Machine Learning in Financial Risk Management

Lecture 1

Spring 2026

Kaila
Aalto University

Poll: Background

- Finance studies
- Financial
- Programming
- Mathematics
- Work experience in finance

presemo.aalto.fi/firma3

Course team

- Ruth Kaila
- Eljas Toepfer

Ruth Kaila

Passing the course

6 lectures and 6 exercise sessions

- Lectures every second Wednesday at 14-15-16.00, **Ruth Kaila**, starting 14.1.2026
- Exercises every second Wednesday at 14-15-16.00, **Eljas Toepfer**, starting 21.1.2026

A 3-credit version, a 5-credit version or a 6-credit version of the course are available.

3 credit version:

- Biweekly work sheets, the fundamental exercises (reflection on the lectures and analysis/solutions of Python exercises done in Jupyter)

grading: **passed/failed**

5 credit version:

- Biweekly work sheets, fundamental exercises (reflection on the lectures and analysis/solutions of Python exercises done in Jupyter)
- Biweekly work sheets, independent exercises (finding your own data and working with this)

grading: **passed/failed**

6 credit version:
same as 5 credit version + essay on a scientific article

grading: **passed/failed**

Biweekly exercises

The instructions for the biweekly exercises are given a couple of days after each lecture. The exercises can be solved individually or in small groups.

The early bird deadline ends 10 days after the exercise has been given. **The final deadline for all the exercises is 16.4.2026**.

To pass the course, the student must complete 6 weekly exercises. However, **if the student returns each exercise within 10 days of the date the exercise was given, 5 exercises are sufficient to pass the course**.

The exercises are each time divided in two groups:

- First, the fundamental exercises, which all should solve. Here the data is provided to you.
- Second, the independent exercises for those, who wish to get 5 or 6 credits. Here you might need to find a data set interesting to you with which to work (we will help you to access data sets).
- The exercises are graded passed/failed

Schedule

14.1 Machine learning in finance, Algorithmic and High Frequency trading

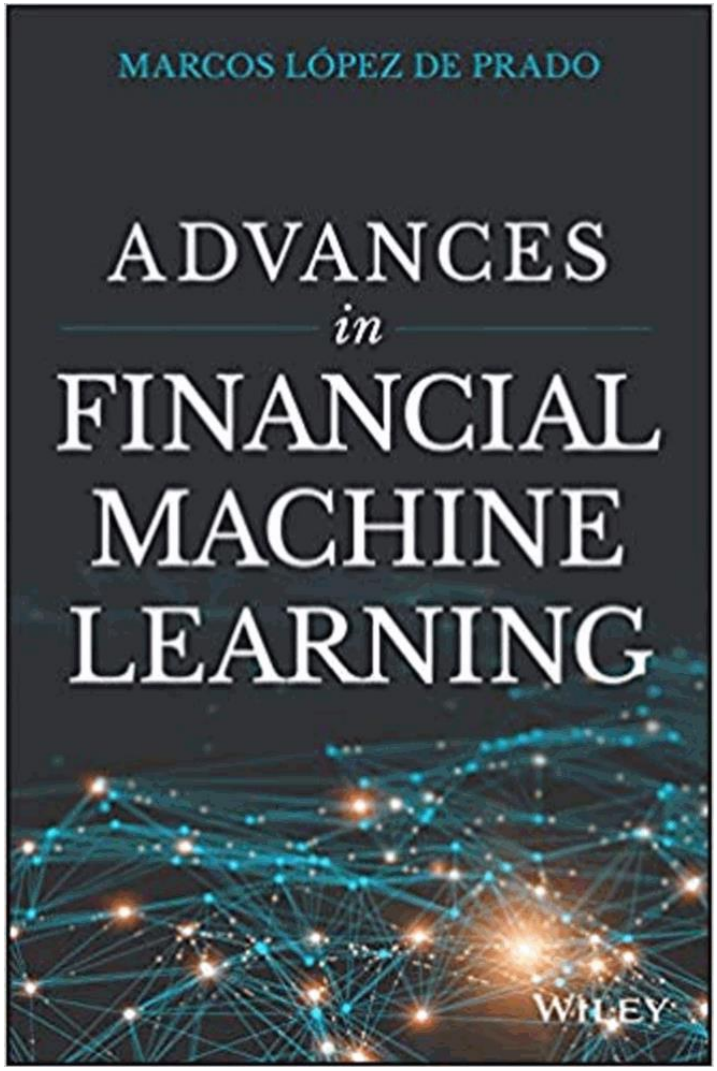
28.1 Structure of data, Classification

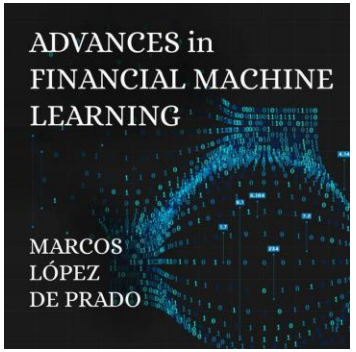
11.2 Overfitting, Regression models, Ensemble methods, Cross-Validation

25.2 Cyber security
Visiting lecture TBA – ***Compulsory***

11.3 Feature importance, Backtesting

25.3 Dimensionality reduction, NLP





Book, e-book (Aalto Learning Center, audiobook)

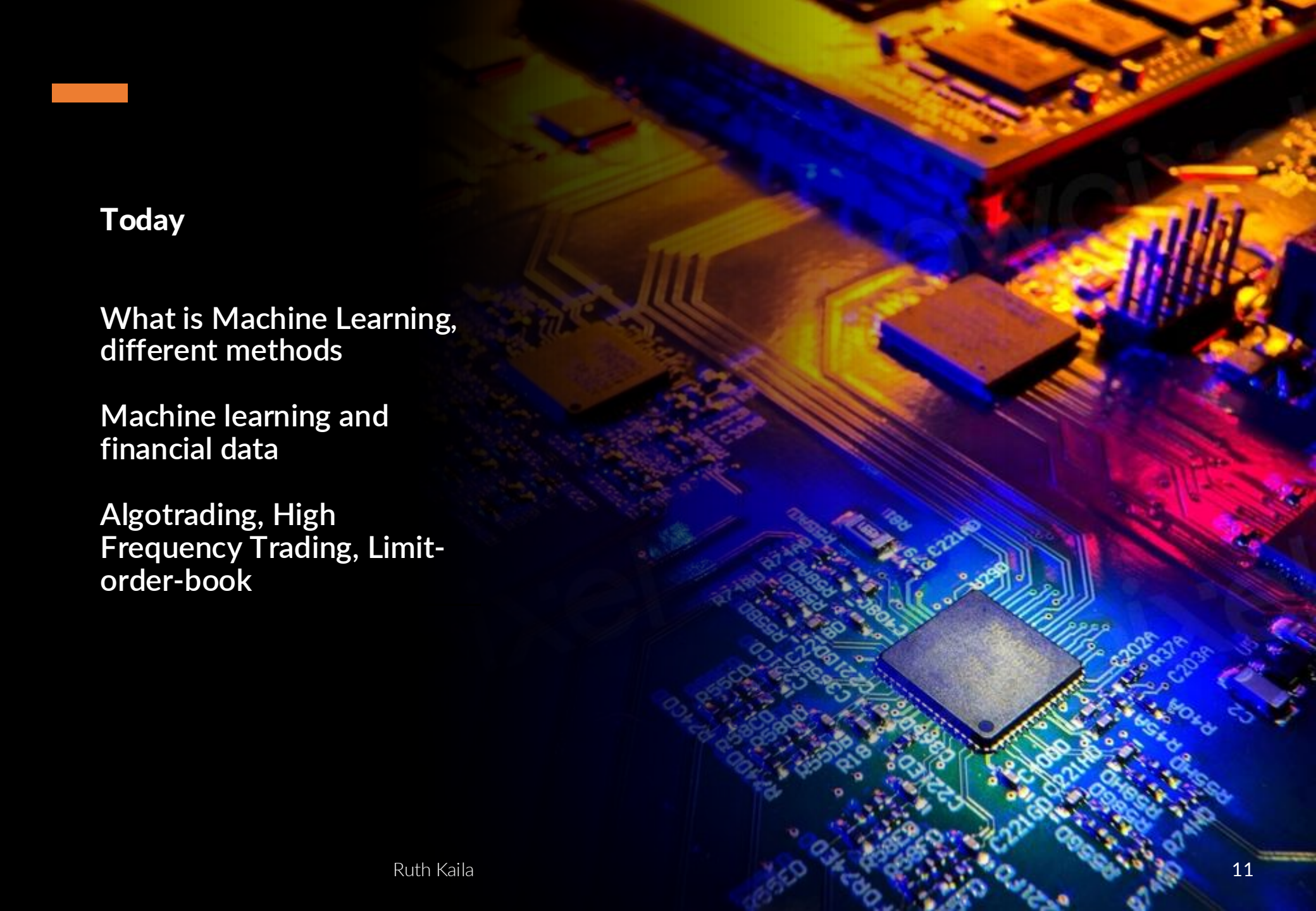
Ruth Kaila

Today

What is Machine Learning, different methods

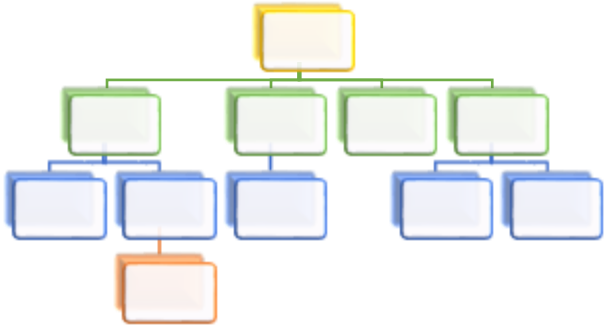
Machine learning and financial data

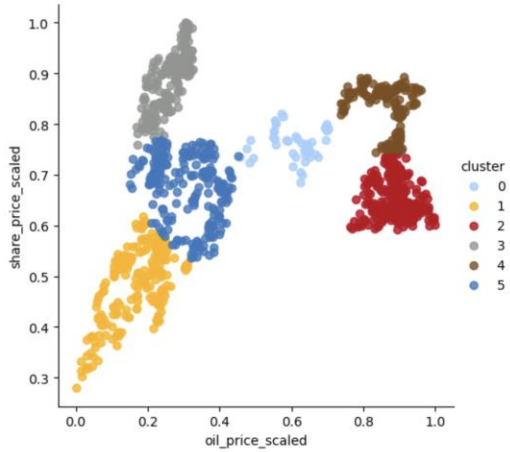
Algotrading, High Frequency Trading, Limit-order-book



The Evolution of Intelligent Systems

Rules-based expert systems designed to mimic the decision-making of a human expert, e.g., Smart houses



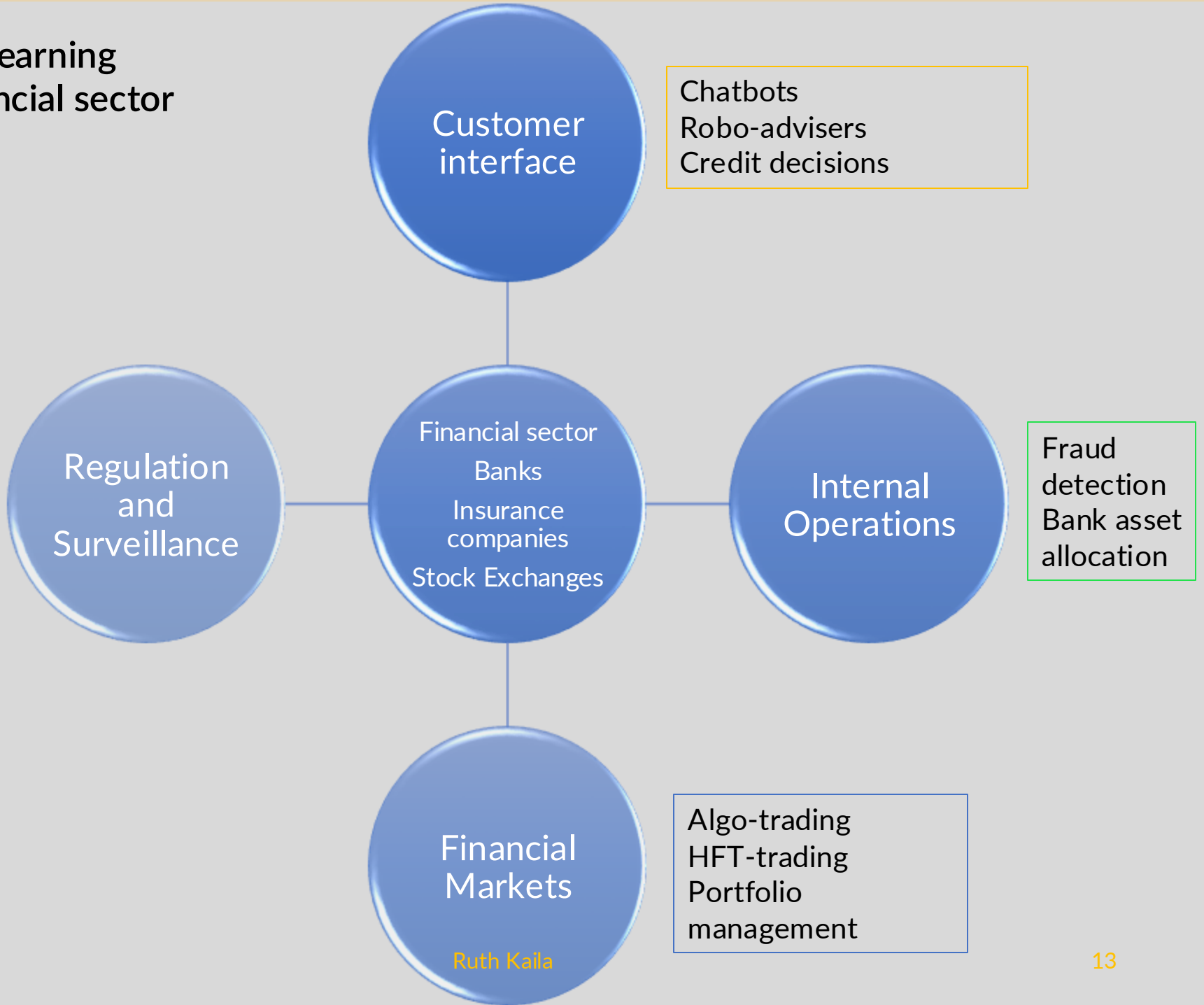


Machine Learning systems learn from data without being explicitly programmed **Deep Learning**

