seasonal Factor} + \epsilon$, and you use data from the past 24 months to estimate the model parameters. After applying the model for a few months, you notice that the sales forecasts are consistently overestimating actual sales. Upon reviewing the original data, you discover that the last six months in the dataset had exceptionally high sales due to a temporary market trend, which skewed the parameter estimates.

Statistical Challenges in Analysis: Key takeaways

- **Out-of-Sample Validation**: A successful strategy should demonstrate strong performance not just on the data it was trained on, but also on new, unseen data.
- **Assess Risk-Adjusted Returns**: Raw returns are insufficient for evaluating performance. Always consider the risk associated with those returns to gain meaningful insights.
- **Rule Popularity May Lead to Distortion**: The more widely known a trading rule is, the higher the chance that its efficacy will be compromised over time.
- **Ground in Theory**: Any strategy derived through data mining should be backed by a sound theoretical framework for longer-term viability.



Question/s of the week

Question: We roll a 6-sided die n times. What is the probability that all faces have appeared?

Answer:

Let *P(n)* stand for the probability that all faces have appeared in *n* rolls.

To determine *P(n)*, we can use the **principle of inclusion-exclusion**: We wish to count the number of roll sequences that do not contain all faces. There are 6^n ways to roll a die *n* times. Of these, 5^n have no 1, 5^n have no 2, etc. Simply adding those will not yield what we are looking for, since there are roll sequences that contain no 1 and no 2 (for example), so we would be counting those twice. As a result, we take that sum and subtract all the roll sequences with both no 1 and no 2, or both no 1 and no 3, etc. And so on.

Finally,

$$P(n) = 1 - \frac{\binom{6}{1}5^n - \binom{6}{2}4^n + \binom{6}{3}3^n - \binom{6}{4}2^n + \binom{6}{5}1^n}{6^n}$$

Disclaimer

This course is for educational purposes only and does not offer investment advice or pre-packaged trading algorithms. The views expressed herein are not representative of any affiliated organizations or agencies. The main objective is to explore the specific challenges that arise when applying Data Science and Machine Learning techniques to financial data. Such challenges include, but are not limited to, issues like short historical data, non-stationarity, regime changes, and low signal-to-noise ratios, all of which contribute to the difficulty in achieving consistently robust results. The topics covered aim to provide a framework for making more informed investment decisions through a systematic and scientifically-grounded approach.