CSSS 2011, July 3-4 ISC Paris

ORDER BOOK MODELS AND THE MYSTERIES OF ZIPF LAW

Luciano Pietronero

Matthieu Cristelli, A. Zaccaria

Institute of Complex Systems, CNR, Rome Italy

University of Rome Sapienza and Centro Fermi, Rome

(WEB page: http://pil.phys.uniroma1.it)

Back to Liquidity problem

Liquidity: seems more important than volume or news for price changes

Microscopic model for the order book & finite liquidity (crisis).

This should be included in a realistic ABM

Order Book & ABM

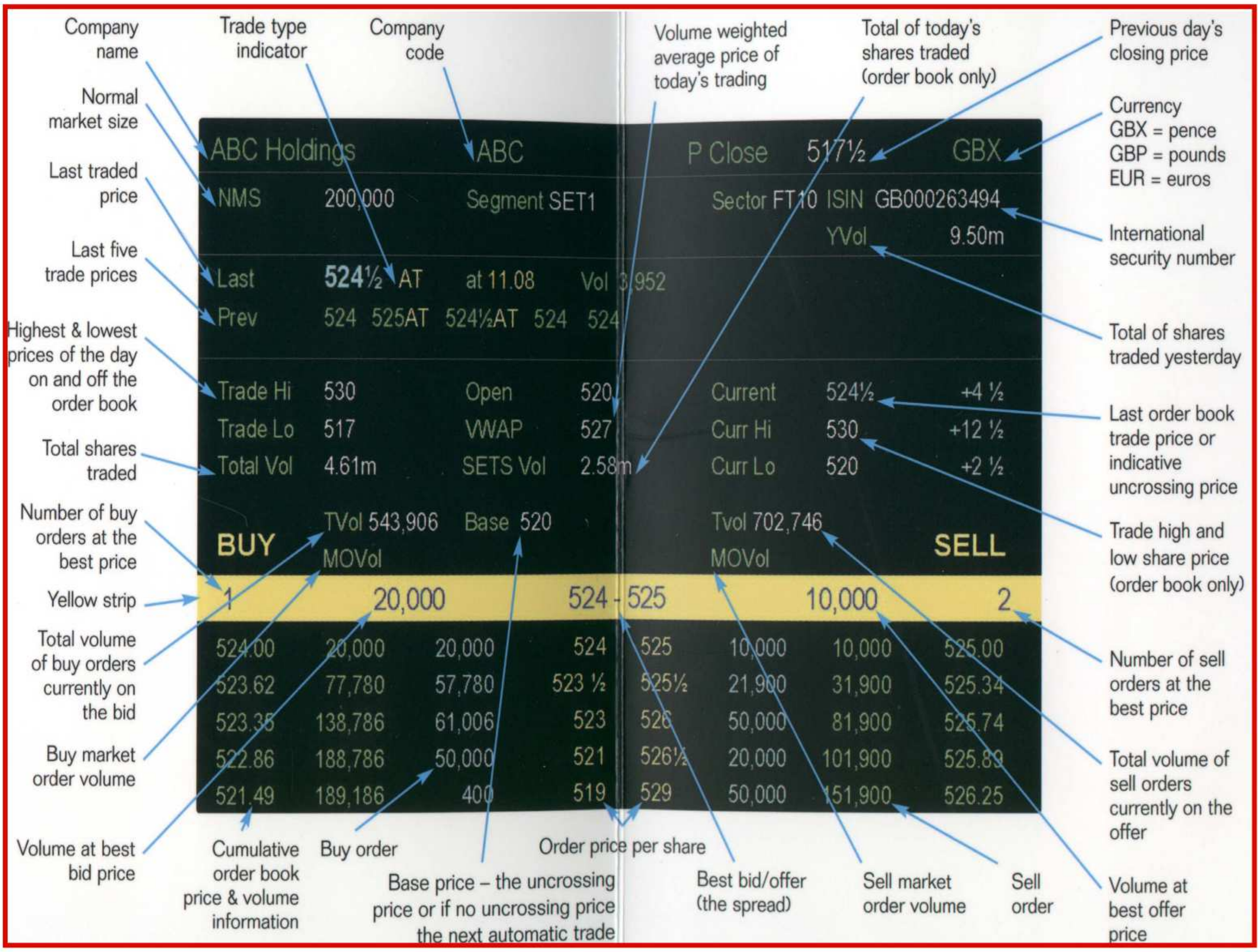
In a typical Agent-Based Model (ABM) the price evolution is a coarse-grained clearing/adjustment mechanism that does not take into account the liquidity of the market

$$\frac{1}{p}\frac{dp}{dt} = \beta\, ED(t) \quad \xrightarrow{\text{real markets}} \quad \frac{1}{p}\frac{dp}{dt} = \boxed{\beta(g)}\, ED(t)$$

Therefore we need to investigate the microscopic mechanisms for price formation in order to find $\beta(g)$

↓

Order book model


Company name
Trade type indicator
Company code
Volume weighted average price of today's trading
Total of today's shares traded (order book only)
Previous day's closing price
Normal market size
Currency GBX = pence GBP = pounds EUR = euros
Last traded price
International security number
Last five trade prices
Total of shares traded yesterday
Highest & lowest prices of the day on and off the order book
Last order book trade price or indicative uncrossing price
Total shares traded
Trade high and low share price (order book only)
Number of buy orders at the best price
Yellow strip
Total volume of buy orders currently on the bid
Number of sell orders at the best price
Buy market order volume
Total volume of sell orders currently on the offer
Volume at best bid price
Cumulative order book price & volume information
Buy order
Order price per share
Base price – the uncrossing price or if no uncrossing price the next automatic trade
Best bid/offer (the spread)
Sell market order volume
Sell order
Volume at best offer price
ABC Holdings ABC P Close 517½ GBX
NMS 200,000 Segment SET1 Sector FT10 ISIN GB000263494
YVol 9.50m
Last 524½ AT at 11.08 Vol 3,952
Prev 524 525AT 524½AT 524 524
Trade Hi 530 Open 520 Current 524½ +4 ½
Trade Lo 517 VWAP 527 Curr Hi 530 +12 ½
Total Vol 4.61m SETS Vol 2.58m Curr Lo 520 +2 ½
TVol 543,906 Base 520 Tvol 702,746
BUY MOVol MOVol SELL
1 20,000 524 - 525 10,000 2
524.00 20,000 20,000 524 525 10,000 10,000 525.00
523.62 77,780 57,780 523 ½ 525½ 21,900 31,900 525.34
523.35 138,786 61,006 523 526 50,000 81,900 525.74
522.86 188,786 50,000 521 526½ 20,000 101,900 525.89
521.49 189,186 400 519 529 50,000 151,900 526.25

Order Book in a nutshell

The elementary mechanism of price formation is a double auction where traders submit orders to buy or sell. The complete list of all orders is called the order book.

Two classes of orders
- market orders (impatient traders)
- limit orders (patient traders)

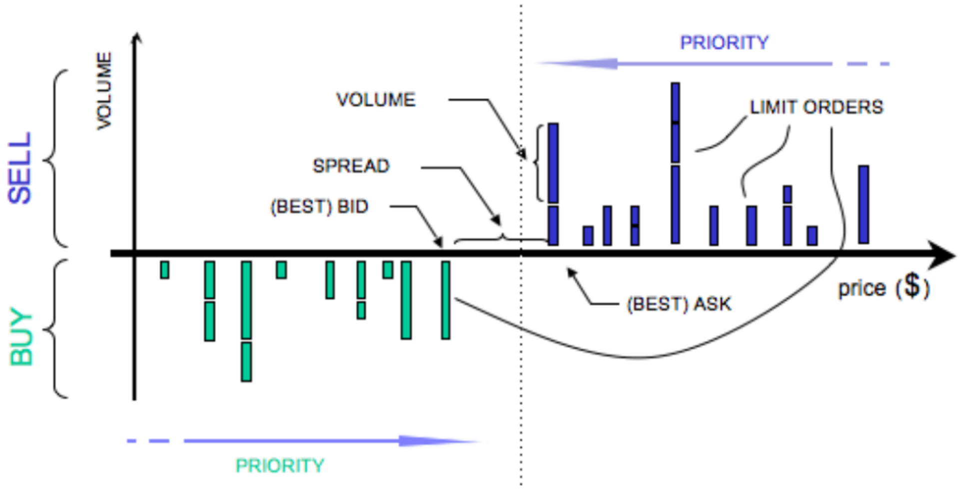
- Market orders correspond to the intention of immediately purchase or sale at the best price (quote) available at that time
- The limit ones instead are not immediately executed since they are orders to buy or sell at a certain quote which is not necessary the best one.

best bid b(t): order of buying with the highest price

best ask a(t): order of selling with the lowest price

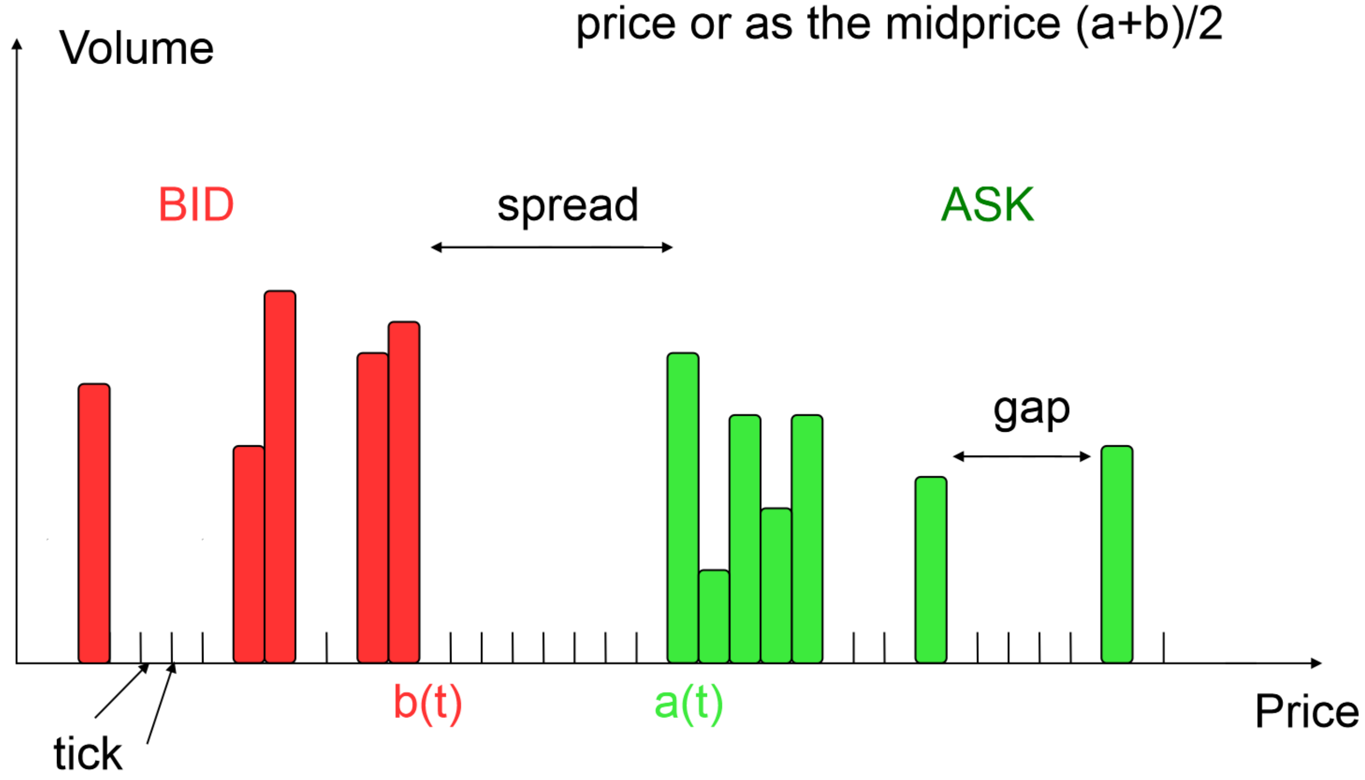
The non-zero difference between a(t) and b(t) is defined as the spread $s(t) = a(t) - b(t)$

The prices of placement of orders, called "quotes", are not continuous but quantized in unit of ticks whose size is an important parameter of an order book.



How an Order Book works

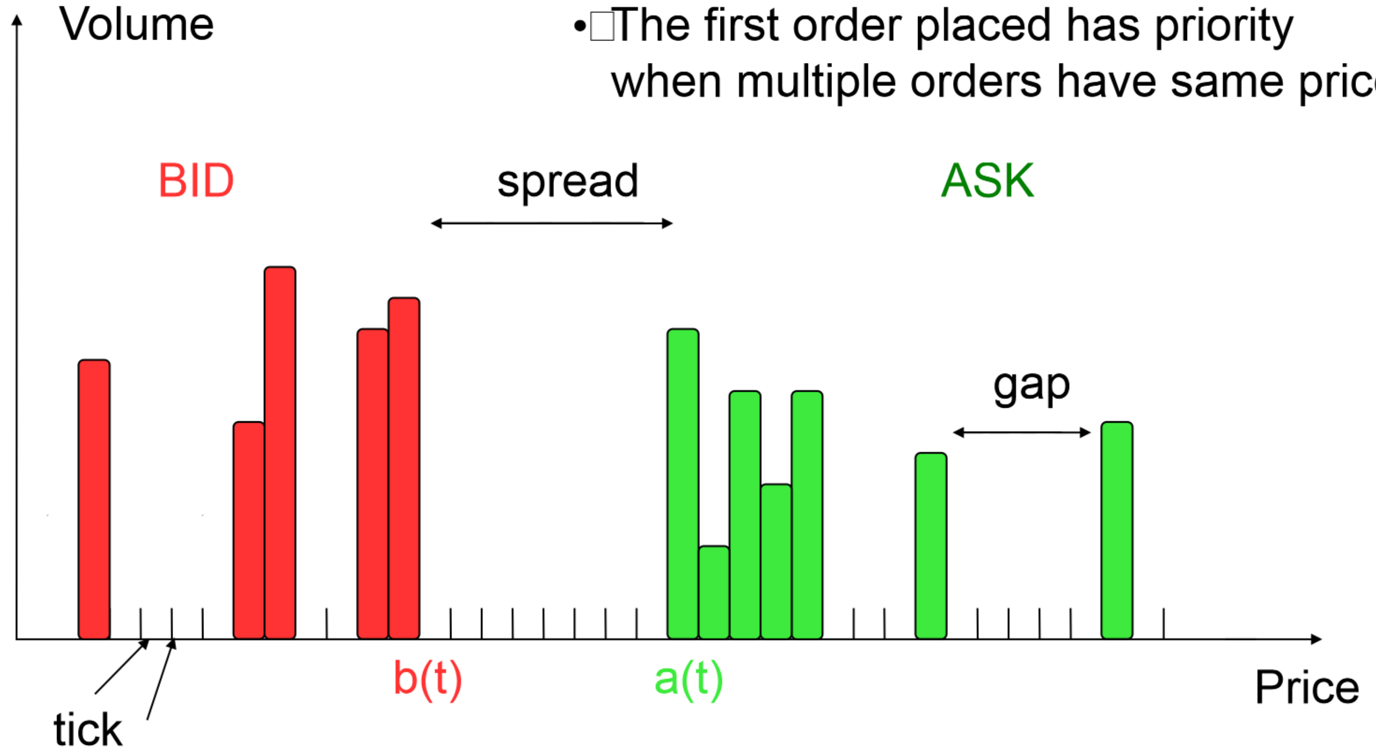
Price is defined as the last transaction price or as the midprice (a+b)/2