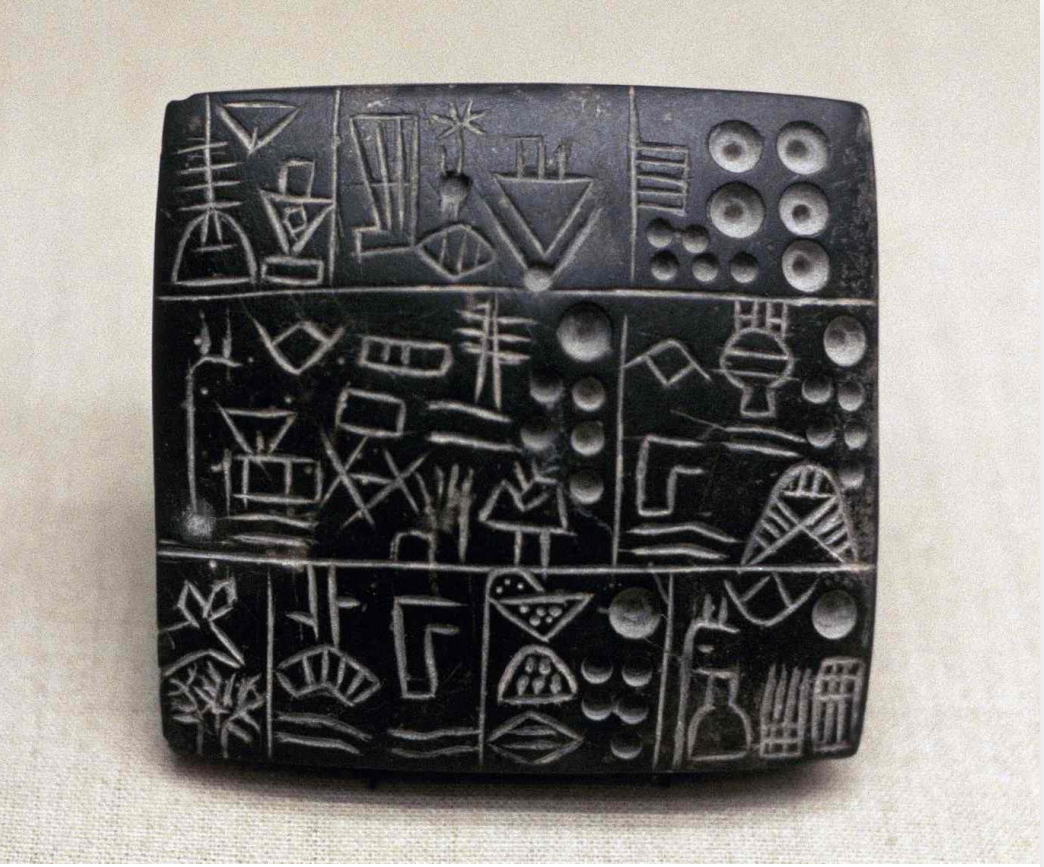
Generative AI models that create new content: text, pictures, music ChatGPT

General AI still theoretical, intellectual capacity

Machine Learning in the financial sector



generates no profit but incurs costs; automating it would be the optimal solution



Picture: Getty Images; Late Uruk period, about 3200 BC
Typically records of trade and labor.

Machine learning

Machine Learning (ML) methods are often simple extensions of traditional statistical methods.

There is no one ML algorithm that gives the best result when applied to different types of data.

Once we have defined the prediction problem, we should choose suitable data and methods. Important is a strong understanding of

- the data available,
- the possible methods,
- *the financial markets.*

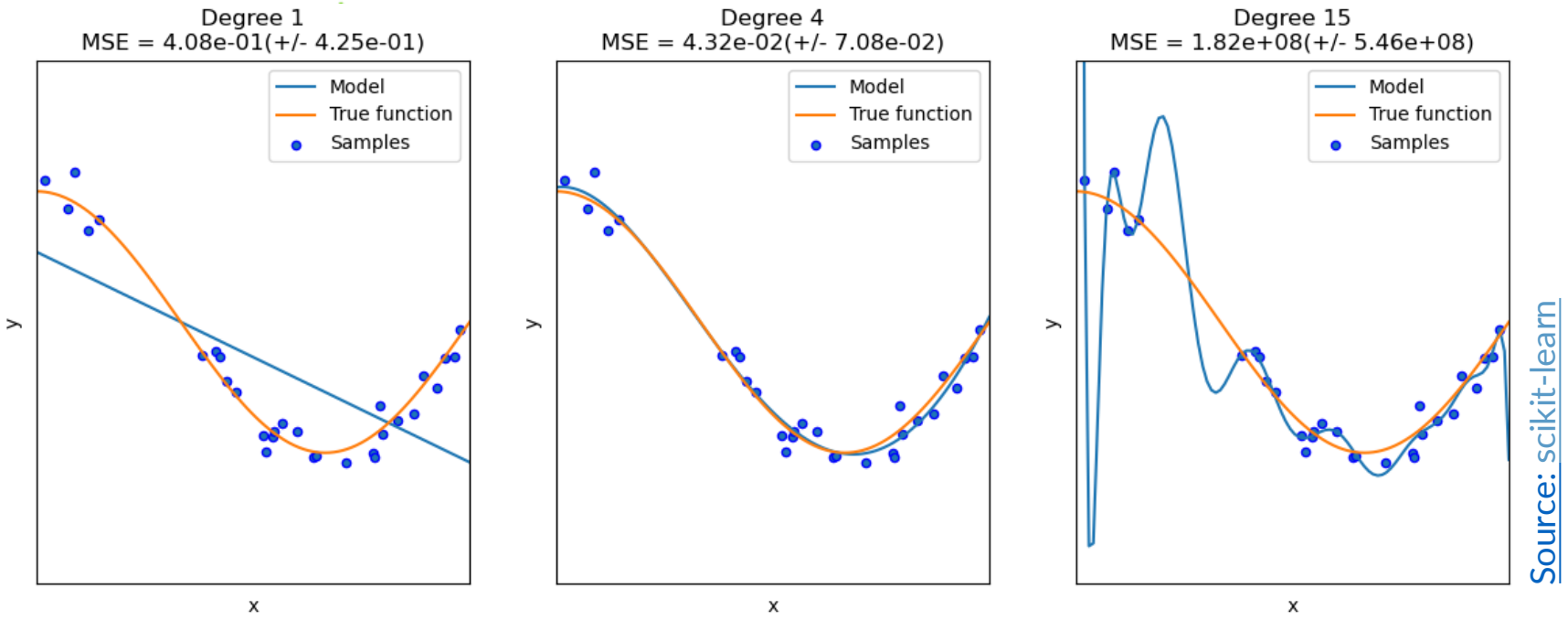
ML methods are very sensitive to over calibration.

Risks

- The datasets might **not contain the information** we are searching (e.g., *alpha*).
- Some signals might be too expensive to be useful, they might have too little investment capacity or may degrade quickly.
- **Overly complex** infrastructure and models—overfitting, instability, wrong relationships and patterns
- Data scientists often lack an understanding of the financial markets and their dynamics.
    - compare quant mathematicians and the financial crisis 2008
- Algorithms can not fully replace human intuition.

Methodology
Overfitting

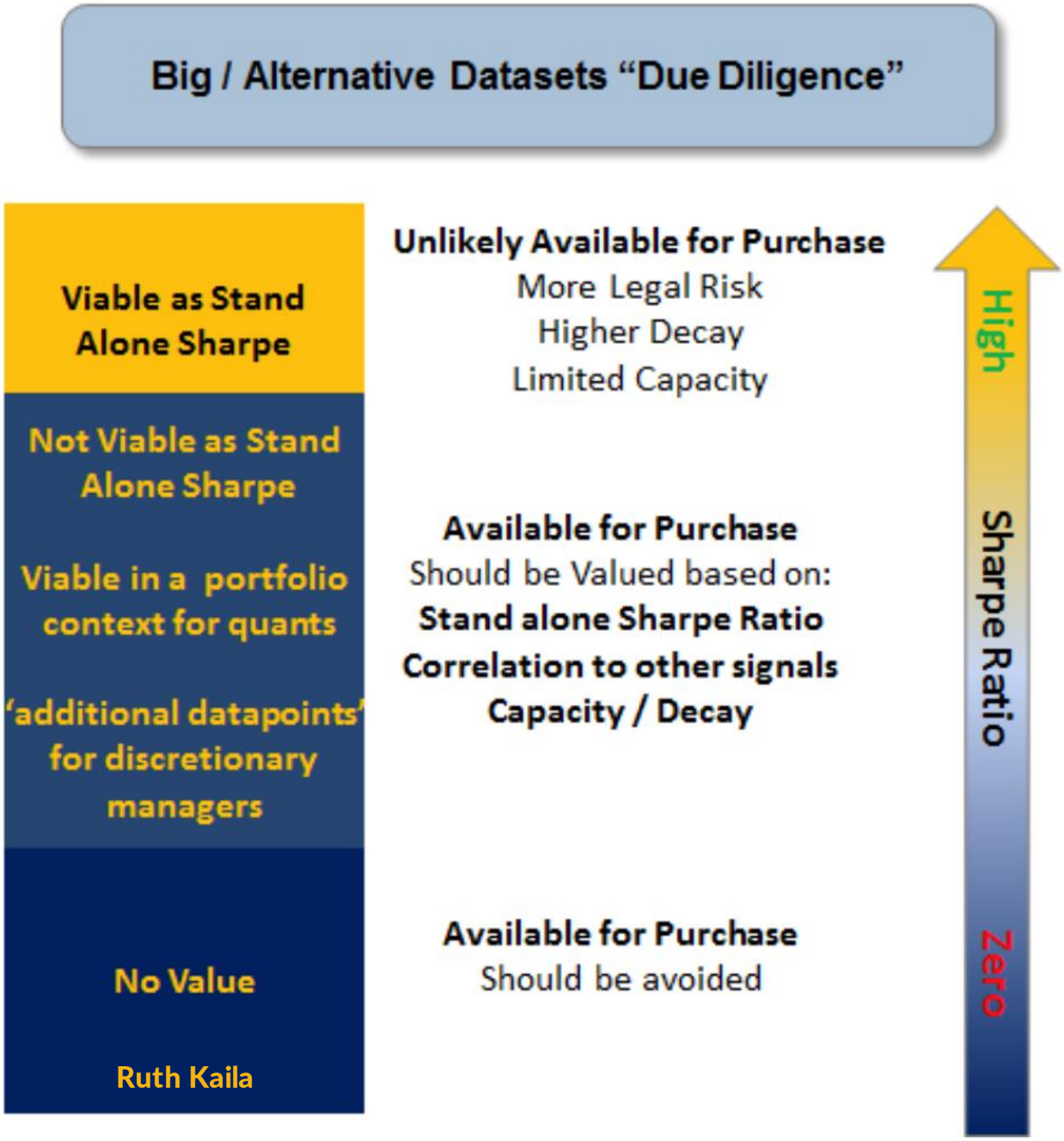
Problem with overfitting: it fits well historical data, but it has poor performance on out-of-sample data forecasting.



Source: scikit-learn

Fitting to historical data

Figure 5: Information content of an alternative data set



It is important to assess the information content of the data.

> *For example, if we try to reach a high alpha (excess return), we should use data containing some relevant information on it.*

Alternative Data used in Financial Markets

Individuals	Business Processes	Sensors
Social Media	Transaction Data	Satellites
News and Reviews	Corporate Data	Geolocation
Web Searches, Personal Data	Government Agencies Data	Other Sensors

Textual form, requires Natural Language Processing.

Often **unstructured** and distributed across multiple platforms.

Often a byproduct of corporate record-keeping such as banking records, supermarket scanner, supply chain data, credit card data.

Often legacy and privacy considerations.

Structured
Can be leading indicators of corporate metrics that are reported at a lower frequency.