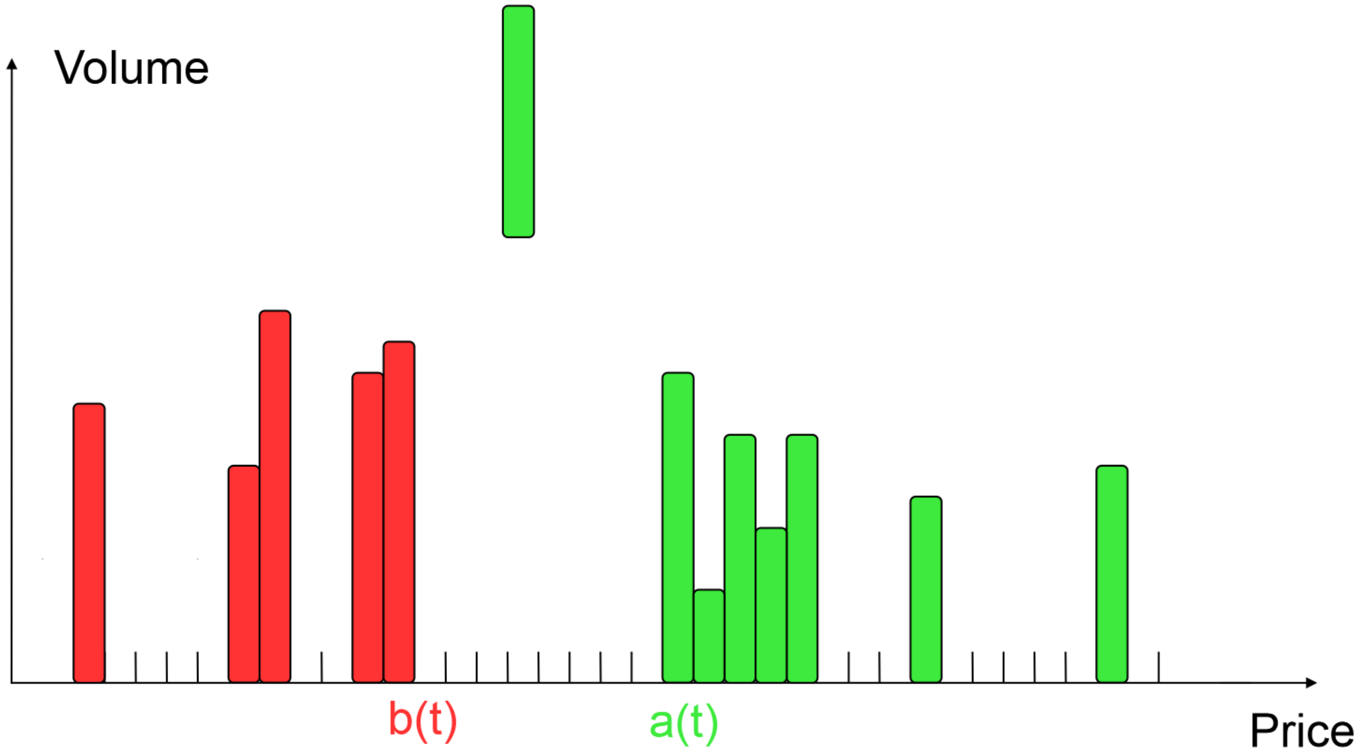
How an Order Book works

Execution priority:

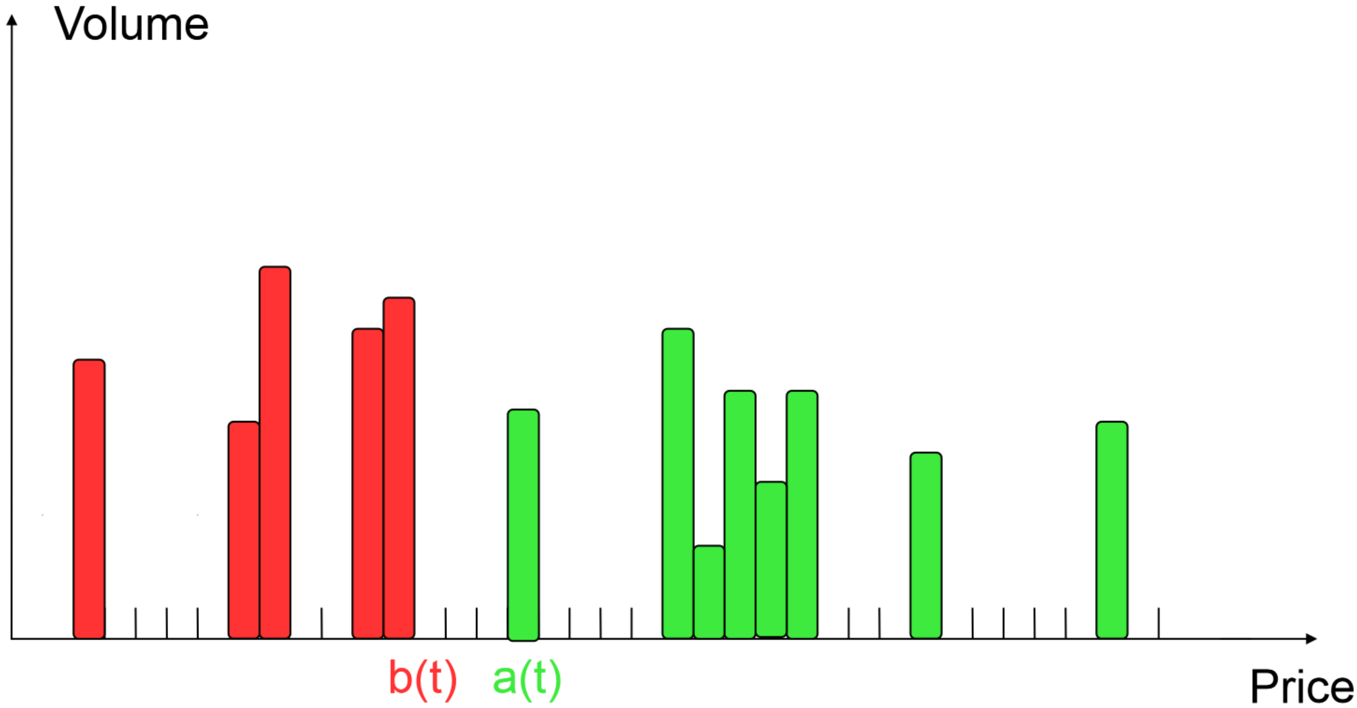
- Lower priced sell orders or higher priced buy orders have priority
- The first order placed has priority when multiple orders have same price.



Limit orders

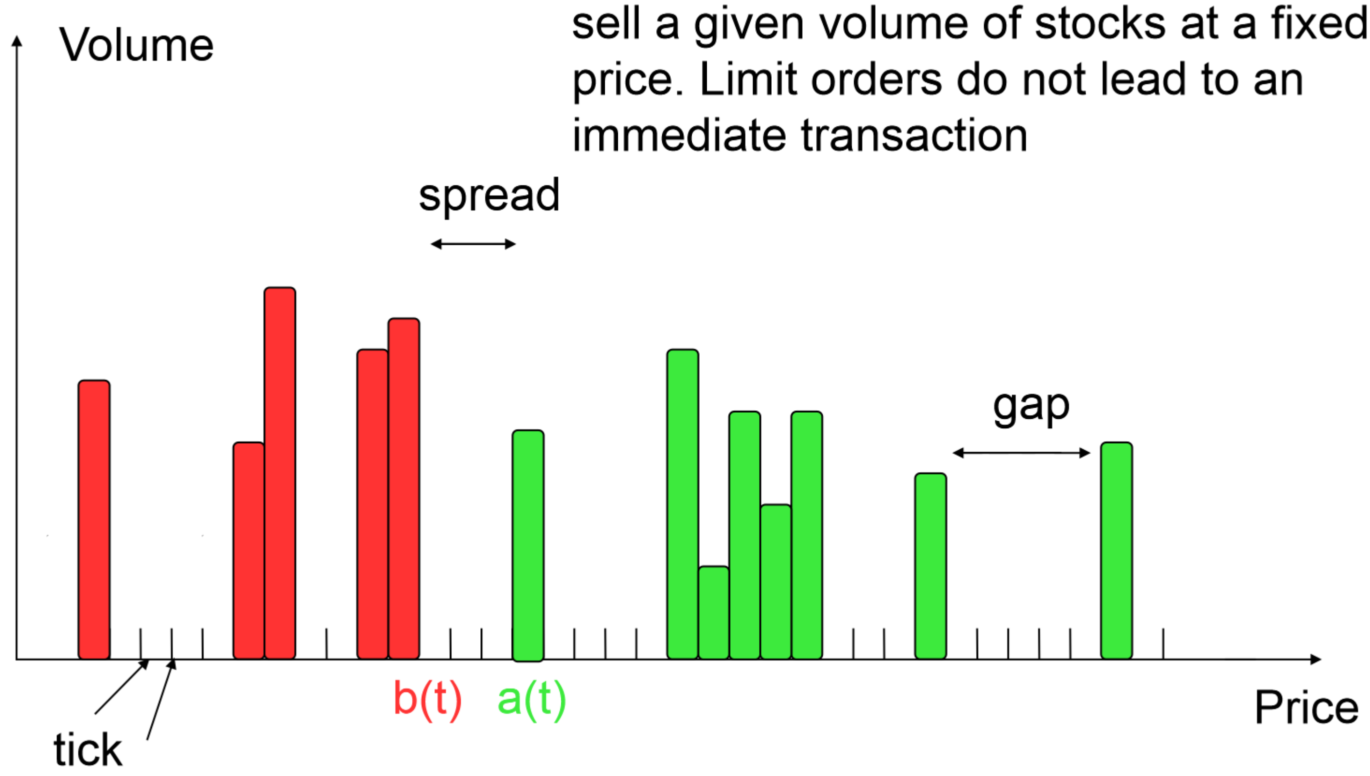


Limit orders

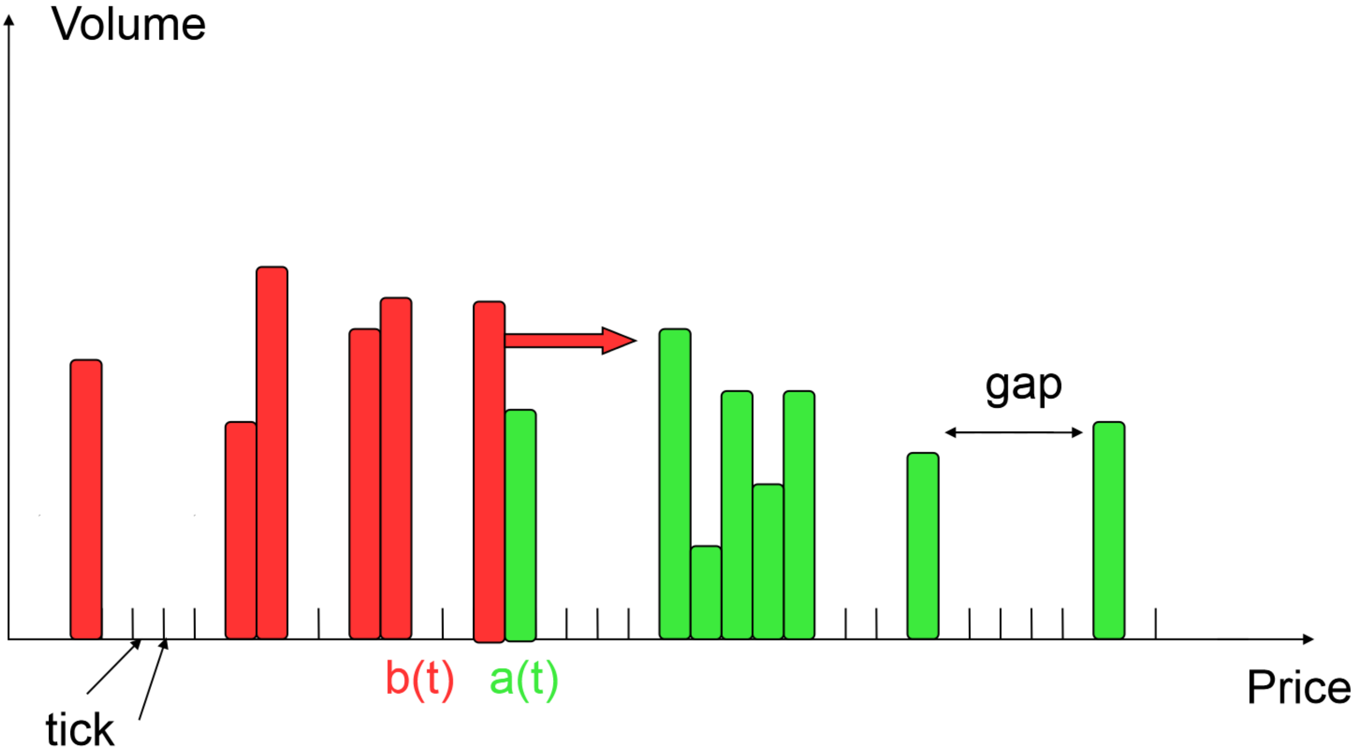


Limit orders

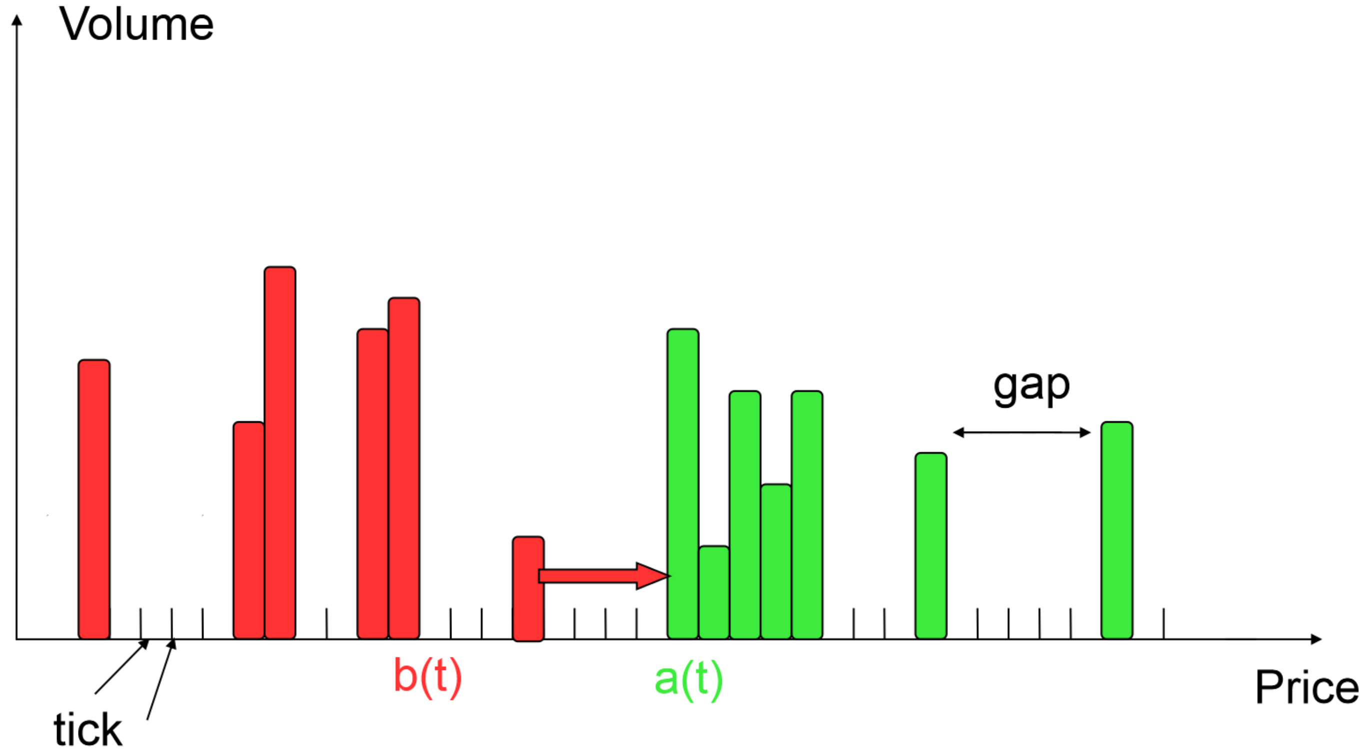
Patient traders place orders to buy or sell a given volume of stocks at a fixed price. Limit orders do not lead to an immediate transaction



Market orders

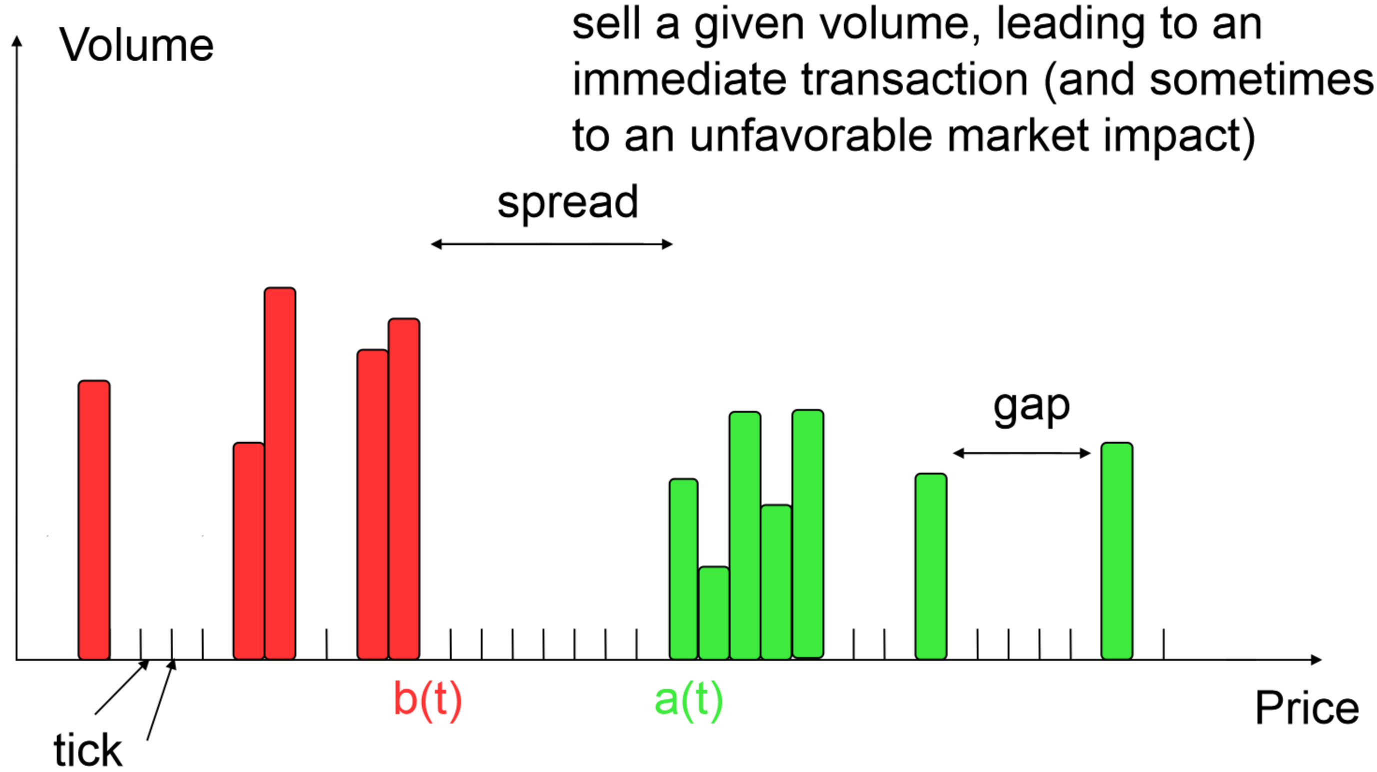


Market orders

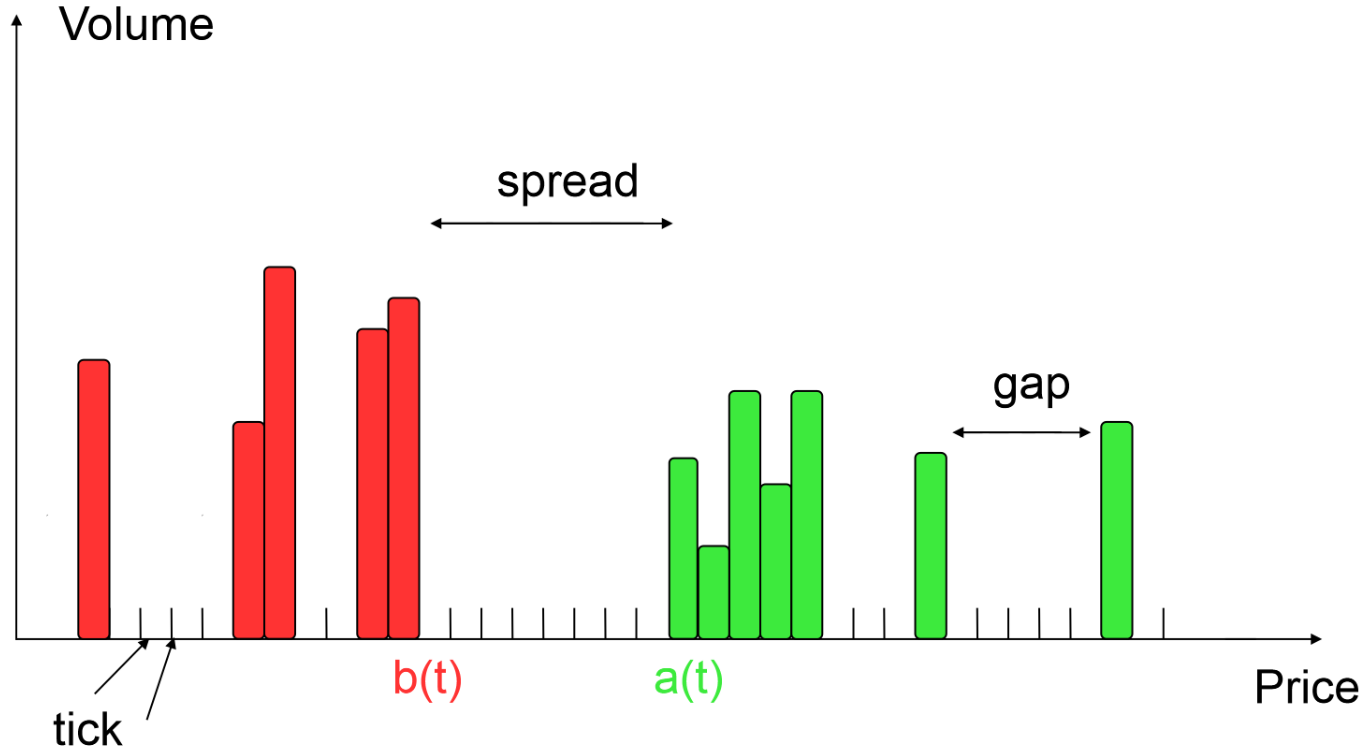


Market orders

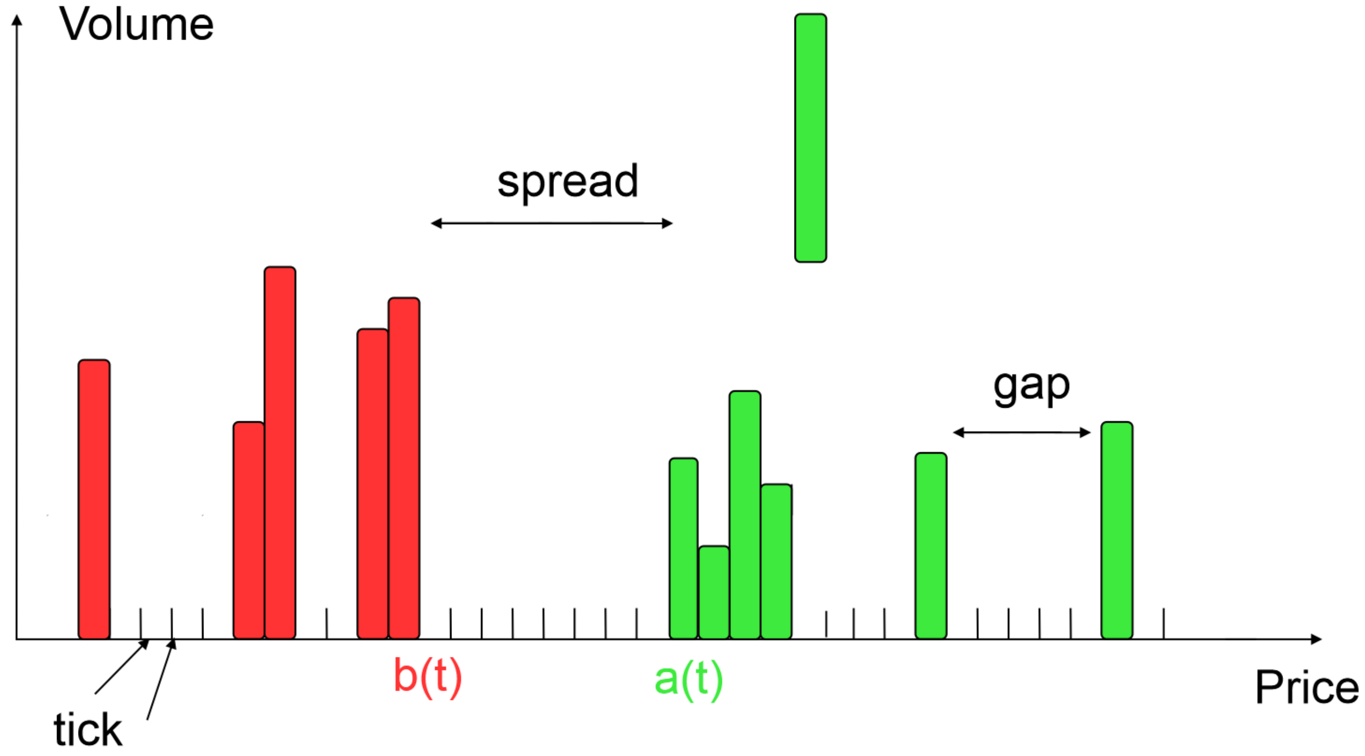
Impatient traders place orders to buy or sell a given volume, leading to an immediate transaction (and sometimes to an unfavorable market impact)



Cancellations

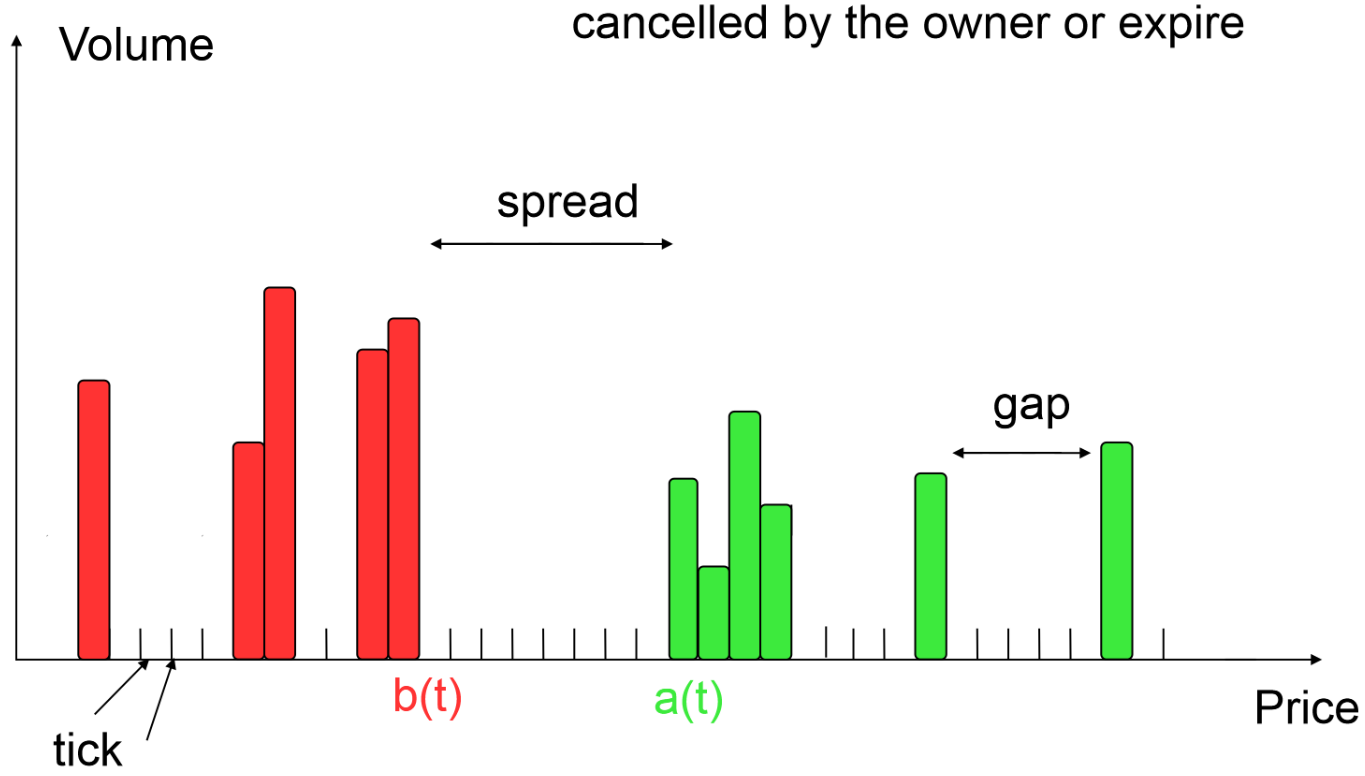


Cancellations



Cancellations

An unmatched limit order can be cancelled by the owner or expire



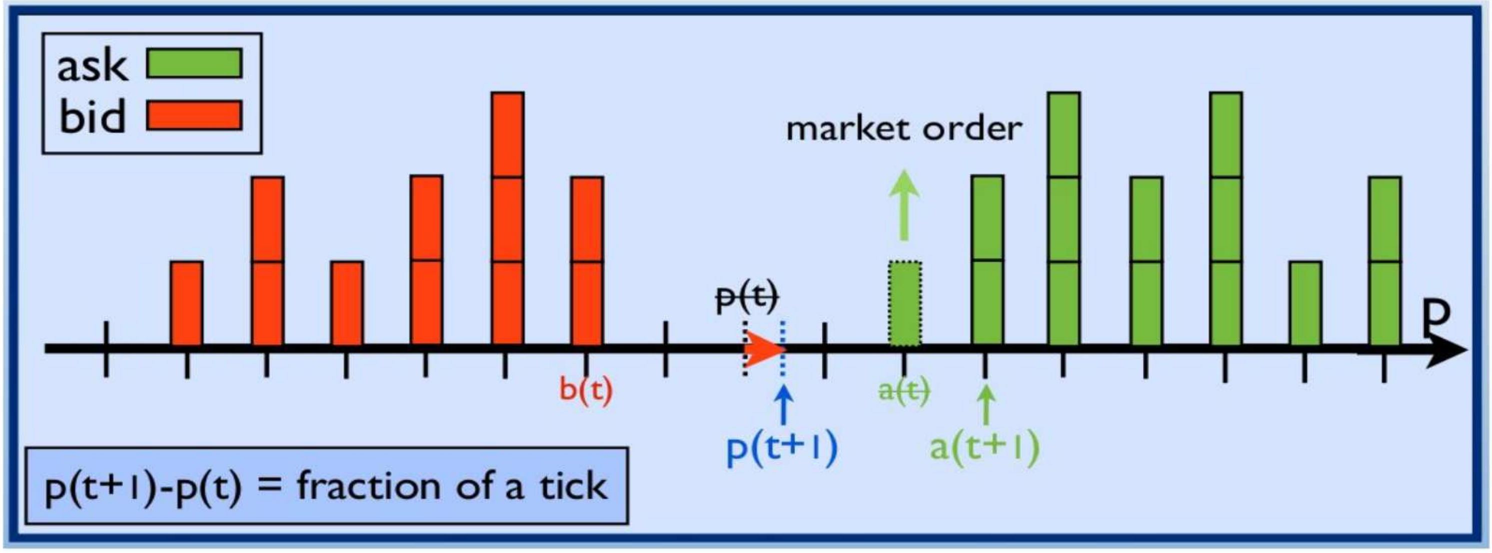
Liquidity

Liquidity is the degree to which an asset or security can be bought or sold in the market without affecting the asset's price. Liquidity is enhanced by a high level of trading activity. Assets that can by easily bought or sold are known as liquid assets.