Collected mechanically from various devices, unstructured, very large sizes.

Requires analysis techniques.

table: JP Morgan

Process from data acquisition to trades

Process from data acquisition to trades

1. Identify and acquire data

- CSV, API, text, html, streaming
- data managing team

→ directly through data owners or third parties

2. Store, structure, preprocess

- Apache Spark, SQL, cloud
- software engineers

→ big data is rarely available in clean format

3. Analyze data via ML, backtesting, visualization techniques

- data scientists, quantitative researchers

4. Implementation of trade ideas, trading signals, risk analysis

- based on reports, alerts, signals
- traders, portfolio managers

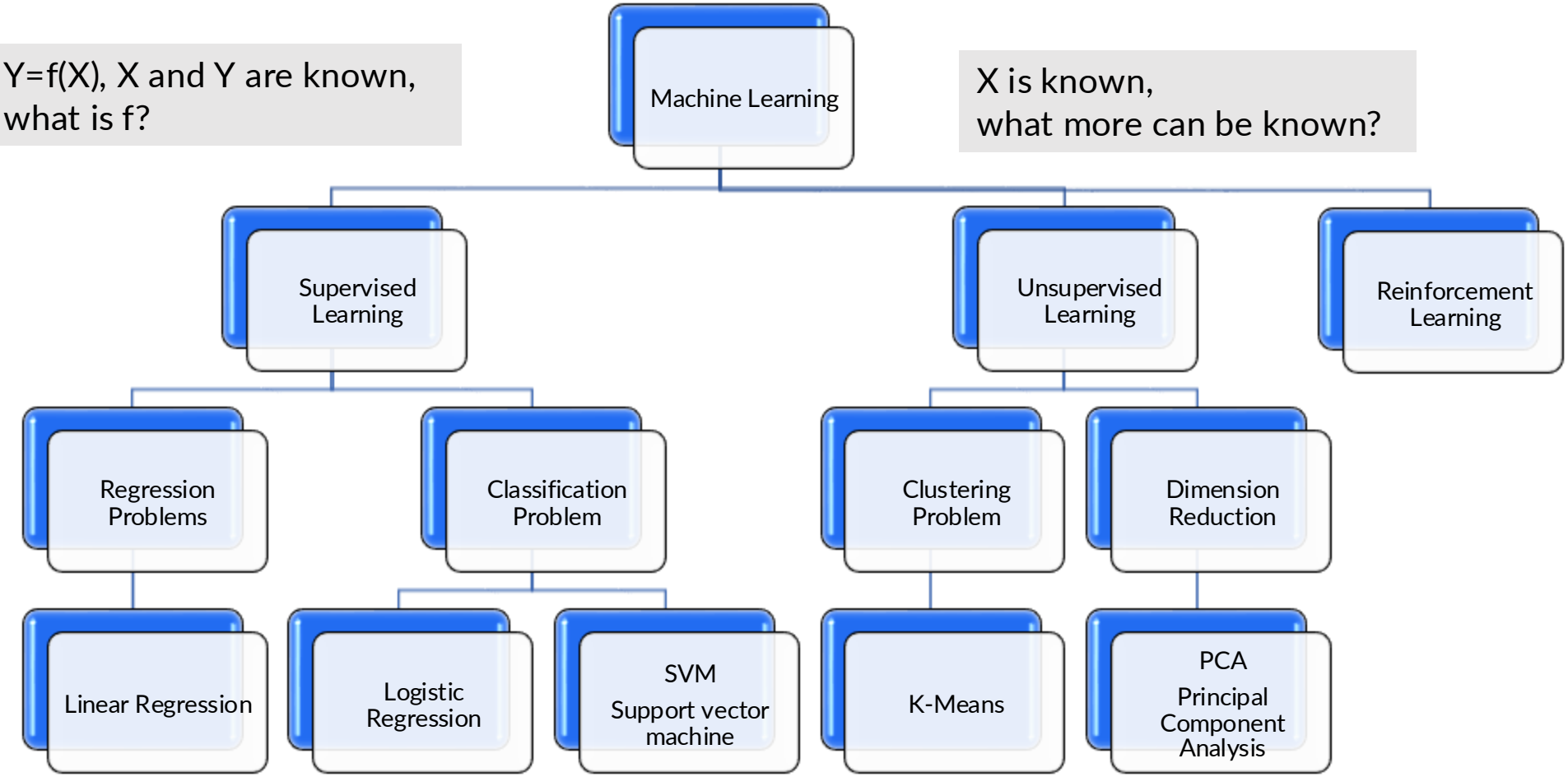
Machine learning is team-work.

Often, it is easier for quants to learn new Machine Learning techniques and tune programs than for data scientists to learn to design good trading strategies.

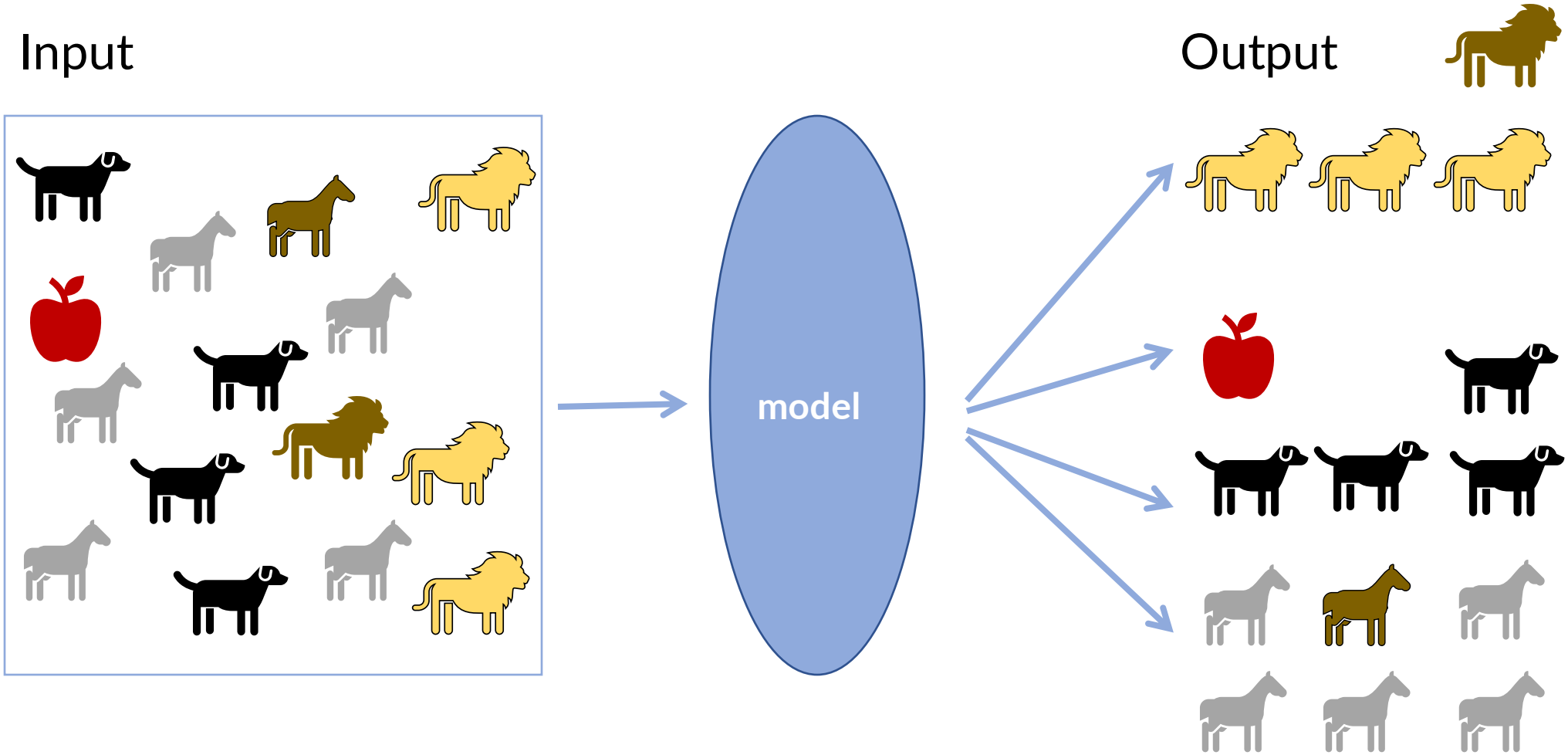
Different types of Machine Learning

Y=f(X), X and Y are known, what is f?

X is known, what more can be known?



Unsupervised learning



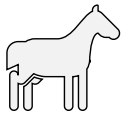
Supervised learning

We train the algorithm with well labeled training data so that the algorithm classifies data or predicts outcomes accurately.

The training set includes inputs and correct outputs, allowing the model to learn over time.

Labeled learning data on animals:
Input and output


cat


horse


horse


horse


lion


cat

Prediction:
This is a horse.

Out-of-sample

Two type of data sets

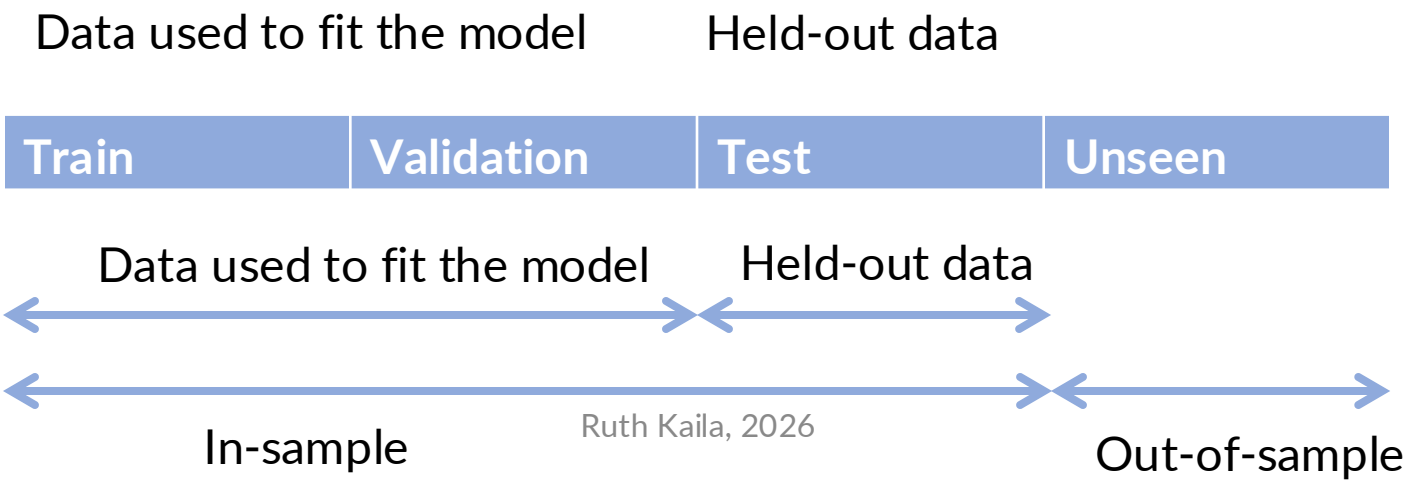
Y=f(X), X and Y are known, what is f?

Training set: (typically 2/3 of the data)

- we select the features (denoted by x)
- we fit model parameters and set the labels (denoted by y)
- Training set errors are estimated from the training set

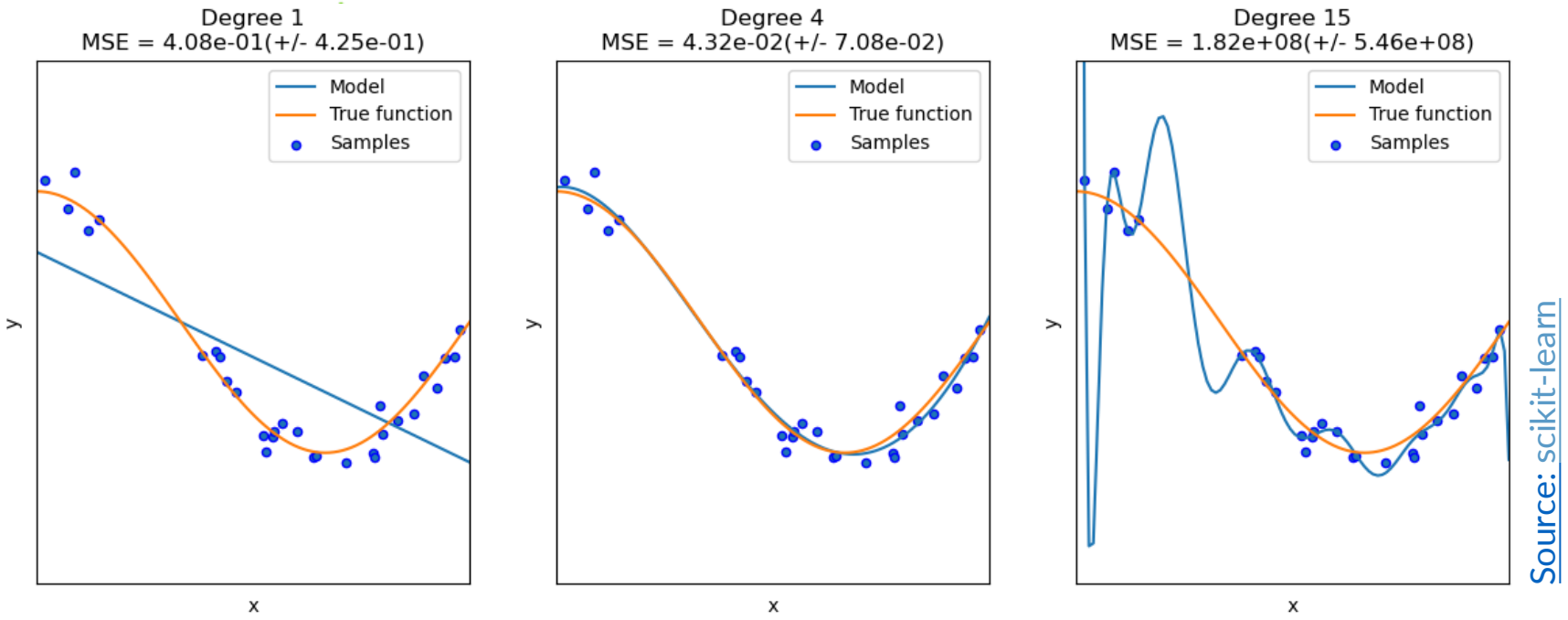
Test set:

- This set is not used for model calibration.
- We test the model's calibration on the test set to ensure it generalizes well to new, unseen data.
- The amount of *overfitting* can be assessed by comparing the model's performance on the test data.