Liquidity is crucial to determine the price impact of a given order, that is the price variation caused by the arrival of the order

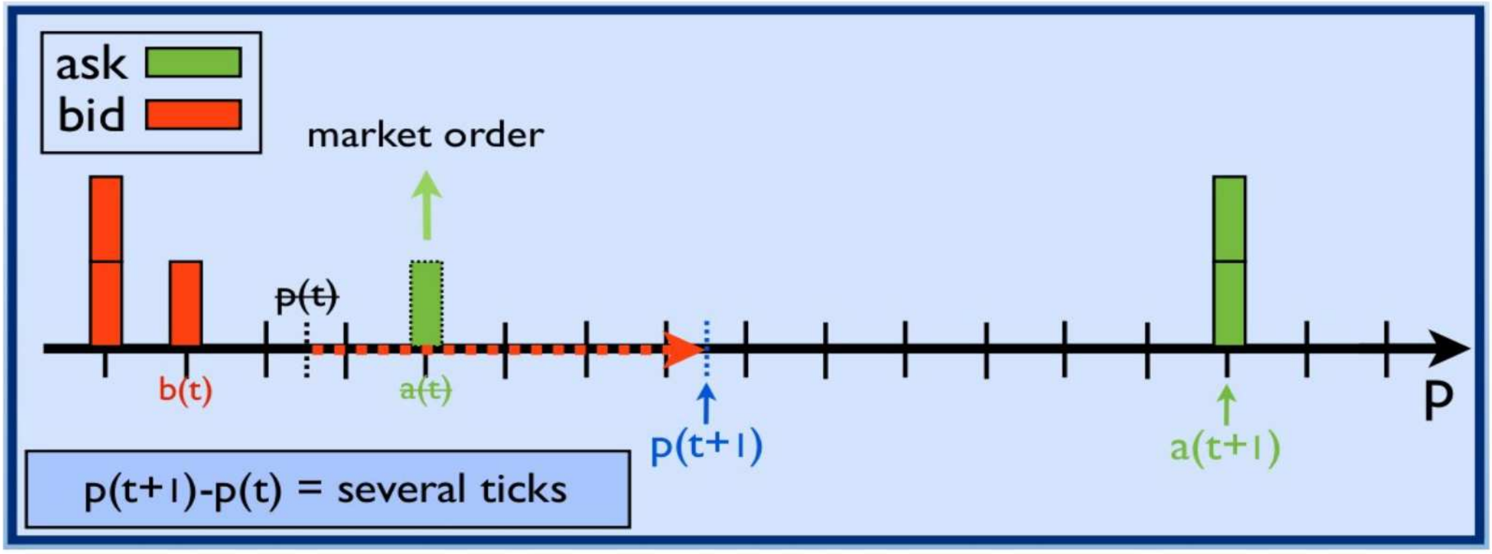
The standard economic theory assumes infinite liquidity i.e. a zero price impact

Liquid vs illiquid market



- Small price variations
- Behavior similar to a continuous system

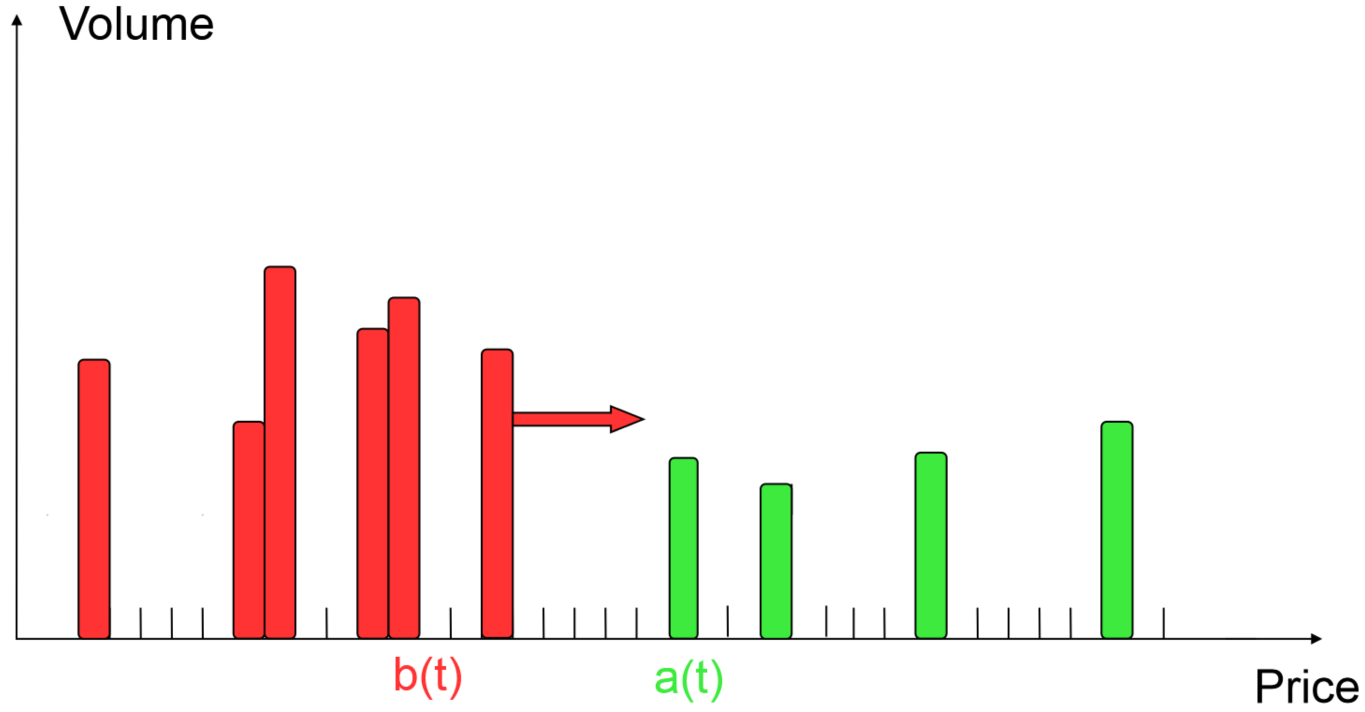
Liquid vs illiquid market



- Large price variations
- The discreteness of the system is crucial

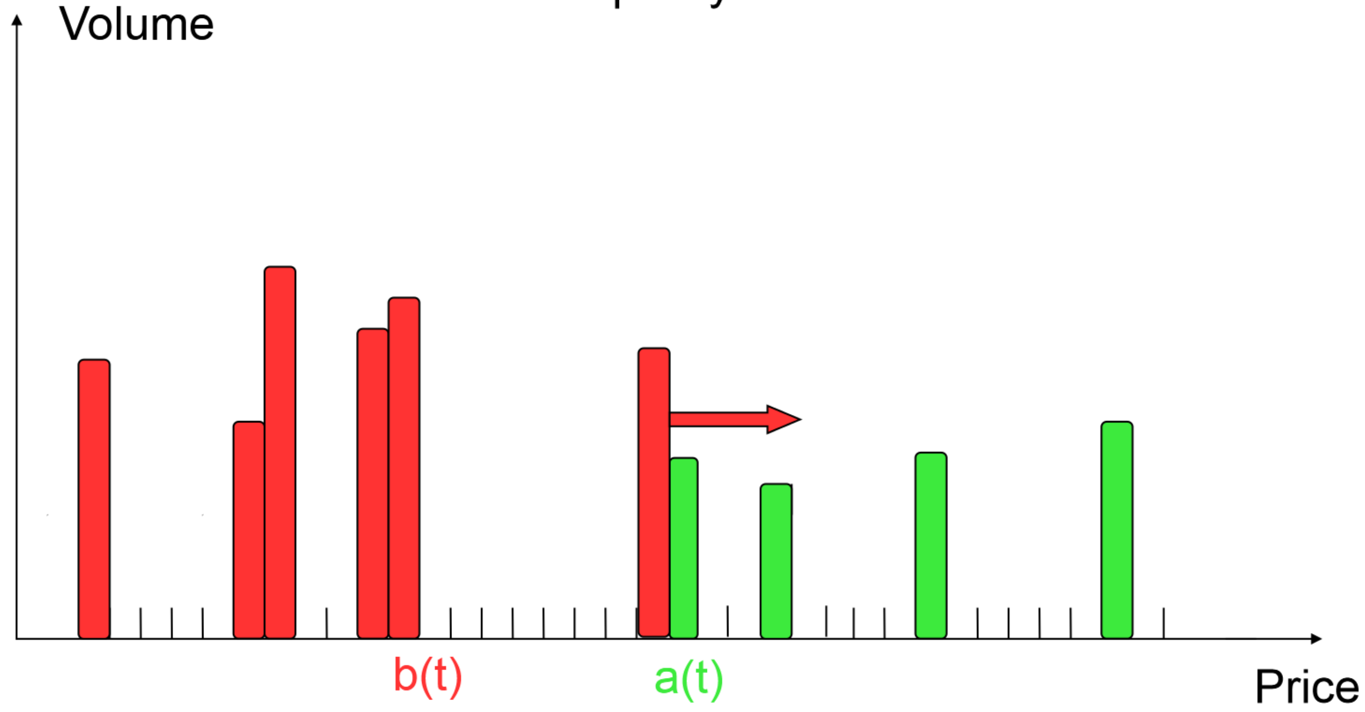
Revealed Liquidity

Suppose a buy market order arrives in the market.

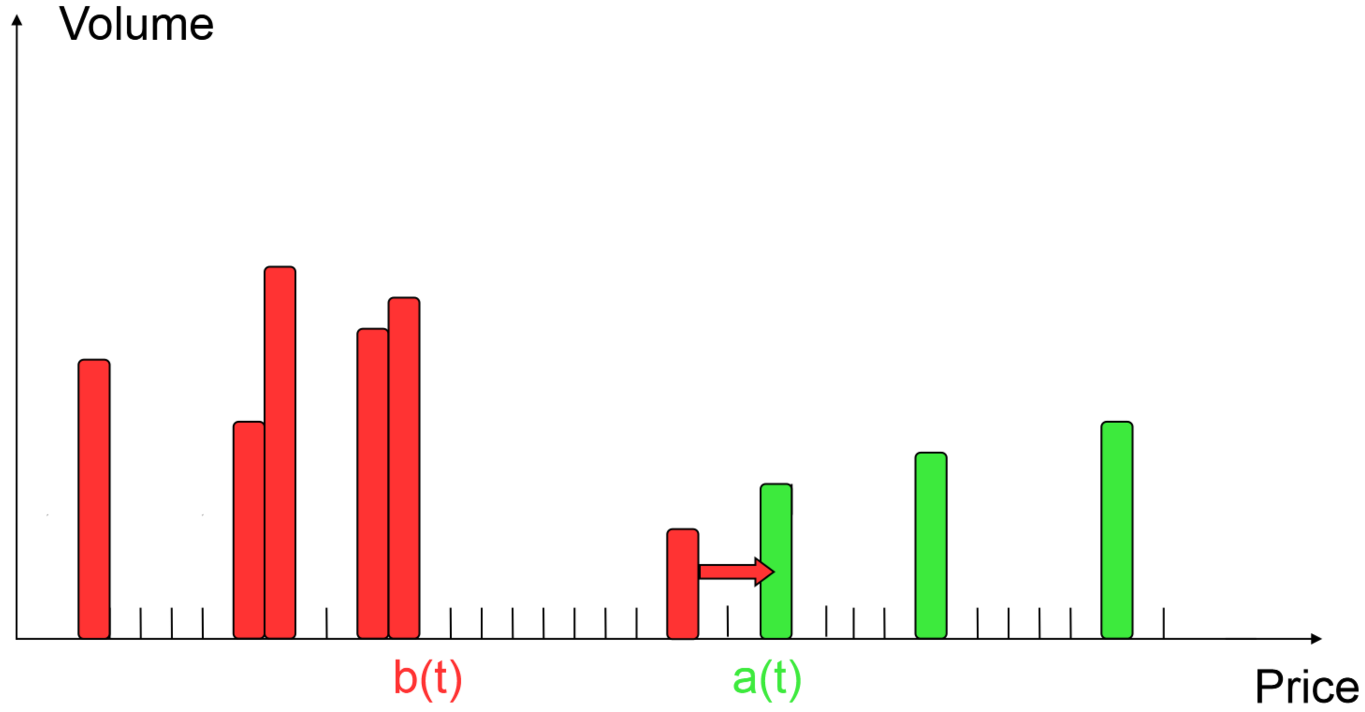


Revealed liquidity

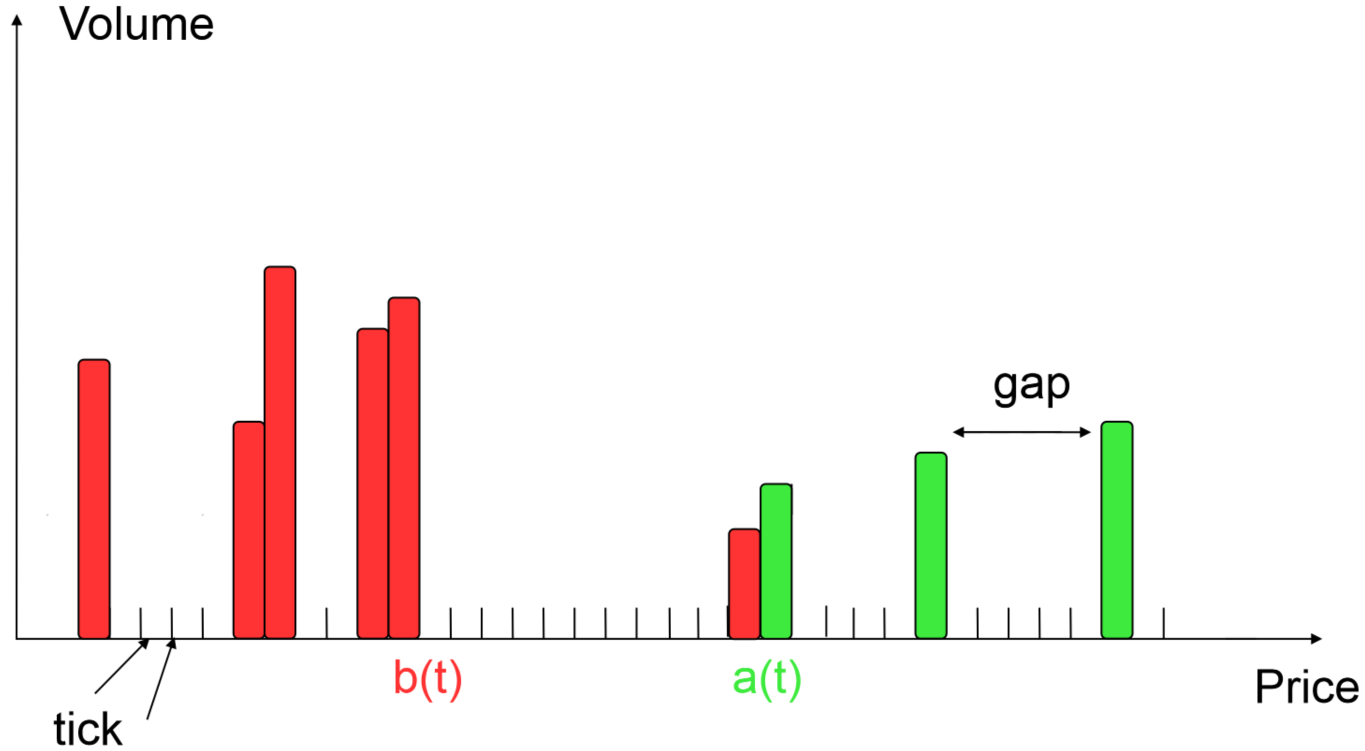
The market order will take liquidity from the ask side of the book.