Methodology
Overfitting

Problem with overfitting: it fits well historical data, but it has poor performance on out-of-sample data forecasting.

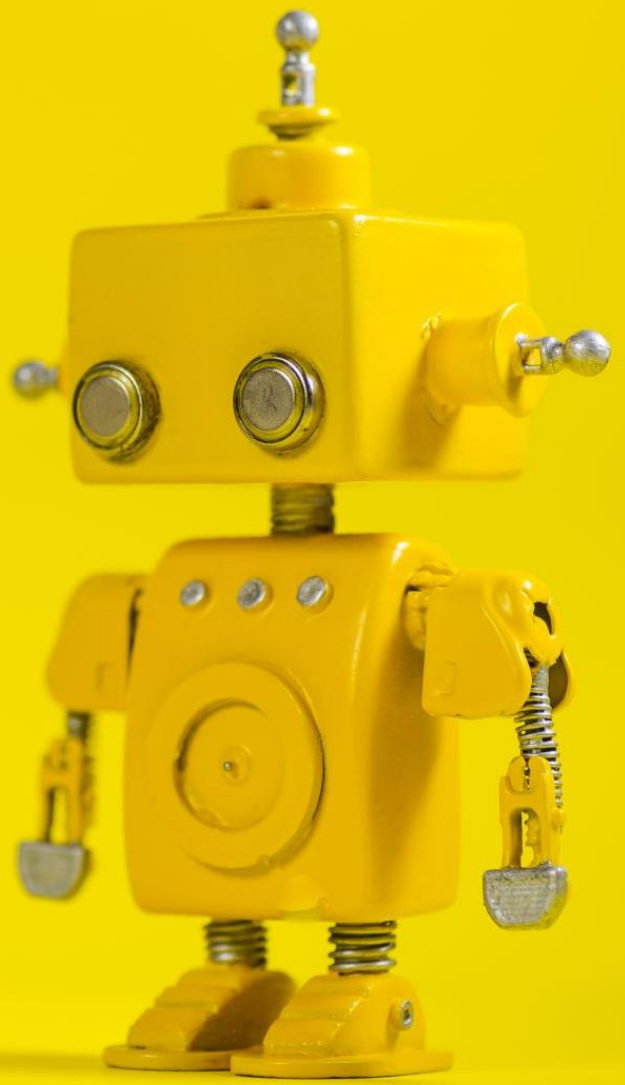


Source: scikit-learn

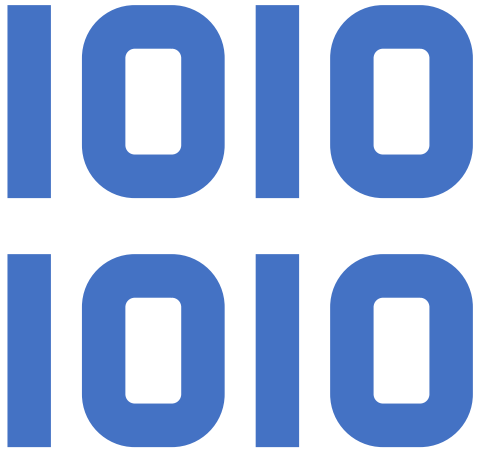
Fitting to historical data

Specialized machine learning courses in Aalto

- Machine Learning CS-C3240
- Machine Learning: Supervised Methods CS-E4710
- Artificial Intelligence CS-E4800
- Machine Learning: Advanced Probabilistic Methods CS-E4820
- Deep Learning CS-E4890
- Gaussian Processes CS-E4895
- Reinforcement learning ELEC-E8125



Ruth Kaila



Introduction to Machine Learning

30 minutes

A Friendly Introduction to Machine Learning

27

Algorithmic trading

Algorithmic trading

Poll

How much of stock trading is algorithmic trading?

- 7 %
- 15 %
- 50 %
- 80 %

Algorithmic trading

The use of programs and computers to generate and execute (large) orders in markets with electronic access.

- Rather than maximize profit, the main objective of algo trading is to **control execution costs and market risk**.

  At least in high-frequency trading, the positions are closed at the end of each day.

- Algorithms started as tools for institutional investors in the beginning of the **1990s**, explosion of algorithmic trading in the beginning of the **2000**.

- 80 % of stock trading is algorithmic trading

(source: https://seekingalpha.com/article/4230982-algo-trading-dominates-80-of-stock-market)

Strategies of algorithmic trading

Name of Algo strategy	Description of strategy
Trade execution algorithms	Designed to **minimize the price impact** of executing trades of large volumes by splitting orders into smaller parcels and slowly releasing these into the market.
Strategy implementation algorithms	Designed to **read real-time market data and formulate trading signals** to be executed by trade execution algorithms.
Stealth/gaming algorithms	Designed **to take advantage of the price movement** caused when large trades are filled, and to detect and outperform other algorithmic strategies.

Source: ASIC 2010

Trade execution algorithms

Institutional clients need to trade large amounts of stocks. These amounts are often larger than what the market can absorb without impacting the price.

Large orders need to be split into smaller orders which will be executed electronically over the course of minutes, hours, day.

- Time frequency: On what time scale are the signals measured?
- Entry signal: What needs to happen so that I enter the market?
- Size: How big positions should be taken?
- Exit signal: What needs to happen so that I exit the market?
- Evaluation benchmark: How is the *performance* measured?

The signals can be related, for example, to the *momentum* or long-term average of the asset, to the volatility of the asset, to news, to twitter feeds etc.

Technical Requirements for Algorithmic Trading — powerful tools

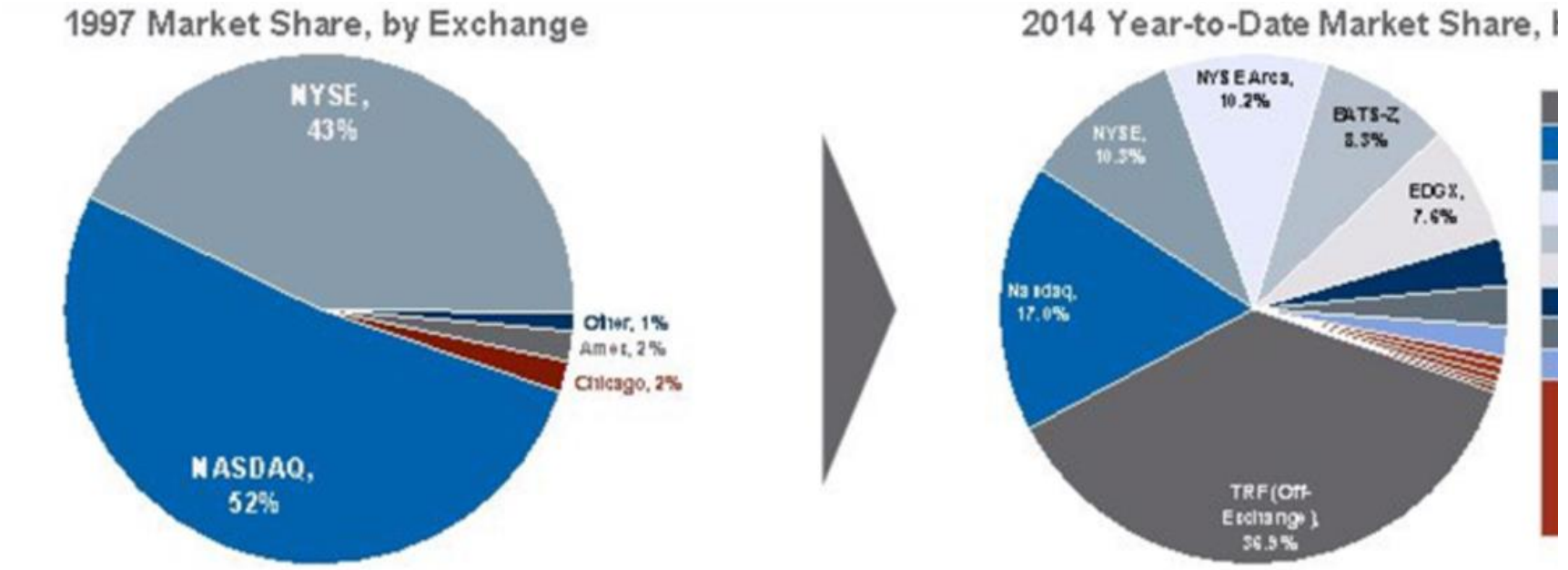
Network connectivity

- Access to ***trading platforms*** for placing the orders
- Access to ***market data*** *feeds*

Trading algorithm and backtesting

- Ability of coding
- Infra, networks
- Historical + real time data

- Being optimized, the trades are executed at the *best possible prices*
- Trades timed correctly and instantly, to avoid significant *price changes*
- *Reduced transaction costs*
- Simultaneous automated checks on *multiple market conditions*
- Risk of *manual errors* in placing the trades is reduced
- *Backtest* the algorithm, based on available historical and real time data
- Reduced possibility of mistakes by human traders based *on emotional and psychological factors*



Picture: valuewalk

Algos help to cope with fragmentation

Fragmented markets (see picture)

- Difficult to get the whole picture of the markets
- Need to search for trading opportunities, compare prices, etc…
- Algos reduce search costs & increase search speed
- More trading opportunities can be identified

(Latency arbitrage for High-frequency trading)

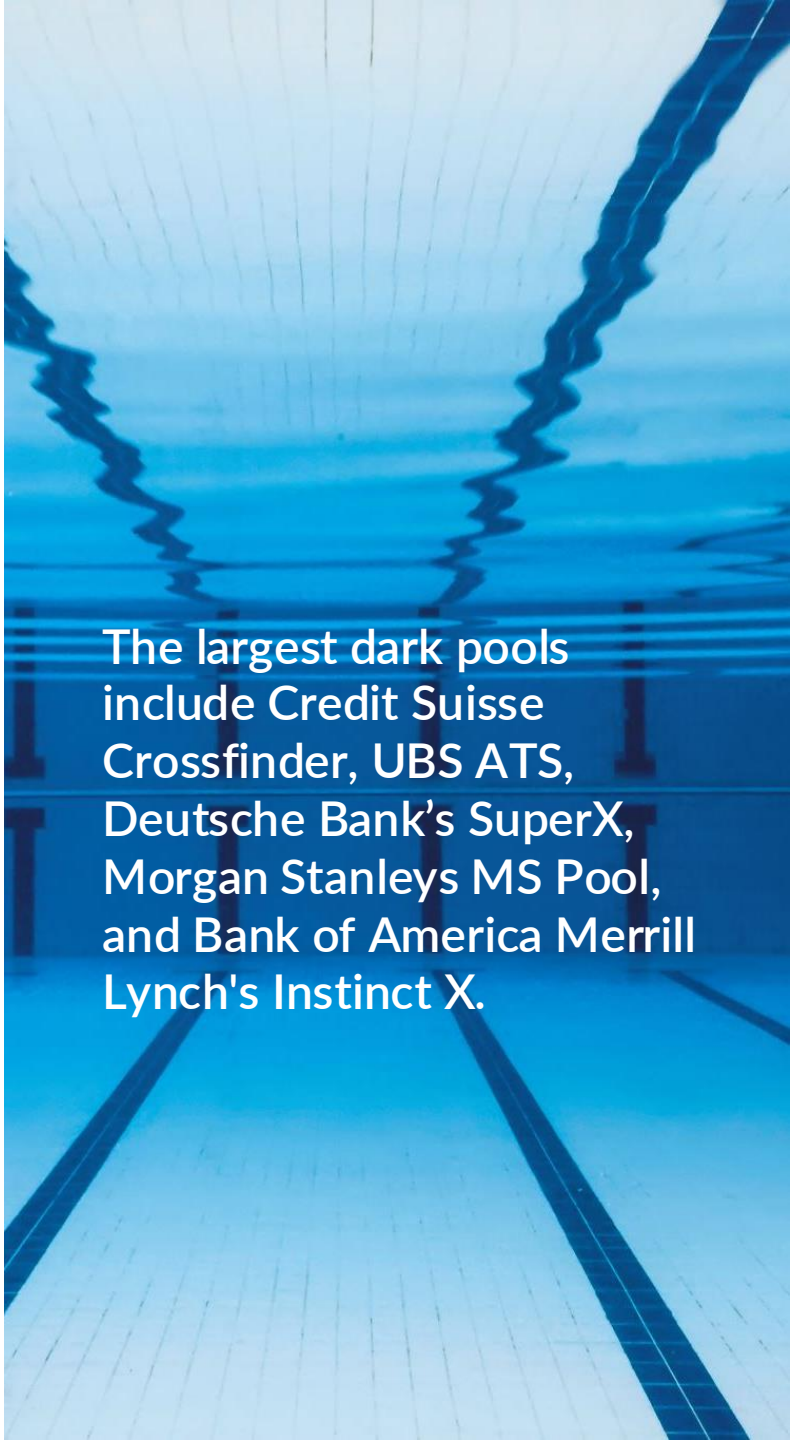
Lit and Dark pools

A Pool - any place where trading takes place.

A Lit Pool – pool where the trading is visible to the public. (e.g., NYSE, Nasdaq)

Dark Pools are alternatives to the exchanges where *price and size of orders are revealed only to participants*

- secret block-trading venues for *institutional traders*
- large sells or buys of an *asset don't affect the asset's market price* (what happens to the price in traditional displayed markets?)
- *lower trading fees* than in the traditional displayed market
- initially used by big *institutional investors to avoid High-frequency HF traders taking profit* from their bids and offers; now days also for HF traders, as dark pools want to grow
- owned by banks, brokers et cetera.
- *risky*: can you trust your dark pool?
- MIFID II Markets in Financial Instruments Directive restricts dark pools in EU



Lit and Dark pools

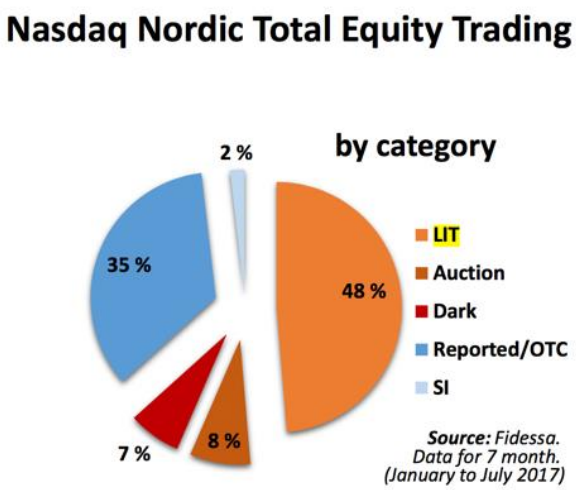
Poll

What percentage of all the trading was done in dark pools in the US in 2017?

1. 3 %
2. 7 %
3. 15 %
4. 40 %

Vote in https://presemo.aalto.fi/firma3

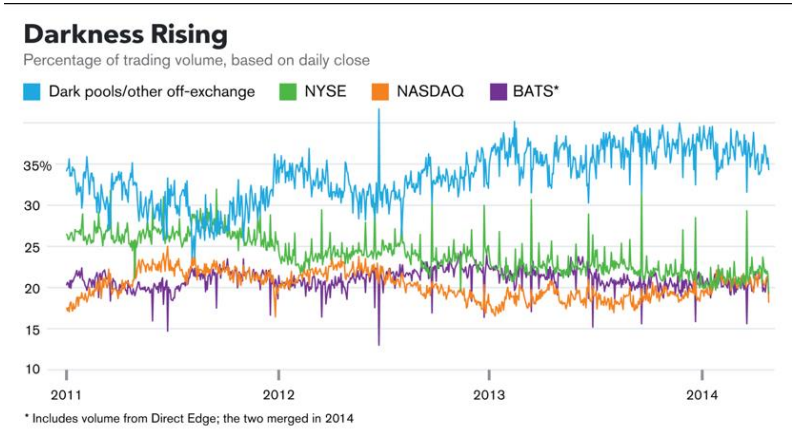
Lit and Dark pools



Source Husman, Nasdaq

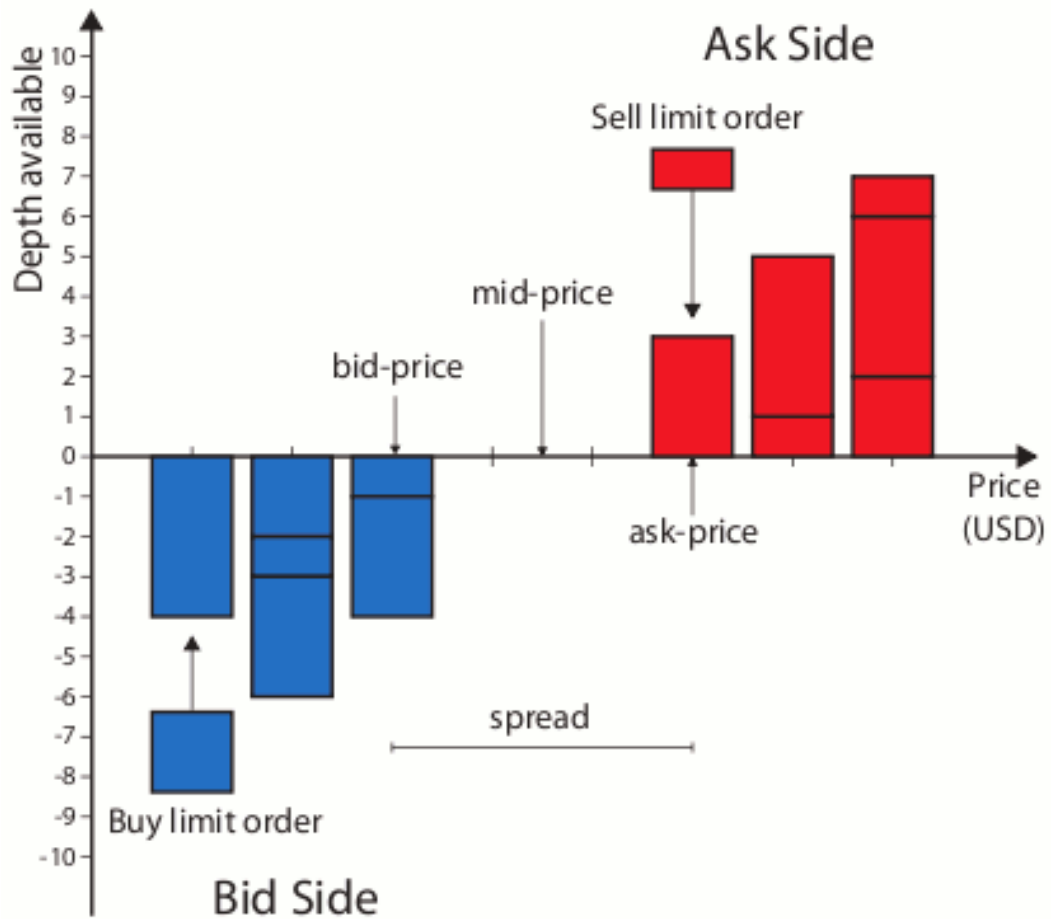
In US:

- Estimates show that dark pools accounted for approximately 40% of all **U.S.** stock trades in spring 2017 compared with an estimated 16% in 2010.
- As of February 2022, nearly 50% of all trading activity took place in dark pools and other off-exchange venues. For specific stocks, such as GameStop, dark pool trading made up more than half of the total trading volume on certain days.



Limit-order-book and Market making

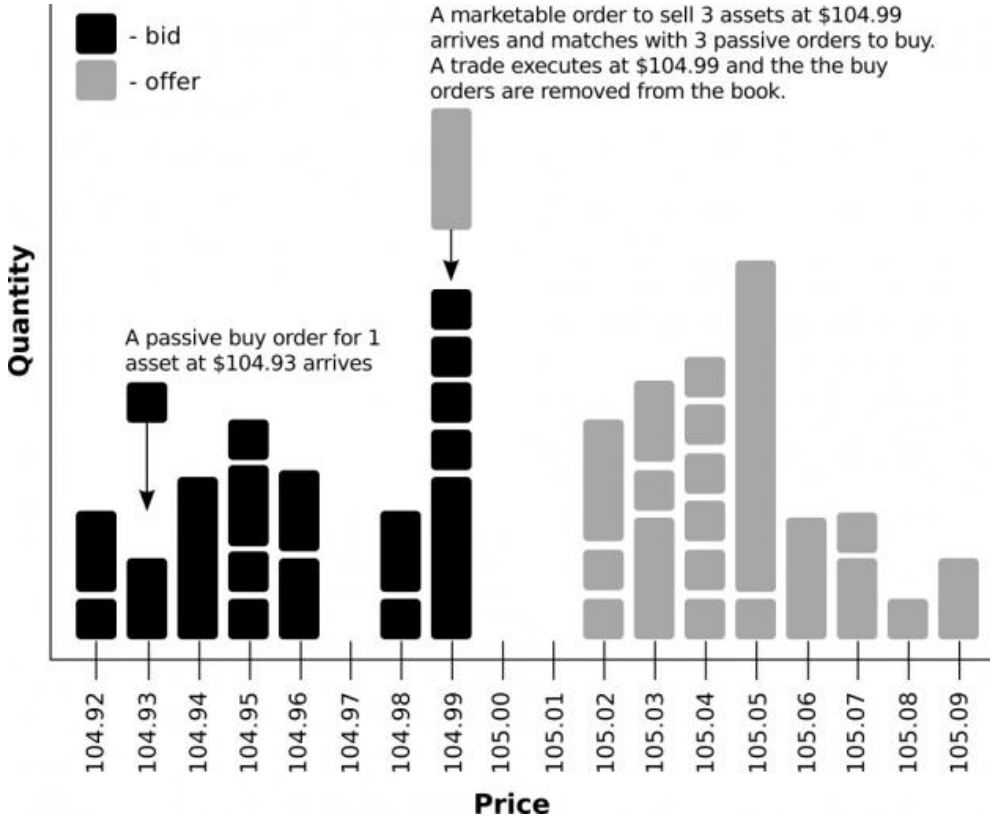
Based on Aldridge, 2013



Limit order book LOB – record of outstanding orders of an exchange

If there is a book order that matches the incoming order, then a trade is executed. The new order is termed aggressive (or marketable) because it initiated the trade, while the existing order from the book is deemed passive.

- The bid price is the highest price a buyer is prepared to pay for a financial instrument
- The ask/offer price is the lowest price a seller will accept for the instrument.



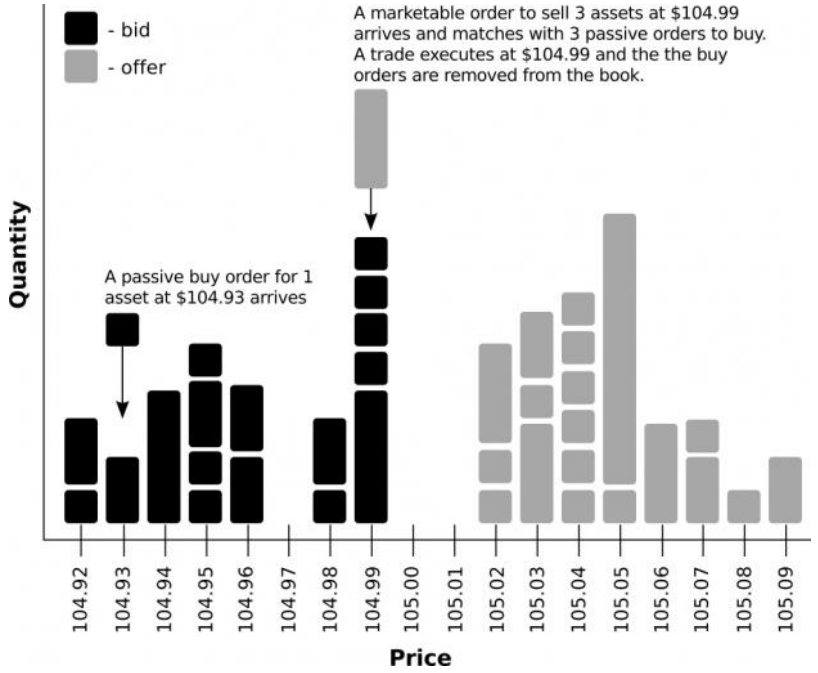
Limit order book LOB, *continued*

For example, let's say there is a stock with a bid price of $10 and an ask price of $11. This means that the highest price a buyer is willing to pay is $10, and the lowest price a seller is willing to accept is $11.

Now, imagine that someone wants to sell their shares of the stock and places a sell order at a limit price of $10.50. This order will go into the sell side of the LOB, as it is above the bid price but below the ask price.

If another person wants to buy the stock and they place a buy order at a limit price of $10.50, this order will match with the sell order in the LOB. The trade will be executed at $10.50, as the buyer is willing to pay that price and the seller is willing to accept it. Both orders are executed, and the LOB is updated accordingly.