Revealed liquidity

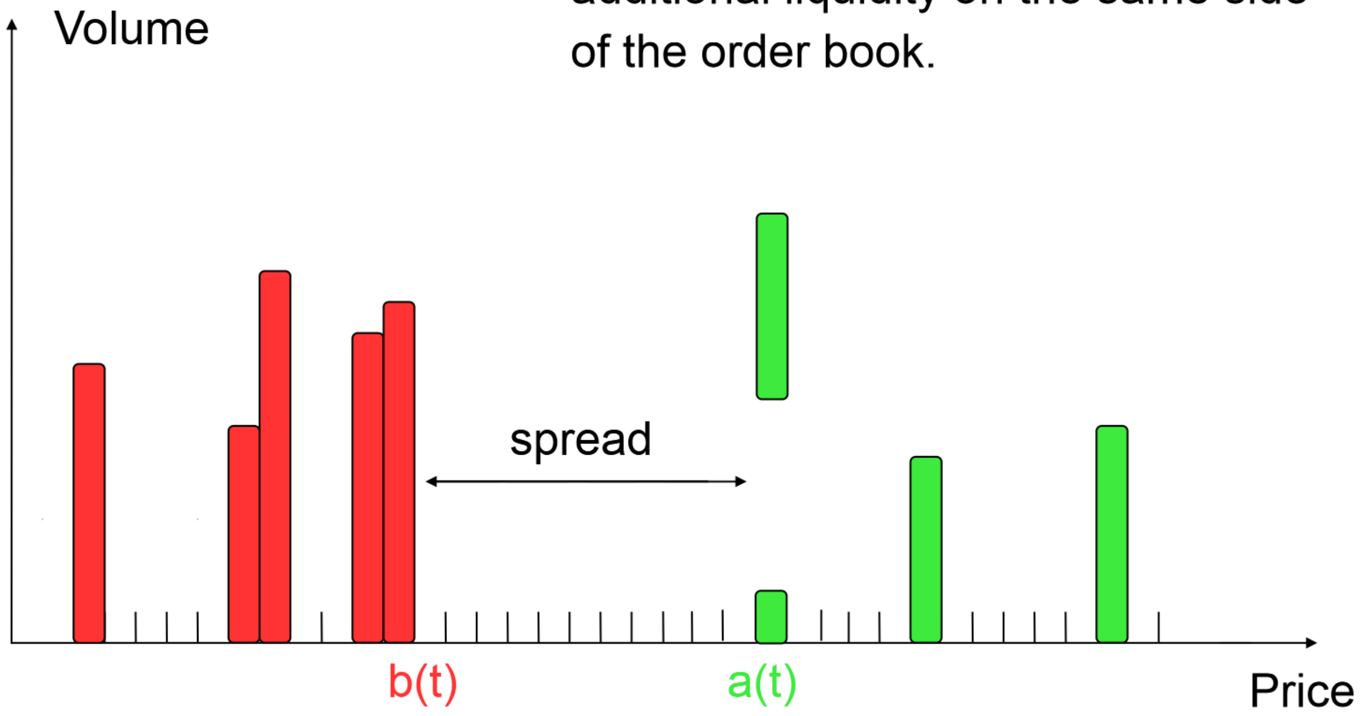


Revealed liquidity



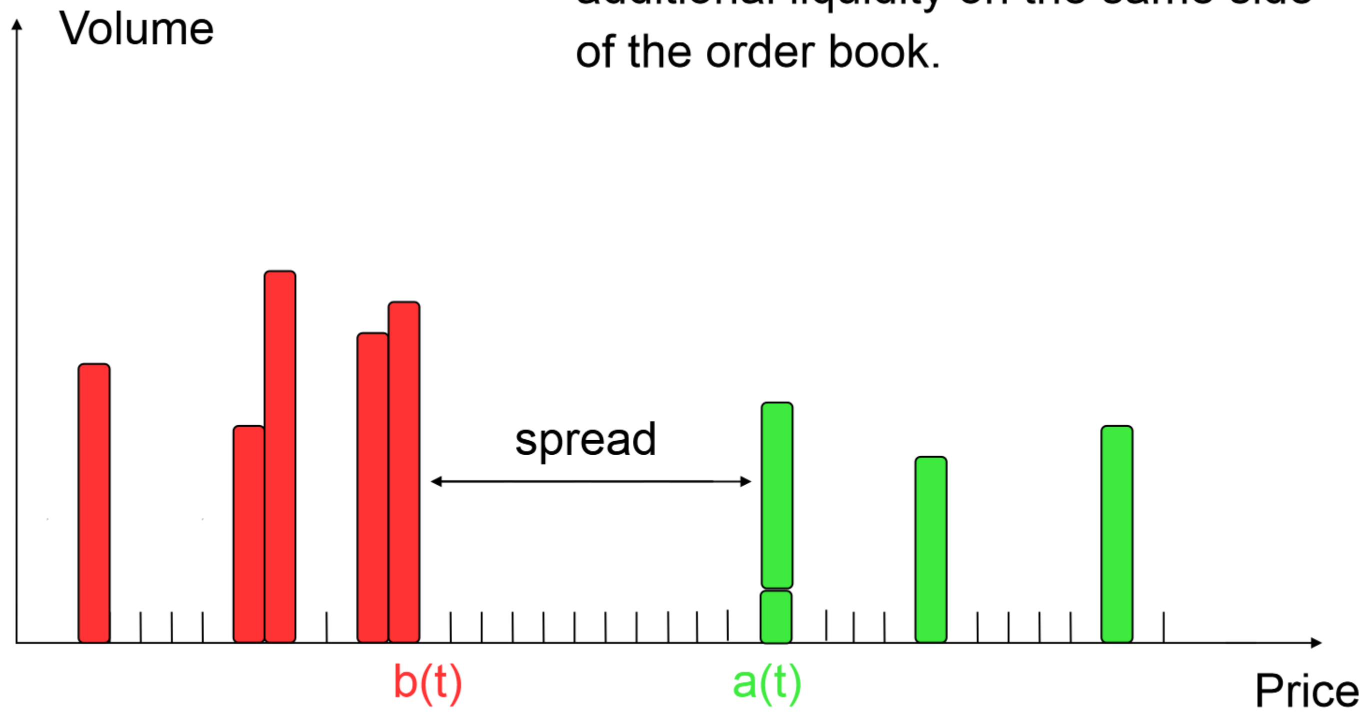
Revealed liquidity

The market reacts providing additional liquidity on the same side of the order book.

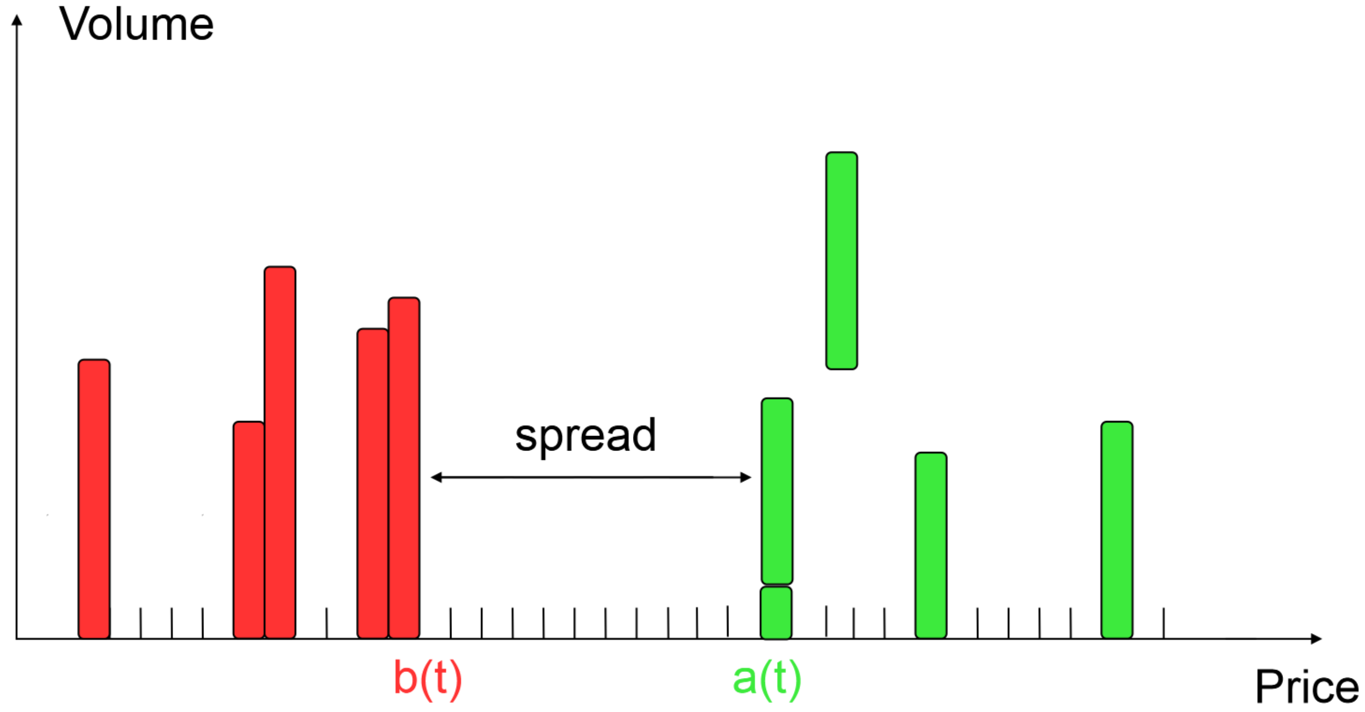


Revealed liquidity

The market reacts providing additional liquidity on the same side of the order book.

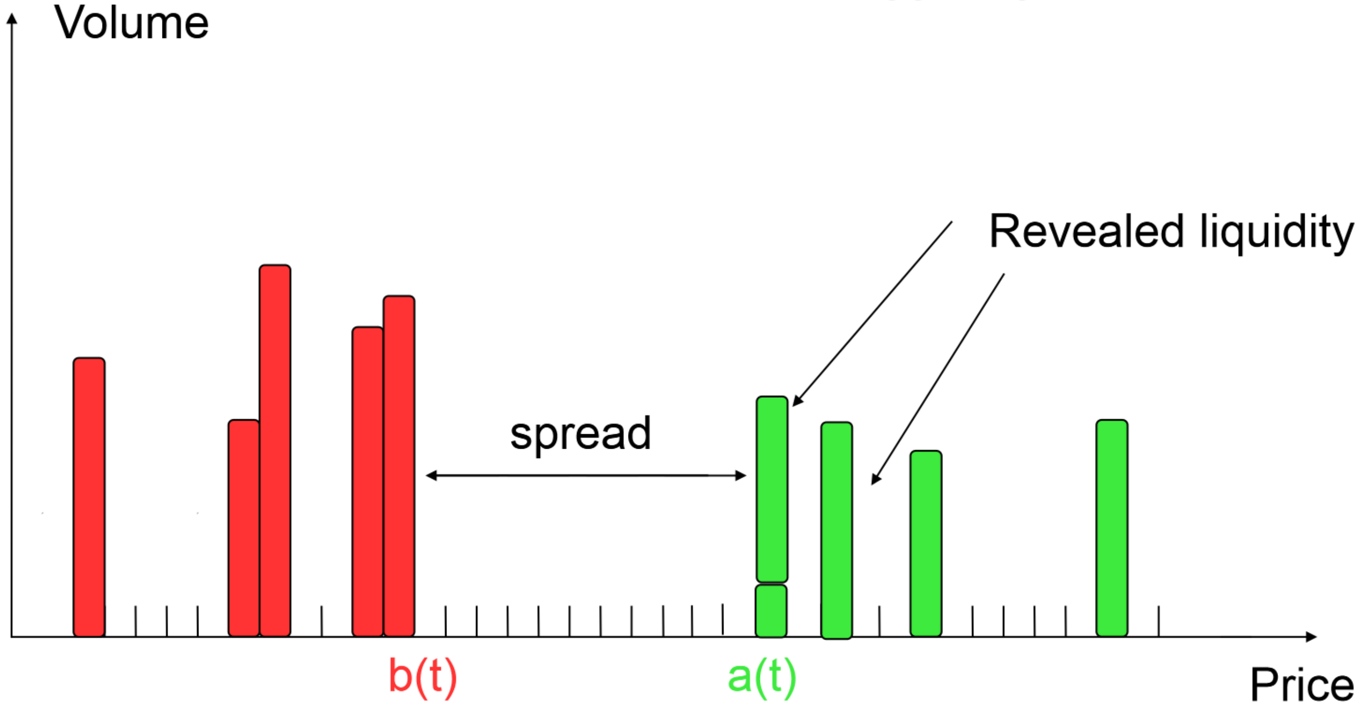


Revealed liquidity

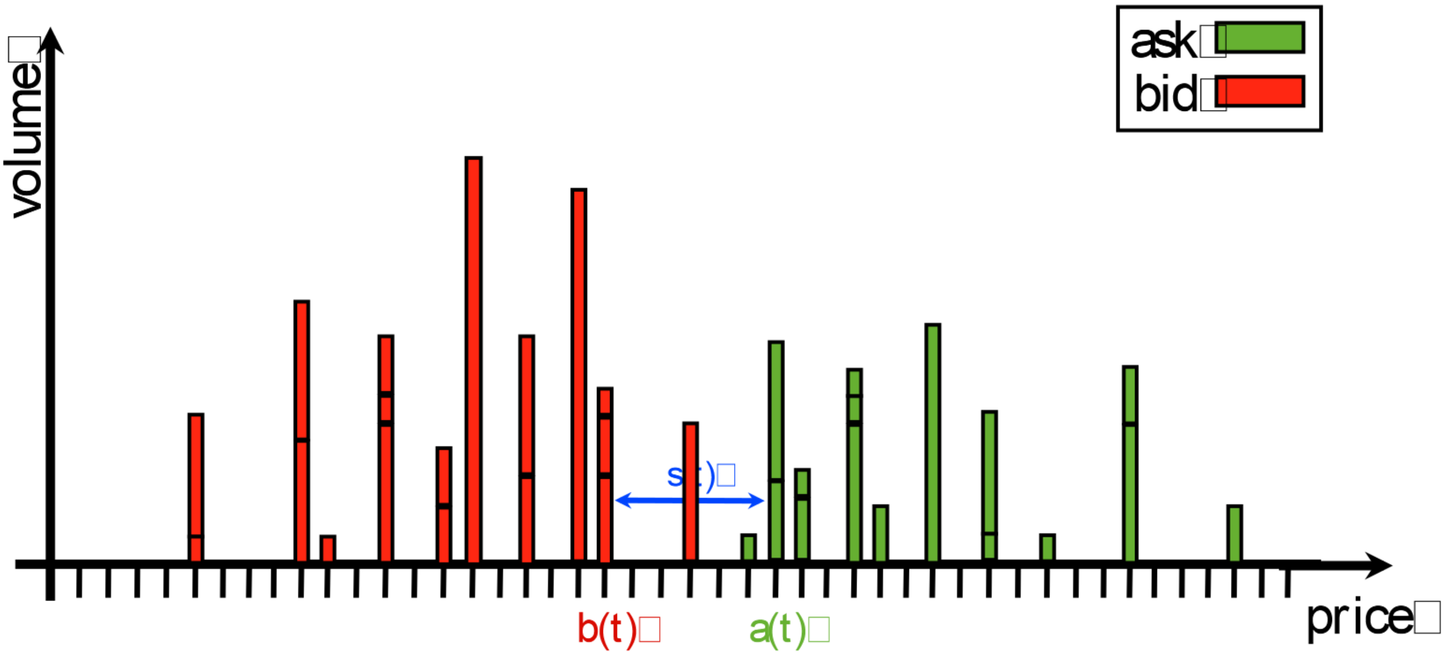


Revealed liquidity

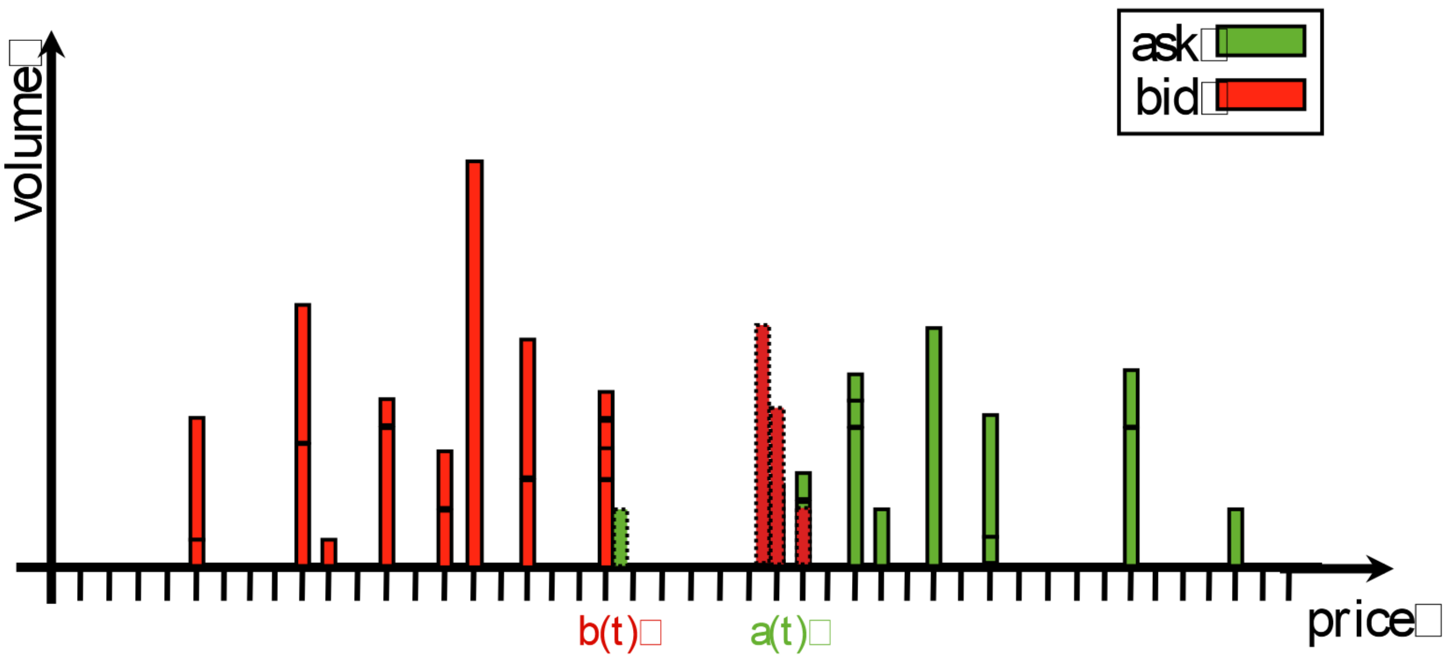
This liquidity was hidden before the arrival of the triggering market order.