However, if the buyer placed a buy order at a higher price, like $11, there would be no match in the LOB and the trade would not be executed. The buy order would stay in the LOB until a seller is willing to accept the higher price.



Market makers

Market makers are entities or individuals who facilitate trading in financial instruments by providing liquidity to the market. They continuously quote bid and ask prices for a particular security, ensuring that there is always a buyer or seller available at any given time.

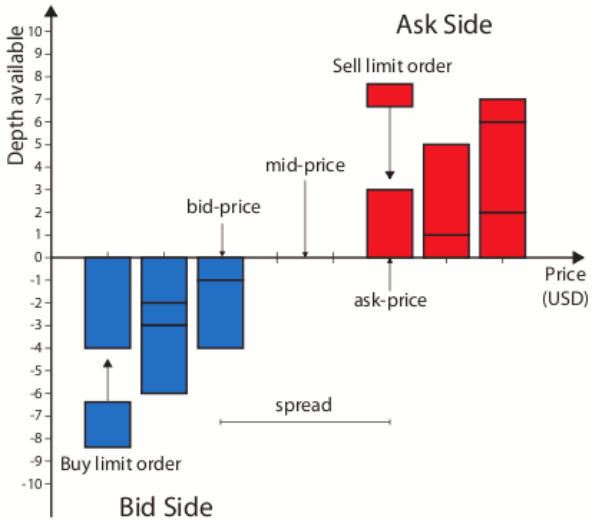
Market makers are typically assumed to provide liquidity, and therefore they are often afforded special trading privileges related to order flow and trade execution. Such **privileges** include

- access to order flow and order flow information,
- direct connections to exchange trading mechanisms,
- low transaction costs, and high transaction speeds.

In return, they are often assumed to perform the social and market function of supplying liquidity, for example by **absorbing temporary order imbalances.**

Here are a few examples of some of the largest market makers:

- Nasdaq, AMEX, NYSE, BNP Paribas, Deutsche Bank, Morgan Stanley, and UBS
- Flow Traders, Citadel Securities, DRW, Jane Street, Optiver, SIG



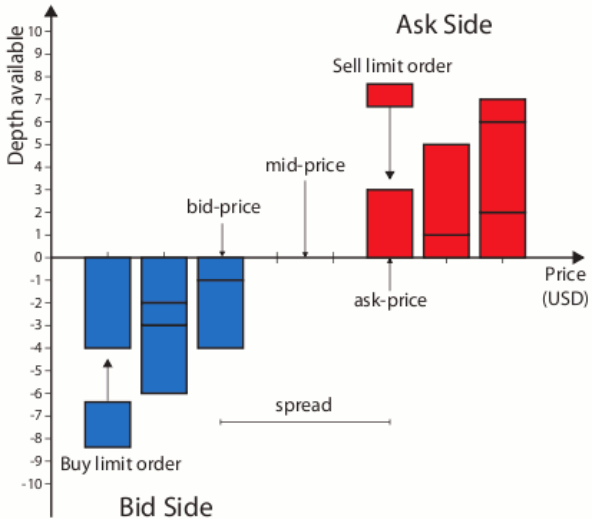
Based on Aldridge, 2013

Market making

Market making describes **placement of limit orders on both sides of the market price**.

A market maker provides liquidity on the markets and is compensated for that.

- A market maker placing a limit buy order just below the market price and a limit sell order just above the market price creates or 'makes' the market.
- If a market *buy order arrives* from another market participant, it is *matched with the market maker's limit sell order*, and the limit sell order is executed (i.e., a short position is recorded in the market maker's account.)
- If a market *sell order arrives* from yet another market participant, it is matched with the *market maker's limit buy order*, and a long position is added to the market maker's account.
- If the size of the long position is added to the market maker's portfolio, the market maker collects the spread as a compensation form providing limit orders, or liquidity, to traders placing market orders. (buys cheap, sells expensive)

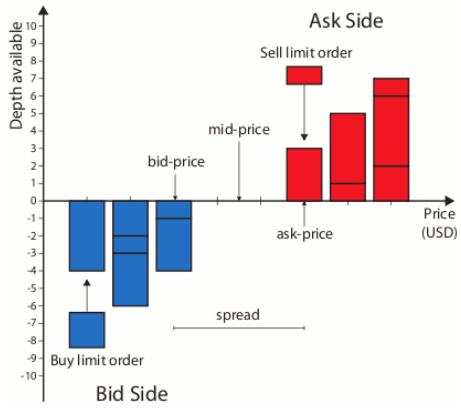


Based on Aldridge, 2013

Simple market making strategies

Fixed offset

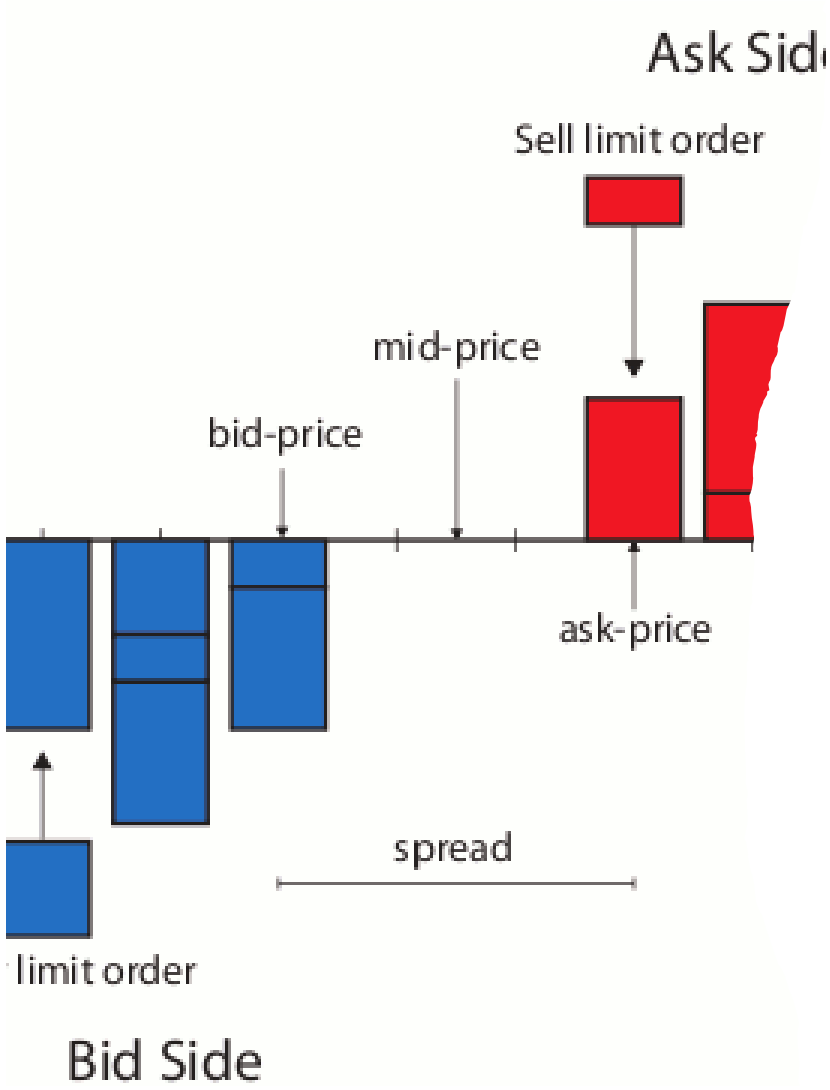
- Continuous placing of limit orders at a p*redetermined number of* ticks away from the market price, on both sides of the market.
- Limit orders placed at current market quotes are likely to be executed.
- In most financial instruments, market makers are allowed to place limit orders *only within 10 percent* of the current market price, to prevent quotes far away from the market from executing at times of extreme volatility. (they would make better profit in this case)
- The smaller the offset of a limit order from the market price, the higher the probability of order execution, and the more frequent the resulting reallocation of the market maker's capital. *Frequency of trading has been shown to be the key to market maker's profitability.*



Volatility-dependent offset

- one way to vary the offset is to make it a function of volatility.
- In high-vola conditions, limit orders farther away from the market are likely to be hit.

Offset is a function of order-arrival rate.



Market making risks

When placing limit orders, the market maker is subject to two kinds of risks:

Inventory risk

- The potential loss the market maker might incur when the value of his inventory declines in price due to natural market movements. Thus, the market maker accumulating a *long position (buying) in a downward-trending market is likely to experience a loss on his position*, at least for a short term.
- When the market maker wishes to close his position (at least in the end of the day), he may face competition from *other parties looking to sell their positions at the same time*.

Risk of adverse selection

- Potential loss due to informational differences between the market maker and a market taker.

Amount of information available about the Limit-Order-Book LOB

This depends on the needs and resources of the traders.

- the only information that is publicly available in real time is **the last traded price or the mid-price** (the point between the current best prices).

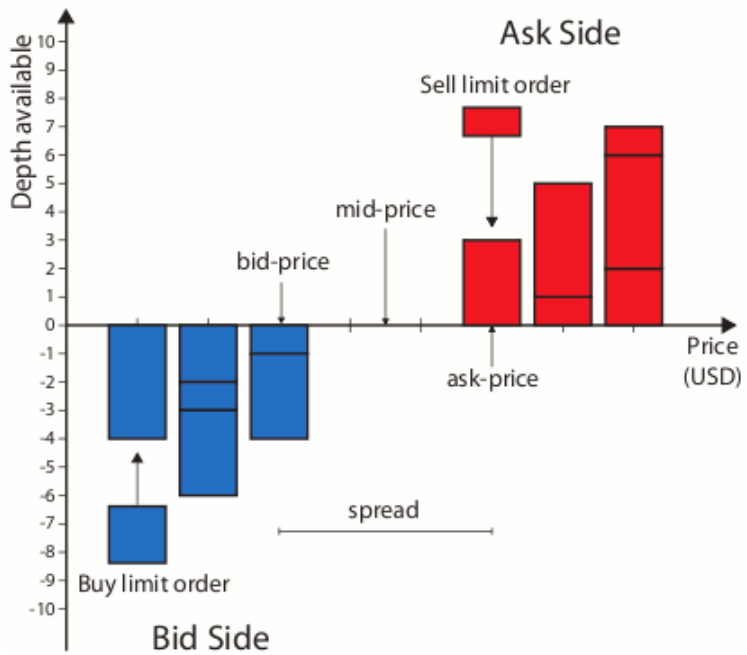
Level 1 market data

- information on the **price and size for the best prices**, along with the **price and size of the last recorded transaction;** needs subscription

Level 2 market data

- **the complete contents of the book** (except for certain types of hidden orders)

Trading in Limit-Order-Book LOB



Trading in a LOB is a highly complex **optimization problem.**

- Traders may submit buy and/or sell orders at different times, prices, quantities and – in today's highly fragmented markets – often to multiple order books.
- Order may also be modified or cancelled at any time.

In addition to limit orders there are also **market orders** (sell or buy at current price whatever it is):

- A market order is a type of order to buy or sell a security at the current market price. When placing a market order, the investor instructs their broker to **execute the order immediately, *without specifying a price*.** The trade is executed as quickly as possible at the best available price at that moment.
- In selling, the broker picks up the next closest bids in the limit order book.
- In buying, the broker utilizes the next closest asks.

Several measures for liquidity

What kind of measures for the liquidity are there? Why are we interested in liquid markets?

Several measures for liquidity

The tightness of the bid-ask spread

Reflects the degree of competition among limit order-placing traders

Market depth at best bid and best ask

This is the ability to **sustain relatively large orders without impacting the price of the security**

Shape of the order book

Level II data can be used to calculate the exact amount of aggregate supply and demand available on a given trading venue.

Price sensitivity of order-flow imbalance

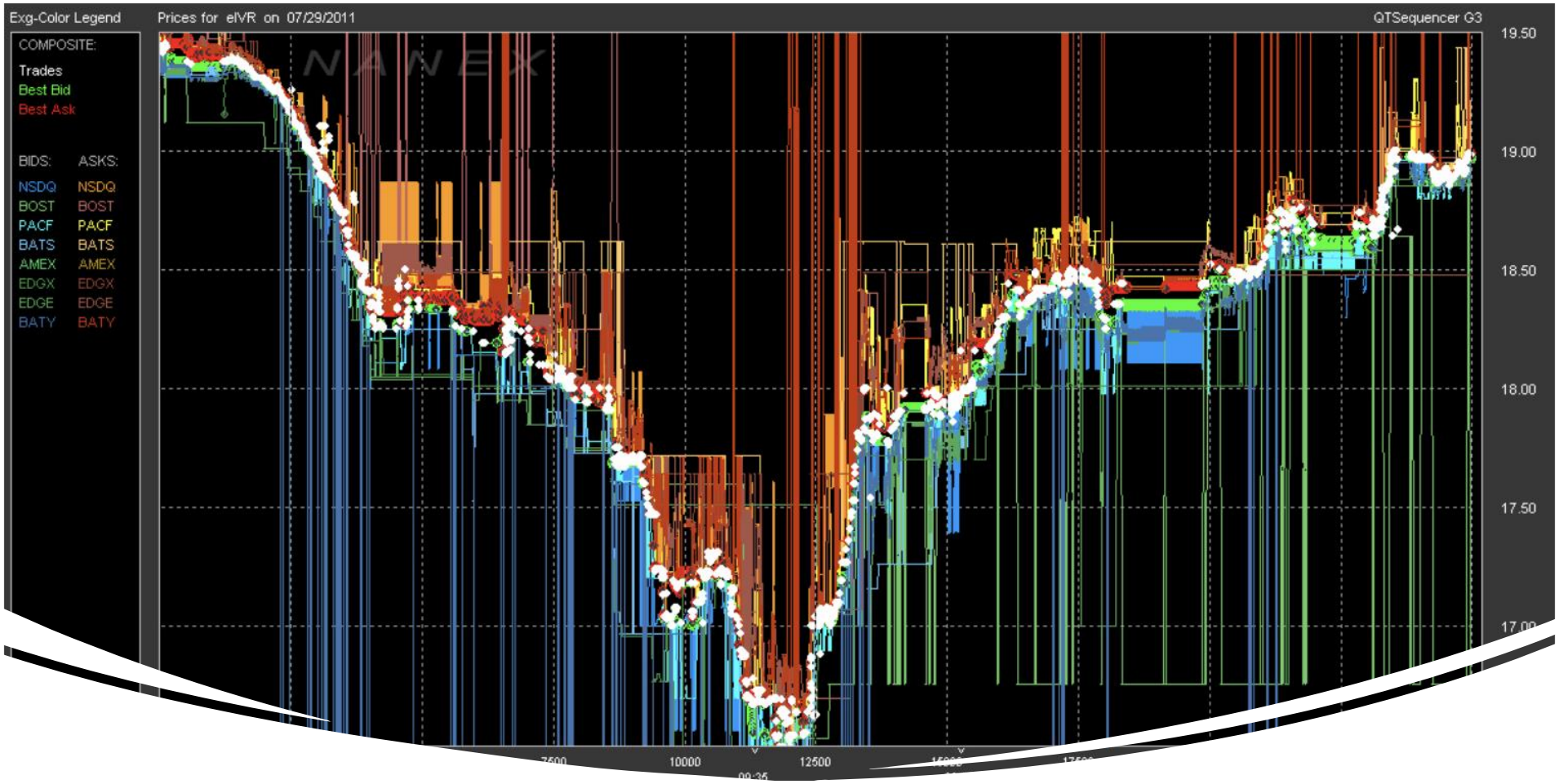
How the **price moves following an order**; low liquidity makes the market eat through the limit order book, moving the price substantially

Market resilience

How f**ast an order book recovers its shape** following a market order

High-frequency trading HFT
Trading in fractions of seconds

Image by Sabrina Jiang © Investopedia 2020



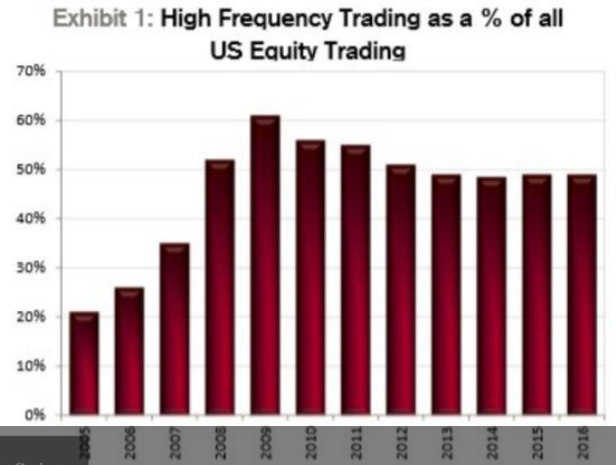
High Frequency Trading

Poll

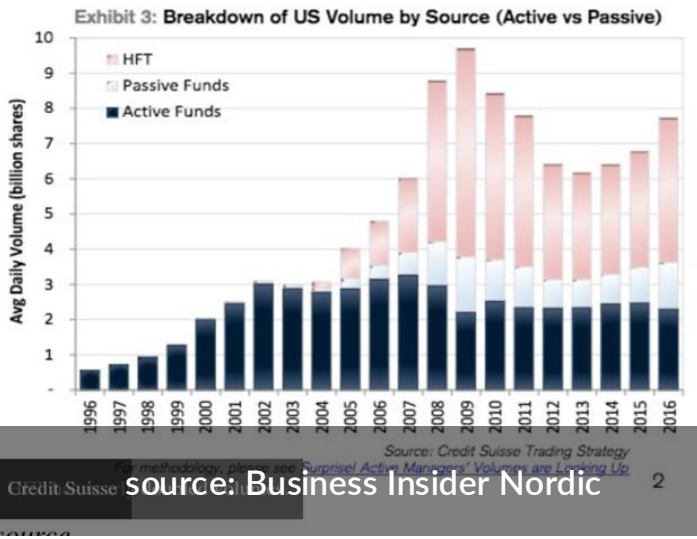
What percentage of all US trading volume was high-frequency trading in 2020?

1. 10
2. 25
3. 50
4. 75

Vote in https://presemo.aalto.fi/firma3



In 2020, 50 %



source: Business Insider Nordic

High Frequency Trading

High frequency trading is a specialized case of algorithmic trading involving the frequent turnover of *many small positions of a security*.

There is *no formal definition* of HFT. The U.S. Securities and Exchanges Commission attributes certain specific characteristics:

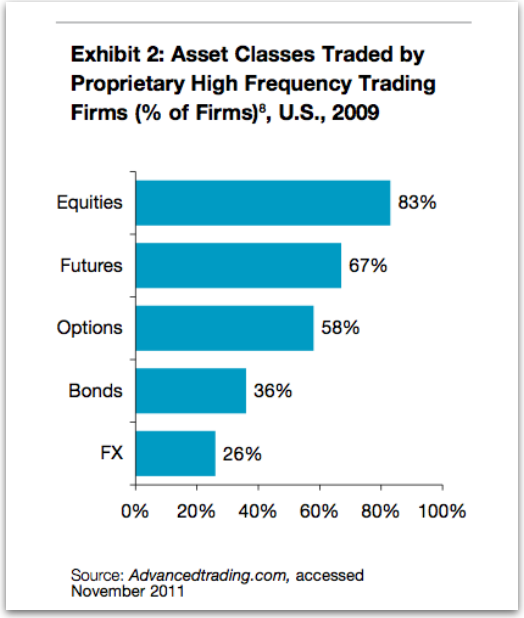
The use of *extremely sophisticated and high-speed computer* programs for generating, routing, and executing orders.

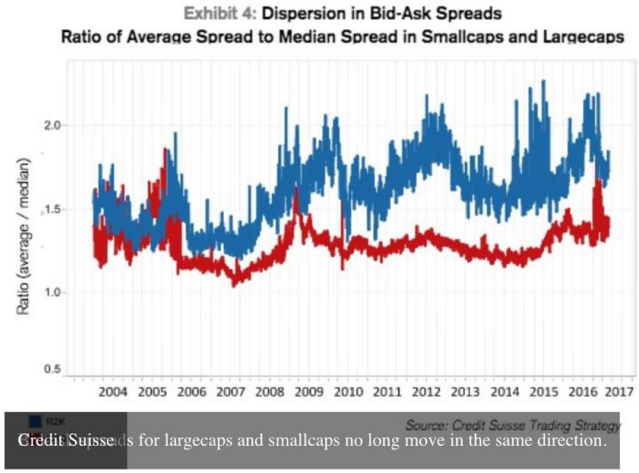
The use of individual data feeds from exchanges as well as co-located servers in order to *minimize network and other types of latencies*.

Maintaining very short timeframes for establishing and liquidating positions, resulting in the *frequent turnover of many small positions* in one or more financial instruments.

Submitting a number of orders that are cancelled soon after submission.

Maintaining *very few*, if any, *overnight positions*.





source: Business Insider Nordic

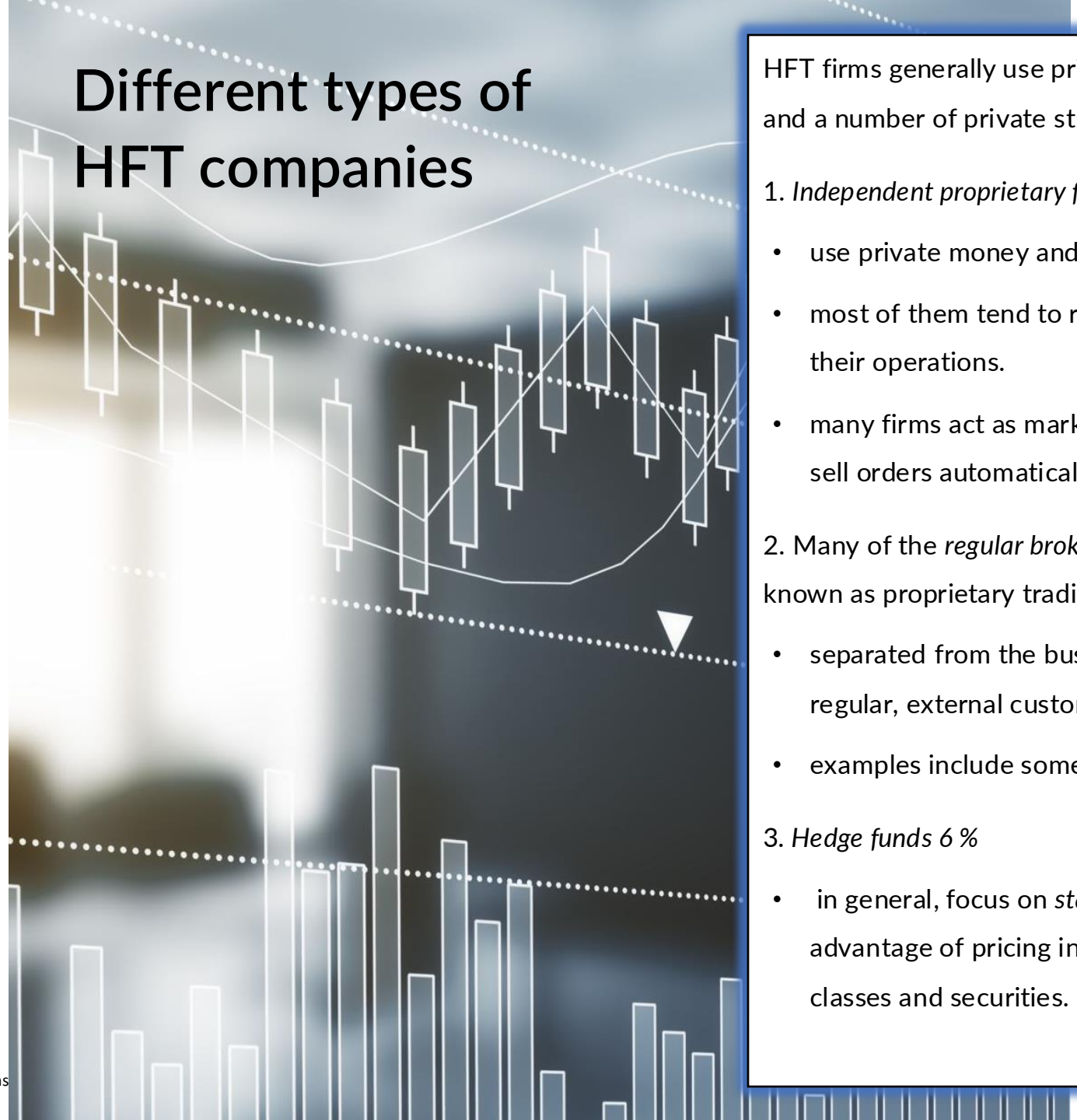
HFT and bid-ask spreads

Along with the rise of HFT,

- the bid-ask spreads for large cap stocks (red) have tightened and
- the bid-ask spread of small caps (blue) have widened.

HFT is concentrating on the trading in the most liquid, biggest stocks. They are providing liquidity to the markets.

Different types of HFT companies



HFT firms generally use private money, private technology and a number of private strategies to generate profits.

1. *Independent proprietary firm 48 %*

- use private money and different strategies
- most of them tend to remain rather secretive about their operations.
- many firms act as market makers, generating buy and sell orders automatically throughout the day

2. Many of the *regular broker-dealer firms* have a *sub section* known as proprietary trading desks for HFT is done. 46 %

- separated from the business the firm does for its regular, external customers.
- examples include some of the largest investment banks.

3. *Hedge funds 6 %*

- in general, focus on *statistical arbitrage* and take advantage of pricing inefficiencies among various asset classes and securities.