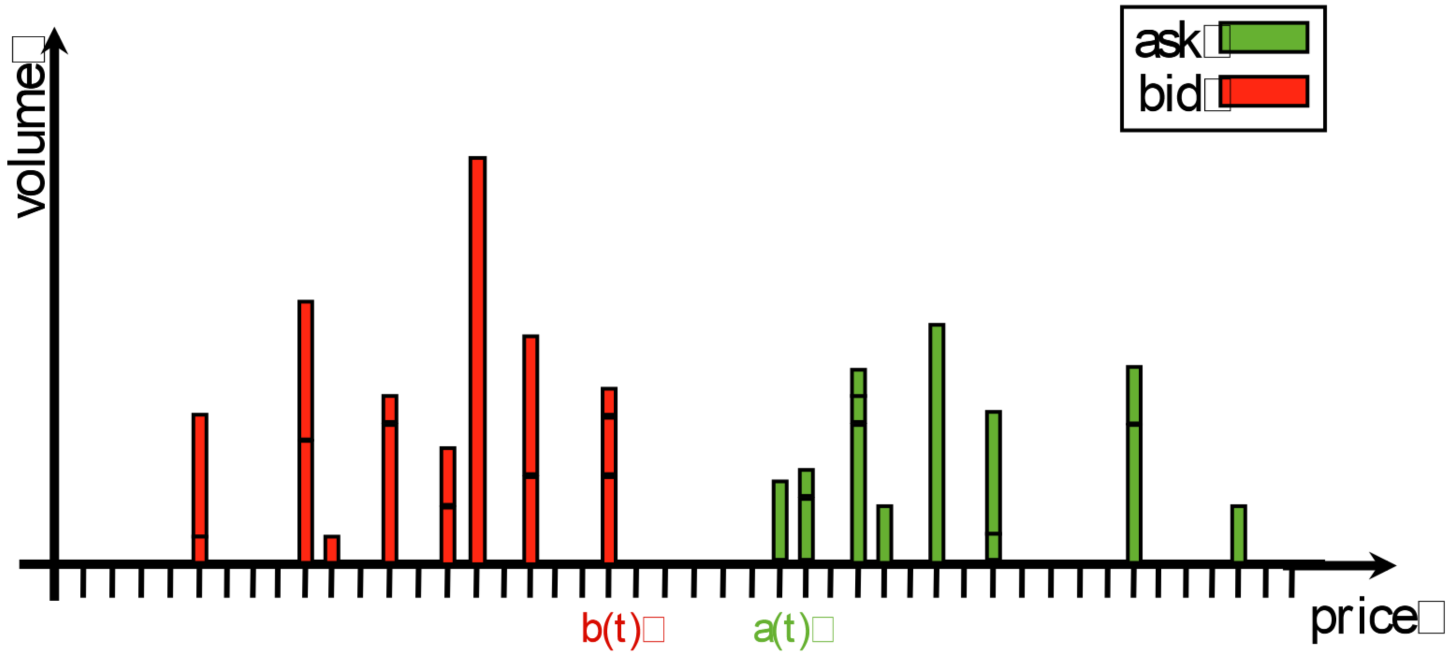
Patient traders: limit orders



Impatient traders: market orders

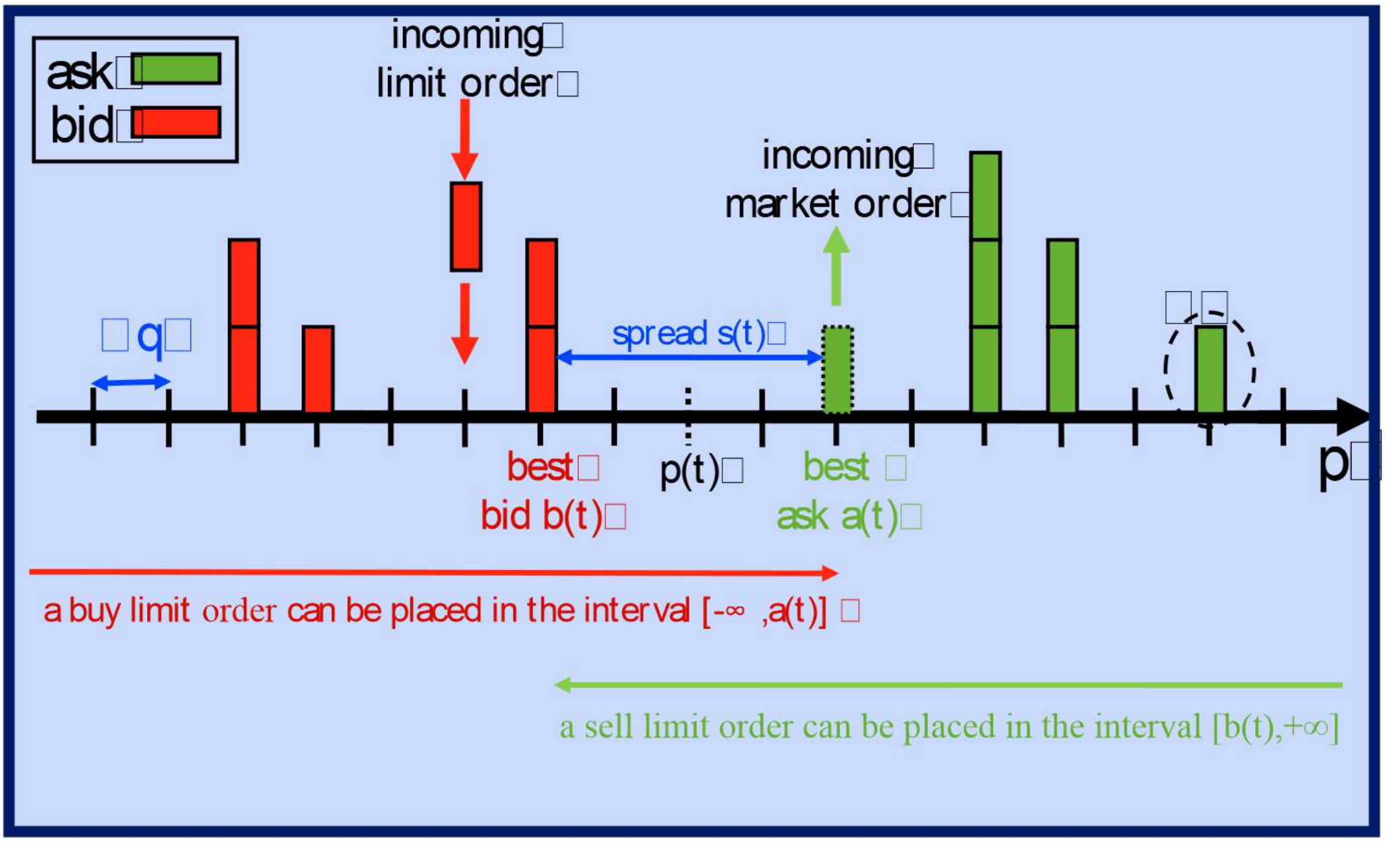


Changing opinion: cancellations



Order Book in a nutshell

(Microscopic version of Farmer’s zero intelligence model)



Order Book regimes

We can identify two different regimes in the order book dynamics

Very liquid market

- Small price variations
- Behavior similar to a continuous system

Illiquid market: discreteness

- Large price variations
- The discreteness of the system is crucial

Order Book Stylized Facts

High frequency data (from 1991 until now)

Model/theory/method/statistical regularities are now investigable in a way similar to the research in physics

Models can be tested

In some sense the order book problem is scientific: some Laws exist.

The amount of information contained by high frequency data is similar to the one from biology or from a modern particle collider such as LHC

We need new statistical tools to extract information in such a huge amount of data: data mining, ...

Order book profile

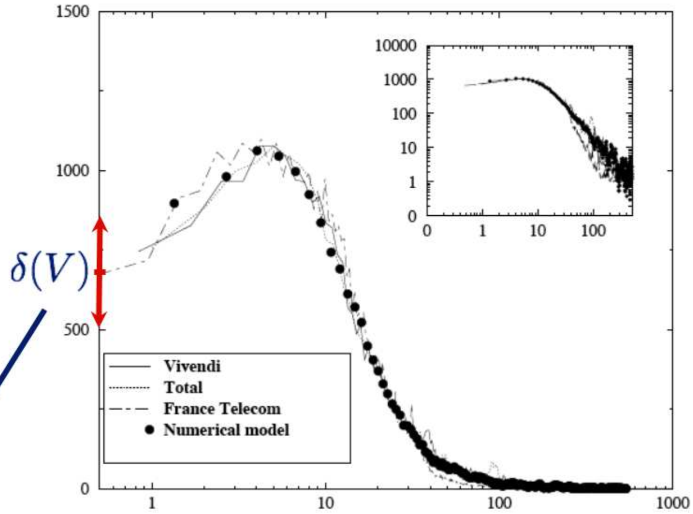


Figure 2: Average volume of the queue in the order book for the three stocks, as a function of the distance Δ from the current bid (or ask) in a log-linear scale. Both axis have been rescaled in order to collapse the curves corresponding to the three stocks. The thick dots correspond to the numerical model explained below, with $\Gamma = 10^{-3}$ and $p_m = 0.25$. Inset: same data in log-log coordinates.

$$\delta(V) \propto V^{\gamma-1} \exp(-\frac{V}{V_0}), \quad \gamma \approx 0.7$$

Wild fluctuations

$$\frac{\sigma(V)}{\bar{V}} \sim 1$$

Cancellation rate

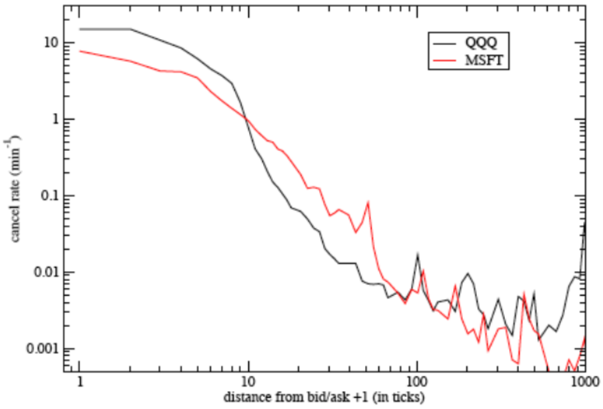
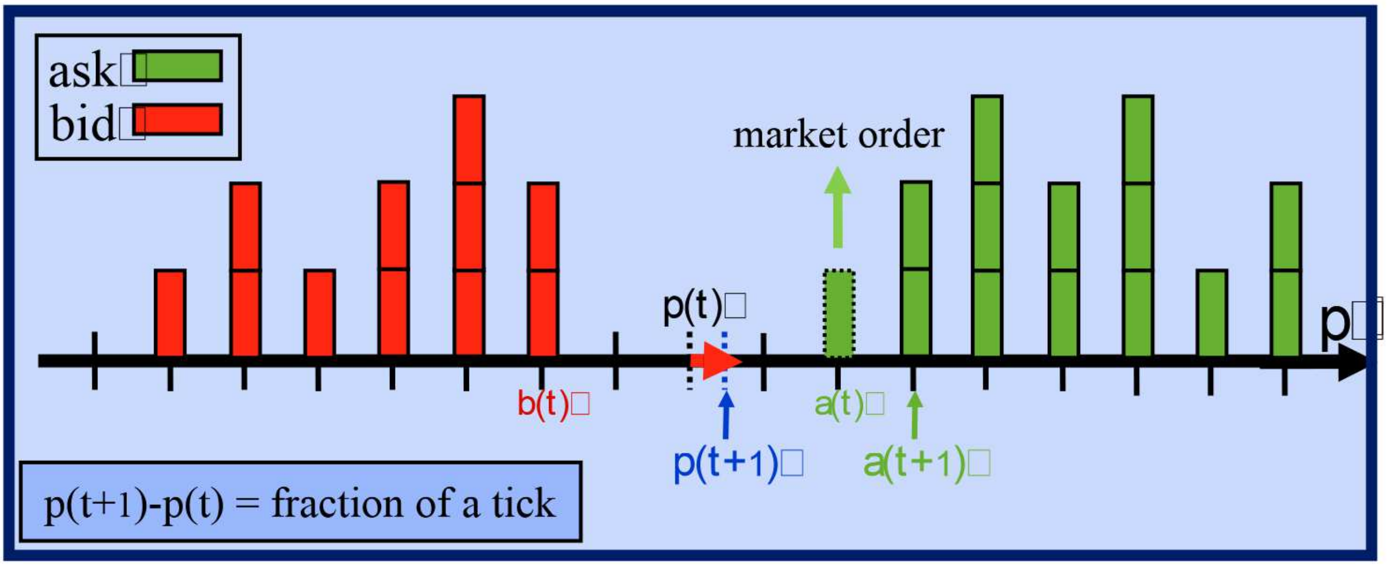


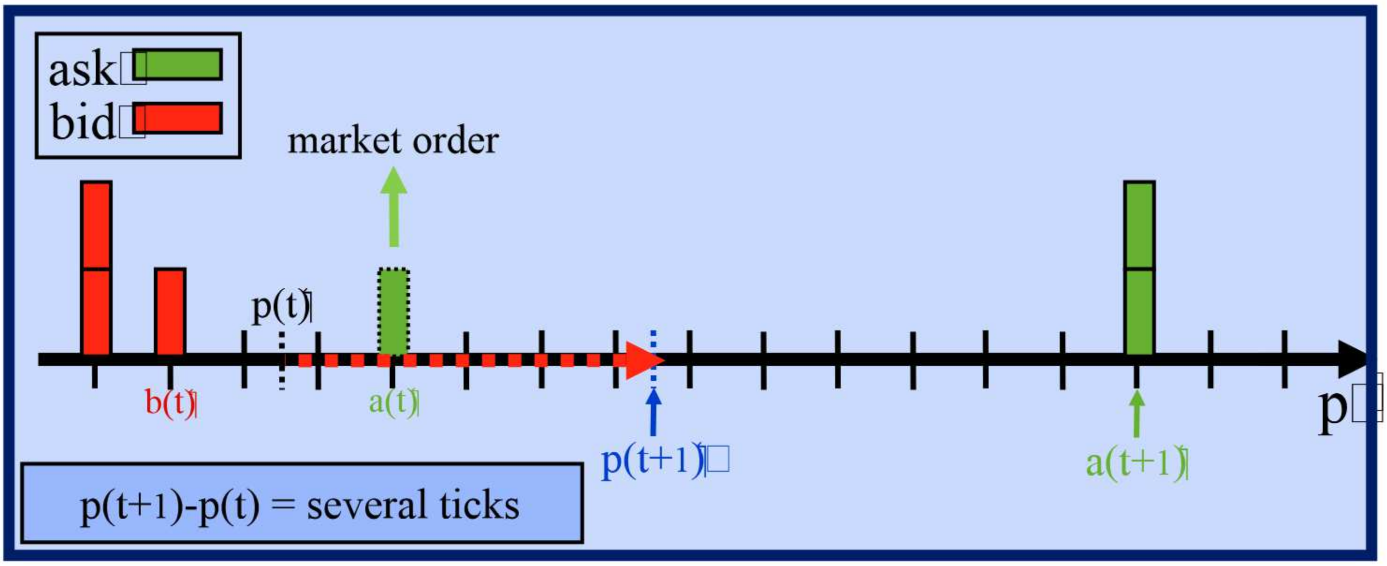
Figure 3: Cancel rate for QQQ and MSFT, as a function of the distance Δ from the current bid (or ask). Note that the cancel rate at the bid/ask is very high (10 per minute), which suggests that most of the orders are automated. Note that the execution rate is only 22% of the cancel rate for QQQ, and 40% for MSFT.