Ideal asset attributes to HTF

Similar to other traders

- Heavy trading volume → narrower bid-ask spreads and abundant liquidity

But

Low volatility

- A trading environment where small amounts of profit can be made with near 100% certainty is preferably to one where larger amounts of profit can be made with a lesser degree of certainty. (typically, traders prefer high vola)

Low stock price—the volume of trading is important

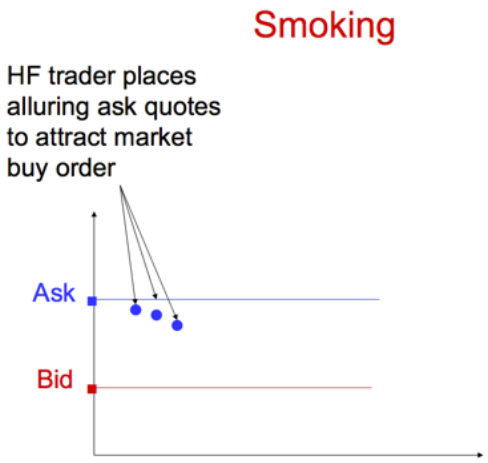
- As *rebates (compensation for market making) are based on the number of shares traded*, so for the same amount of funds deployed in a trade, an HFT firm can earn a larger total rebate on a lower-priced stock than on a higher-priced stock. (typically, traders are indifferent to the price)

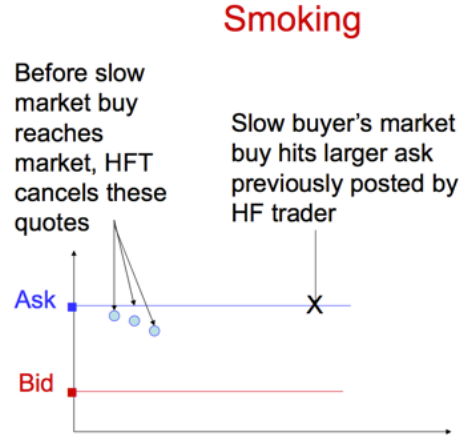
Name of HFT strategy	Description
Electronic market making	Liquidity-providing strategies that mimic the traditional role market makers once played. These strategies involve making a two-sided market aiming at profiting by earning the bid-ask spread. This has evolved into what is known as Passive Rebate Arbitrage.
Statistical arbitrage	Traders look to correlated prices between securities and trade on the imbalances in those correlations.
Liquidity detection	Traders look to decipher whether there are large orders existing in a matching engine by sending out small orders ("pinging") to look for where large orders might be resting. When a small order is filled quickly, there is likely to be a large order behind it.

Source: Aldridge 2010

Two HFT techniques: smoking and spoofing

Additional material





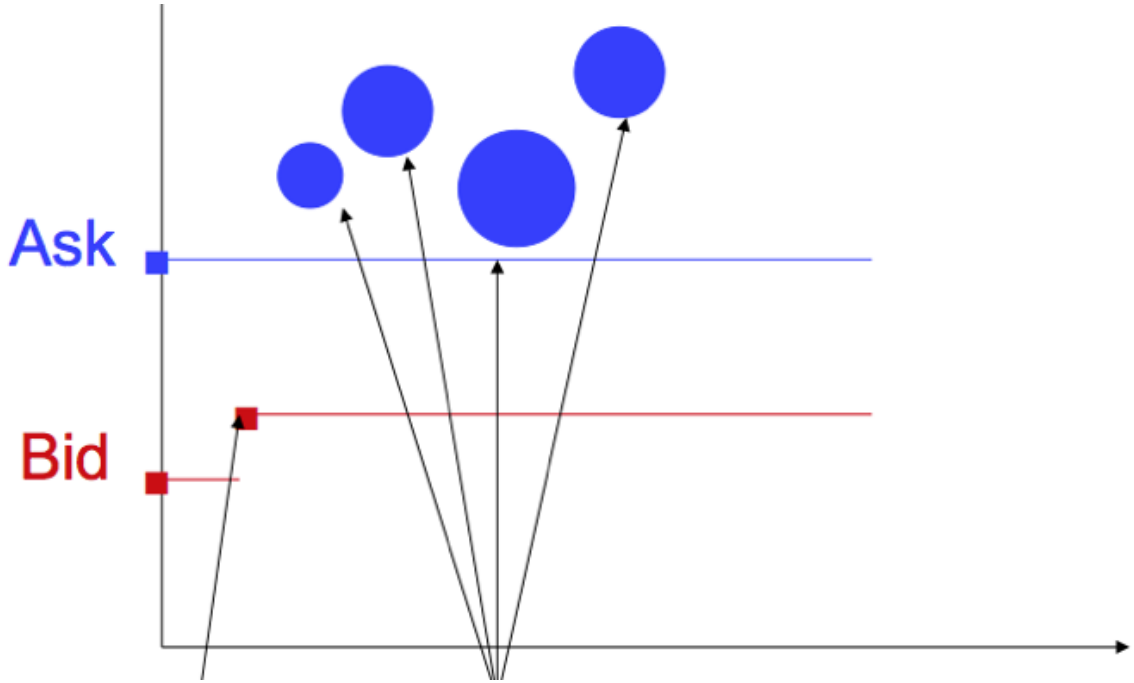
Luring with a good offer

high frequency traders first post locked limit orders to attract slow traders. Then they rapidly revise these orders onto less generous terms, hoping to execute profitably against the incoming flow of slow traders' market orders.

- limit order: execute at limit price
- Market order: execute at market price, what ever it is

- The **bid price**: the highest **price** a buyer is prepared to pay
- the **ask price** is the lowest **price** a seller will accept.

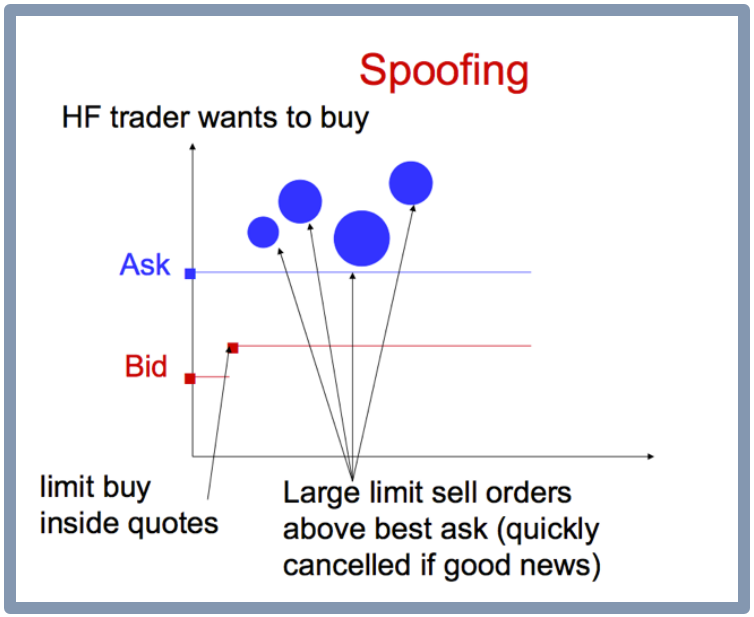
Sens



Poll

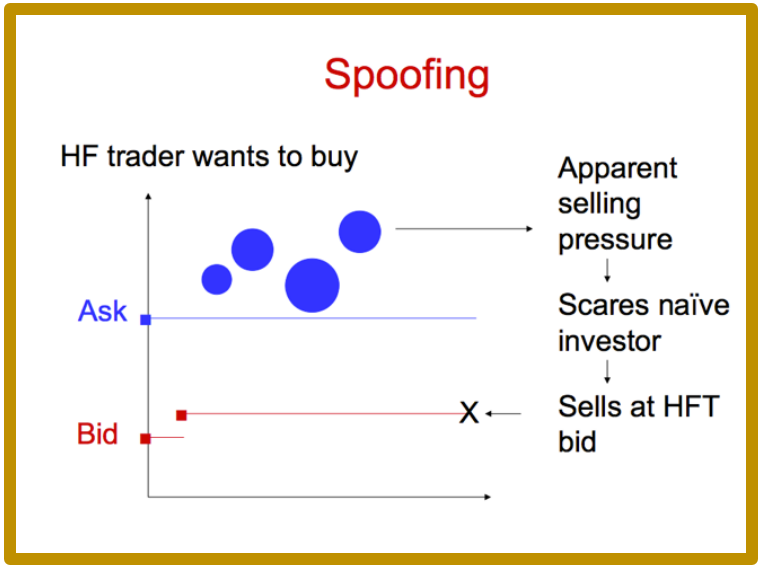
You can see the following information in the limit order book.
Would you

1 sell
2 buy

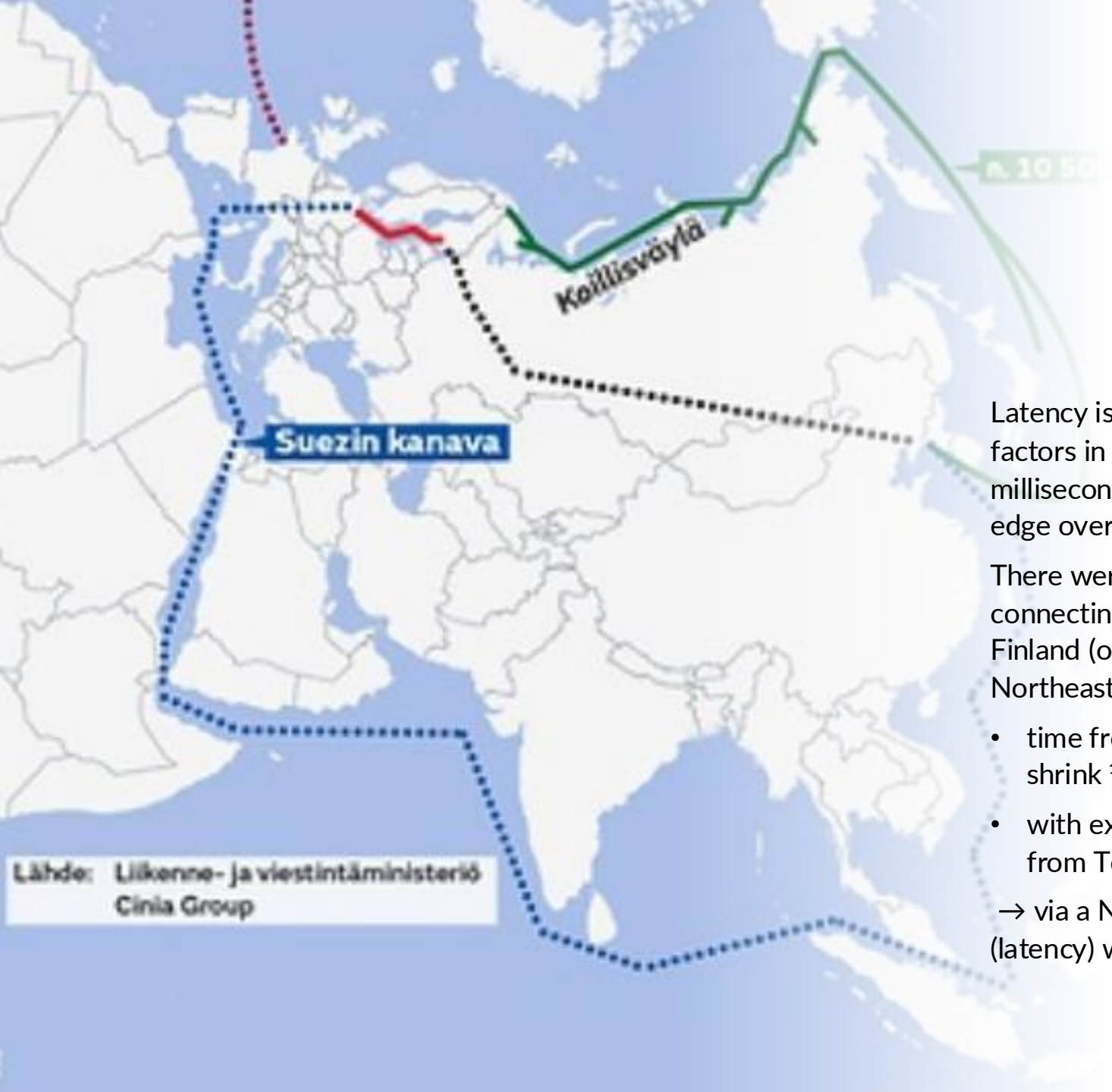


"Spoofing": Suppose the high frequency trader's true **intention is to buy**. Paradoxically, he or she will initially place limit orders to sell in the order book. These orders are not intended to be executed. Therefore, they are placed above the best ask. And, since the high frequency trader is faster than the other market participants, he or she can rest assured he or she will have time to cancel the sell orders before they are executed, if good news reach the market.

With this assurance in mind, the high frequency trader places a sequence of limit sell orders above the best ask, potentially for very large amounts. The hope is to scare the market and induce some naïve participant to sell against the limit order to buy the high frequency trader will have discretely placed meanwhile. (a much lower price)



- Responsible for transmitting approximately 99% of international data, including financial transactions.
- Important in Algo Trading and HFT: low latency and market access (various exchanges, simultaneous trading)
- Hibernia Express for London to New York reduced transatlantic latency by 5 ms.
- **Japan-US Cable Network connects** Asia-Pacific markets to North America
- Risks: cable breakage and geopolitical risks.
    - HFT firms invest in backup routes to mitigate risks.
    - Exploration of alternative technologies, such as satellite communication, though cables remain faster.



Latency is one of the most crucial factors in HFT. Advantages of a few milliseconds can give an investor edge over its competitors.

There were plans for a new cable connecting Europe and Far East via Finland (or Norway) and the Northeast passage.

- time from Berlin to Tokyo would shrink ¼.
- with existing connections time from Tokyo to Helsinki is 247 ms.

→ via a Northeast passage the time (latency) would be 153 ms.


See you next time!