Evidence of automated trades

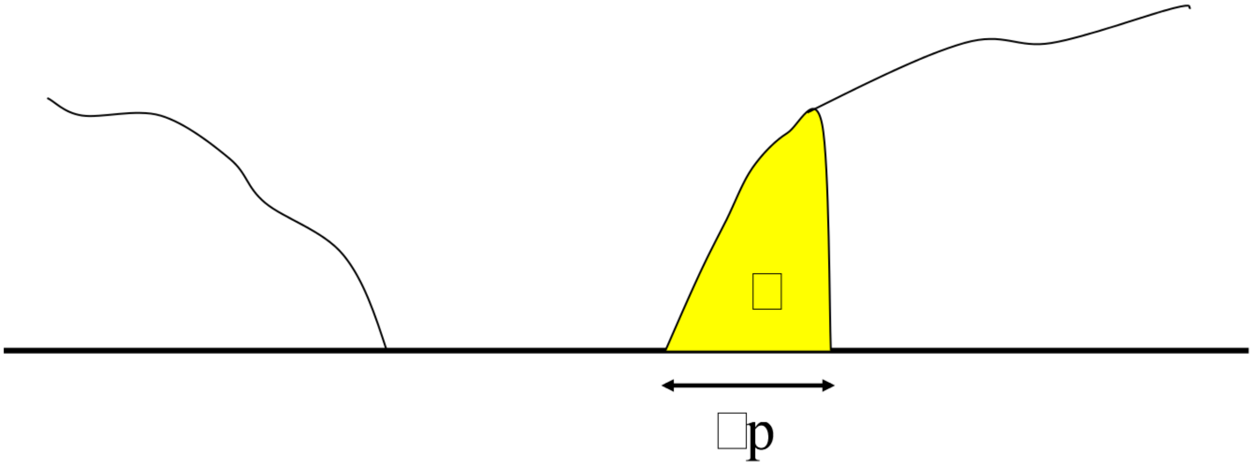
Liquid market



Illiquid market



Price Impact Function



Price Impact Function

The Price Impact Function (PIF) can be considered as the response function of a stock, that is ...

If an agent submit a "virtual" market order of volume □ at time t, what will be the average price change at time t+ □?

$$\phi(\omega,\tau) = E[\Delta p(\tau)|\omega] \longrightarrow \phi(\omega,\tau) \approx \psi(\tau)\Phi(\omega)$$

The Price Impact Function of real market is a concave function with respect to the order volume

$\Phi(\omega) \sim \omega^{\beta}$

NYSE, LSE	$\beta \approx 0.2 - 0.4$ small/medium order size	Farmer, Lillo, Mantegna
NYSE	$\beta \approx 0.5$ large order size	Stanley, et al.
Citygroup	$\beta = 0.6$ large order size	Almgren

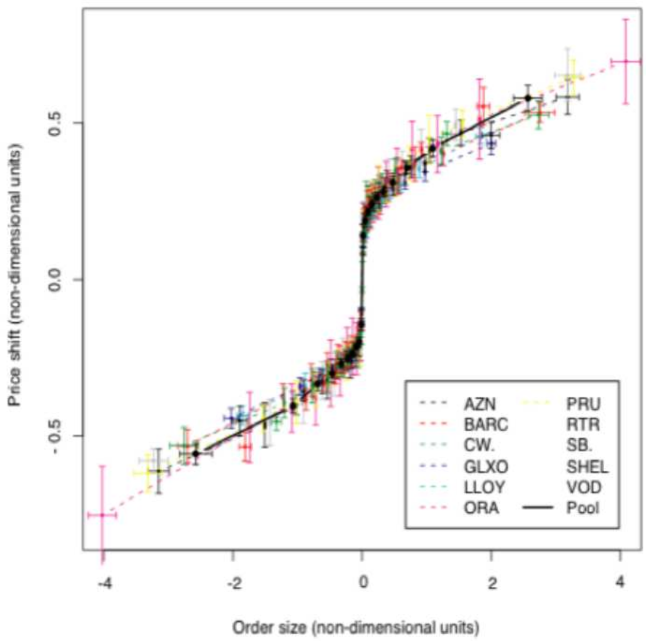
$\Phi(\omega) \sim \ln(\omega)$ small/medium order size	Paris Bourse	Bouchaud, Potters, et al.

Markets are not in a linear response regime

The response is concave (this is a sure feature) but the shape depends on the market, on the order size, on the liquidity, ...

Concavity is due to the shape of the profile, to the response of the investors to a liquidity demand (liquidity provider), ...

$$\beta \approx 0.2 - 0.4$$



Price Impact Surface

We want to study the role played by liquidity/granularity in price response but the normal PIF is calculated averaging on order book configurations with different liquidity/granularity

We define the Price Impact Surface (PIS) which is instead a function of volume and liquidity/granularity

$$E[\Delta p(\Delta t)|\omega, g] = \phi(\omega, \Delta t \to 0, g)$$

where g is a measure of liquidity/granularity

Price Impact Surface - 1

Model result for τ=400 and k=4

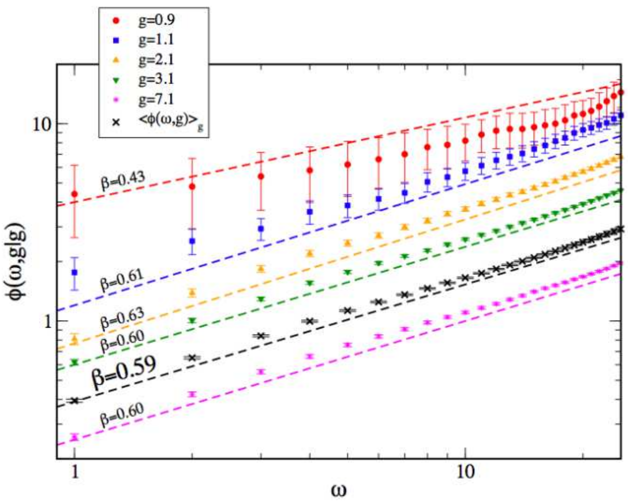
$< \phi(\omega, \Delta t \to 0) >_g$ is concave as one measured in real order book

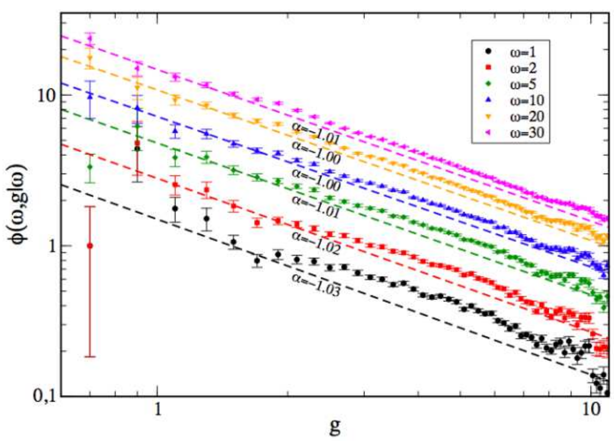
$\phi(\omega, \Delta t \to 0, g|g)$

$\phi(\omega, \Delta t \to 0, g|\omega)$

$< \phi >_g \sim \omega^{0.58}$	$\sim \omega^{0.5 \div 0.7}$

$\sim g^{-0.8 \div -1.0}$

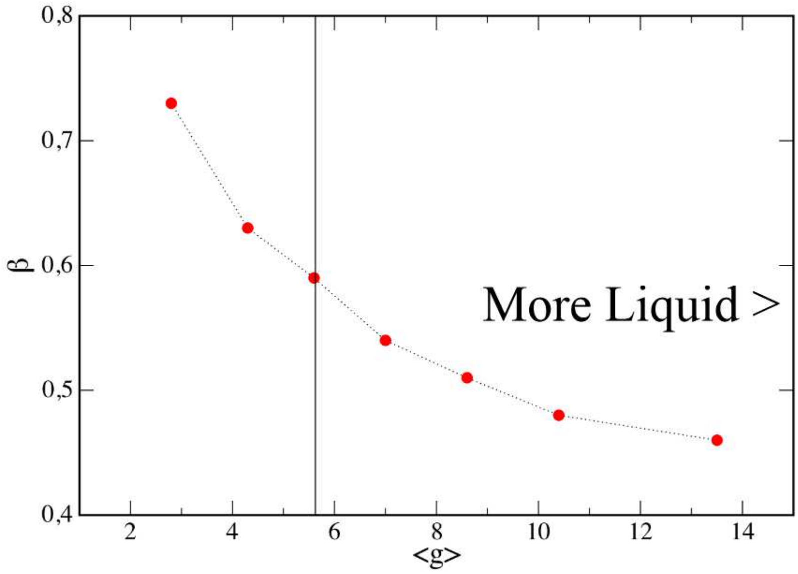




Equity Market Impact Function: Different results ???

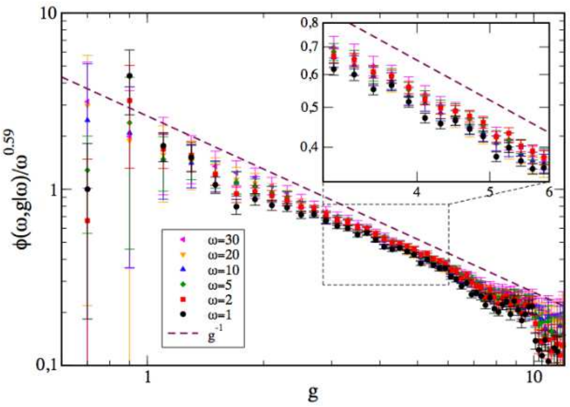
- R. Almgren et al + F. Abergel (Paris): exponent = 0.6
- Lillo, Farmer, Mantegna: exp = 0.2 - 0.4
- J.P. Bouchaud et al: log behavior

Present Model:
0.6 seems to be natural
but prefactor is also
important



Price Impact Surface - 2

If we rescale the PIS with the average impact function $<\phi>_g$ that is proportional to $\omega^{0.58}$ we observe a quasi-collapse in a unique curve (in particular for small values of the order volume)



Relation to ABM:

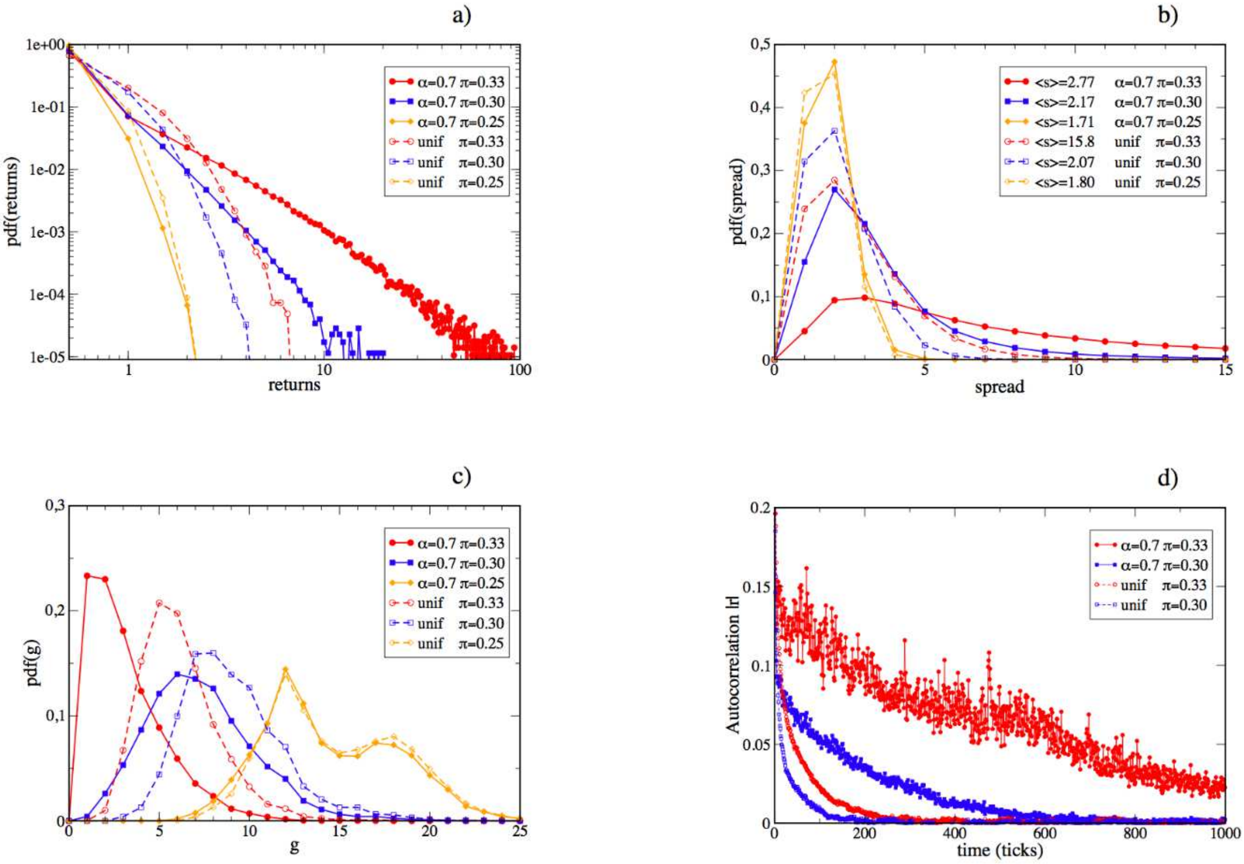
Small N leads to more Sparse orders and more granularity.

N dependent price formation

Therefore the PIS can be approximately factorized as:

$$\phi(\omega, \Delta t \to 0, g) = \Phi(\omega, \Delta t \to 0)\Psi(\Delta t \to 0, g)$$

Stylized Facts of Order Book



Key Concepts:

TO IDENTIFY FROM REAL DATA

- Market sentiment, stabilizing vs destabilizing
- The effective independent agents N* in a market
- Analysis of Herding, Contagion, Correlations
- Liquidity analysis of order book
- Network oriented approach - Direct interaction vs global Trust.
- Coherence problem, similar behavior

BASIC STRATEGIC PROBLEM

- Efficiency vs. Robustness

The Mystery of Zipf's Law and Rank-Size Laws

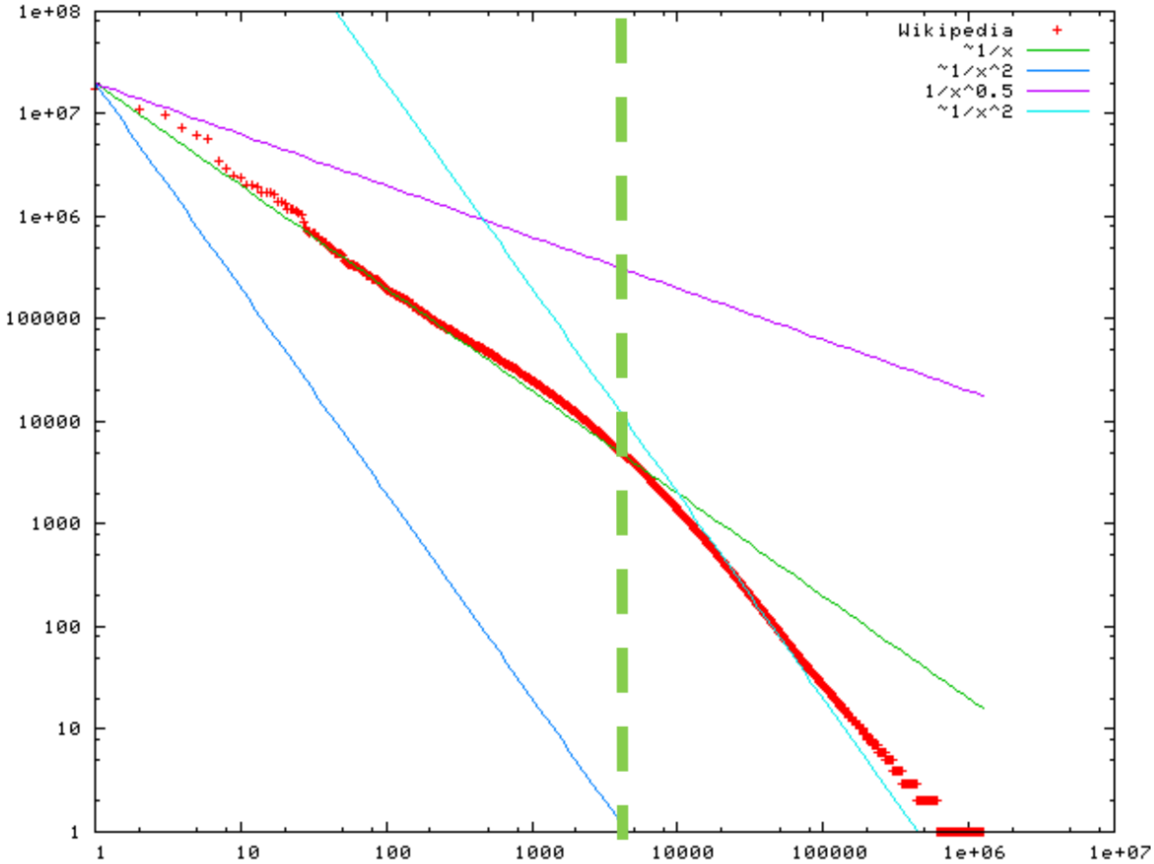
Outline

- Introduction
  - Rank-Size Law (i.e. cities)
  - Zipf's Law (i.e. word frequency)
  - Details matter (works for men but not for women)

- Models for city sizes
  - Simon Model
  - Multiplicative process+ reflecting boundary
  - Inverse square law
  - Coherence and correlations in the sample

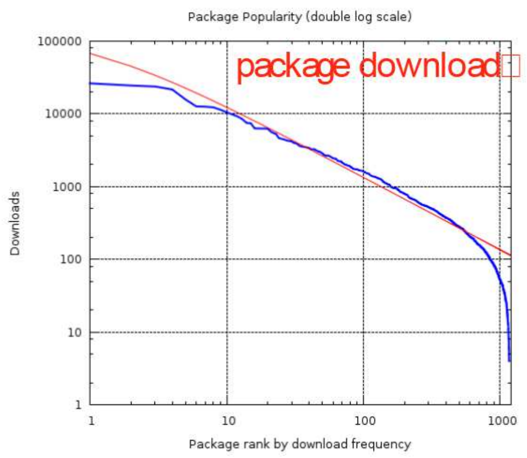
- Microscopic origin of the inverse square law
  - Widespread occurrence of the inverse square distribution
  - Aggregation processes

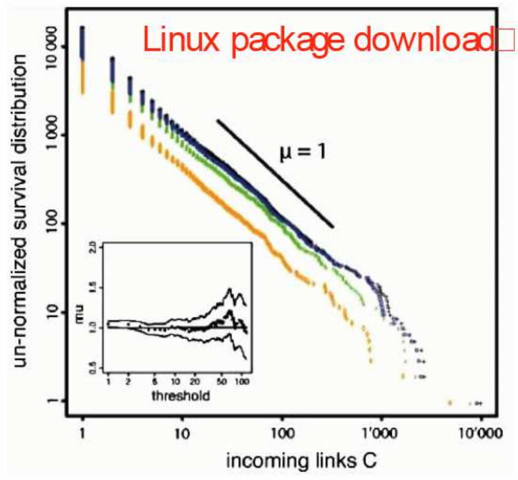
Zipf's Law

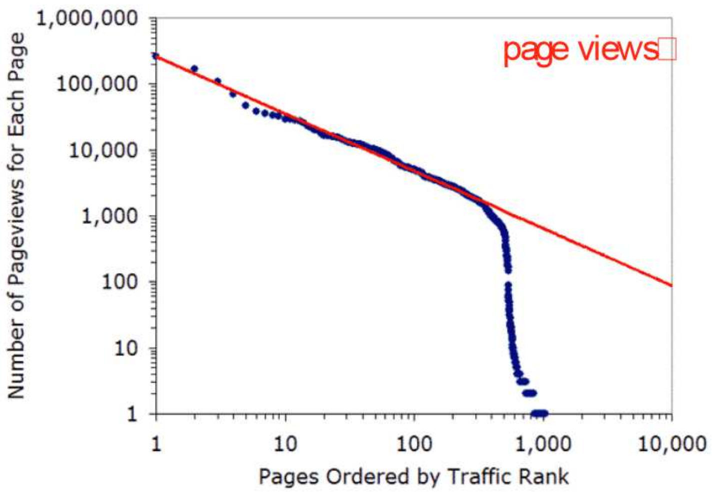


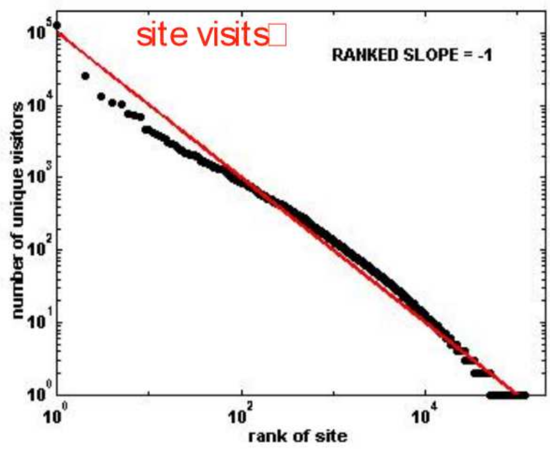
Plot of word frequency in Wikipedia (2006) - source: Wikipedia

Zipfs Law (frequency)

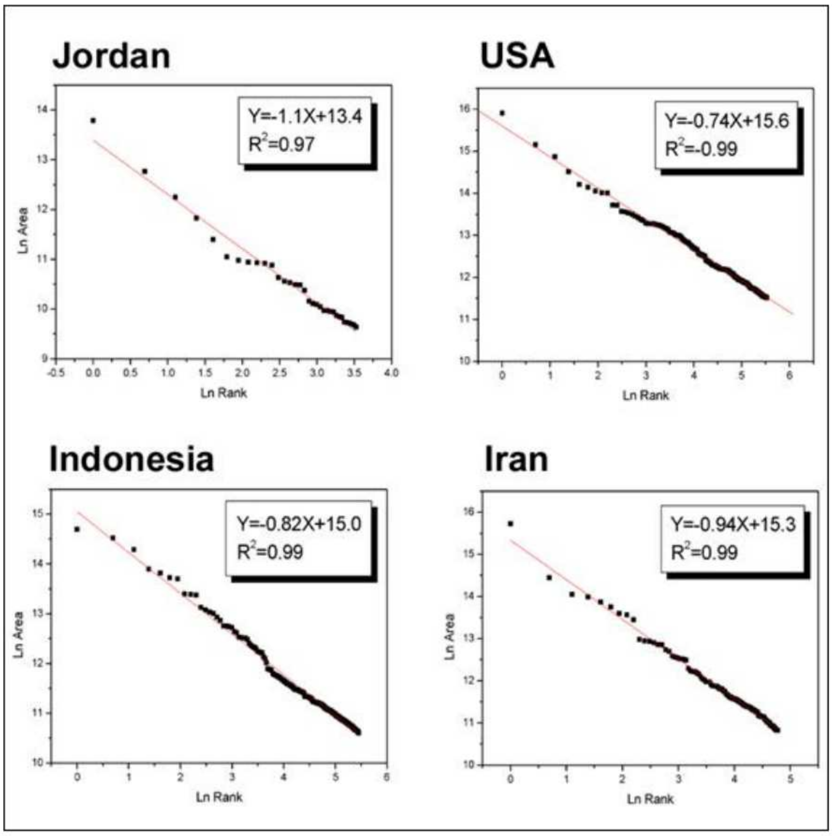


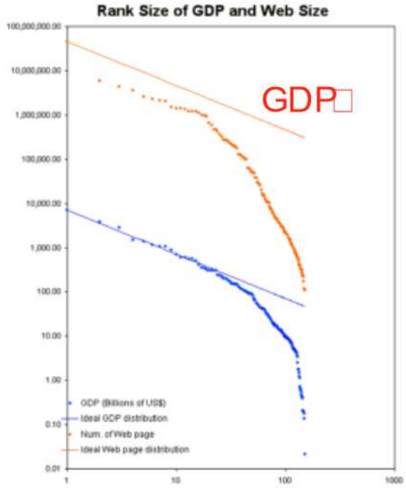


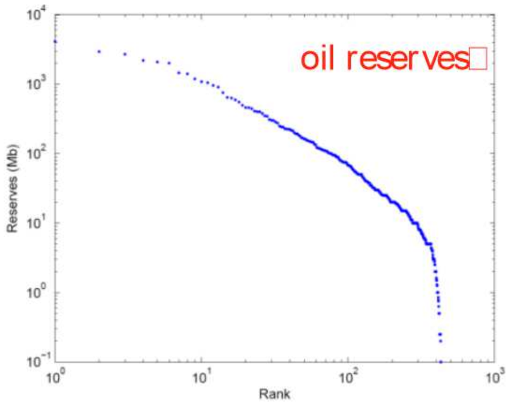




Zipfs Law (Rank-Size)







Zipf's Law $\neq$ Rank-Size Law

Zipf's Law is an empirical law which states the frequencies of many types of data follow a discrete power law probability distribution

Rank-size rules describe the regularities observed in the size of social phenomena when the size is plotted versus the rank

$$f(k;\alpha,N) = H(N,\alpha)\frac{1}{k^\alpha}$$

$$\alpha = 1 \qquad k \in \{1,2,\ldots,N\}$$

$$x(k) = f(k)$$

$$f(k) \sim \frac{1}{k}$$

In both cases k represents the rank but

$\frac{1}{k}$ is a probability

$\frac{1}{k}$ is a size, a lenth, etc

We focus our attention on Rank-Size Law

Non-universality of 1/k behavior: details matter

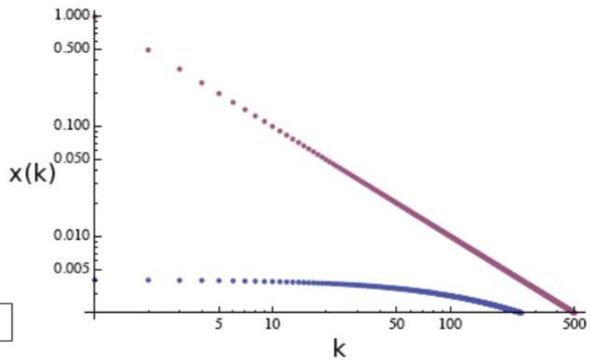
Rank-size distribution or the **rank-size rule** (or **law**) describes the remarkable regularity in many phenomena including the distribution of city sizes around the world, sizes of businesses, particle sizes (such as sand), lengths of rivers, frequencies of word usage, wealth among individuals, etc. All are real-world observations that follow power laws such as those called Zipf's law, the Yule distribution, or the Pareto distribution. If one ranks the population size of cities in a given country or in the entire world and calculates the natural logarithm of the rank and of the city population, the resulting graph will show a remarkable log-linear pattern. This is the rank-size distribution.[1]

If x(k)=1/k holds on a set, the rule cannot hold on a subset or in general on a larger set

For instance on an ordered subset we obtain: $x_M = x_1,\ x_2,\ x_3,\ \ldots,\ x_{k'},\ \ldots,\ x_N = x_m$

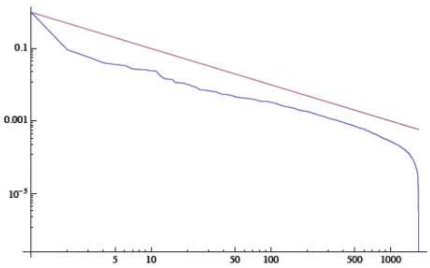
$$x(k) = \frac{x_M}{k} \longrightarrow x(k') \approx \frac{x'_M}{\frac{k'}{k^*}+1}$$

$$k' = k - k^*$$
$$x'_M = x(k')$$

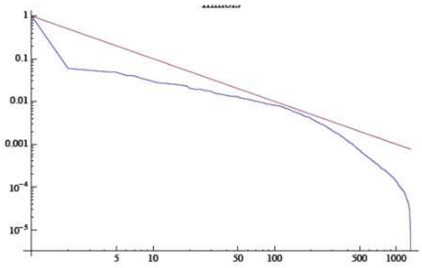


The dataset details matter, 1/k is not a universal behavior

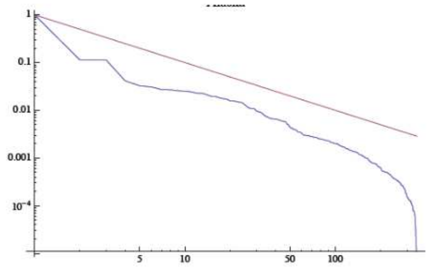
Some pratical examples: USA



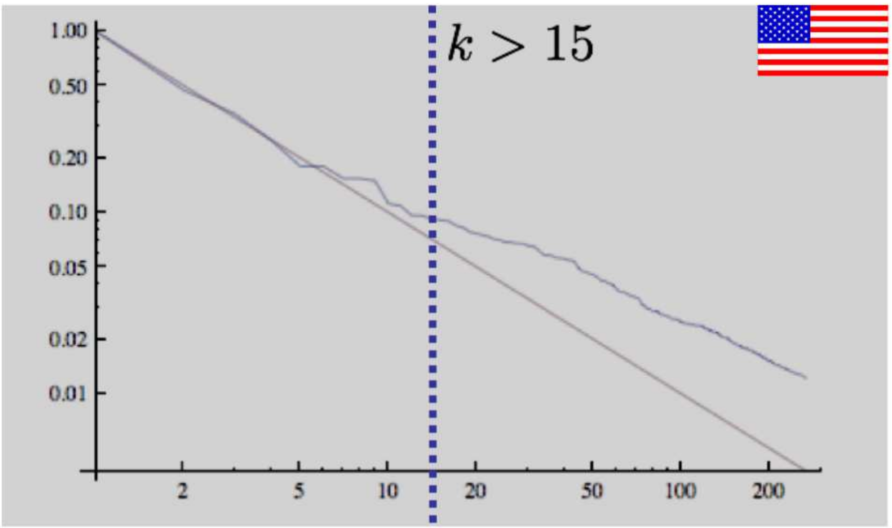
New York



Illinois



Alaska

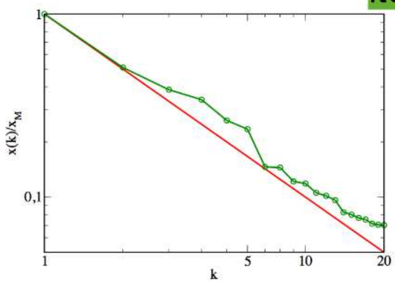


x(k)=1/k approximately holds for United States of America but the rule does not hold in each state

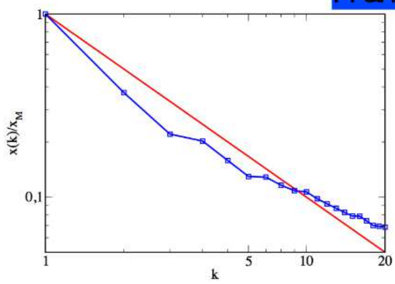
USA are the "right" set

Some pratical examples: EU

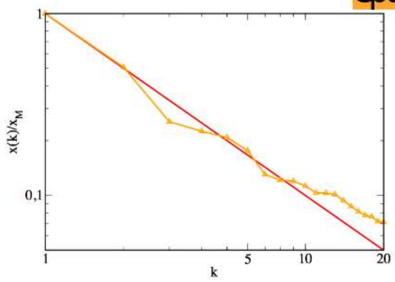
Italy

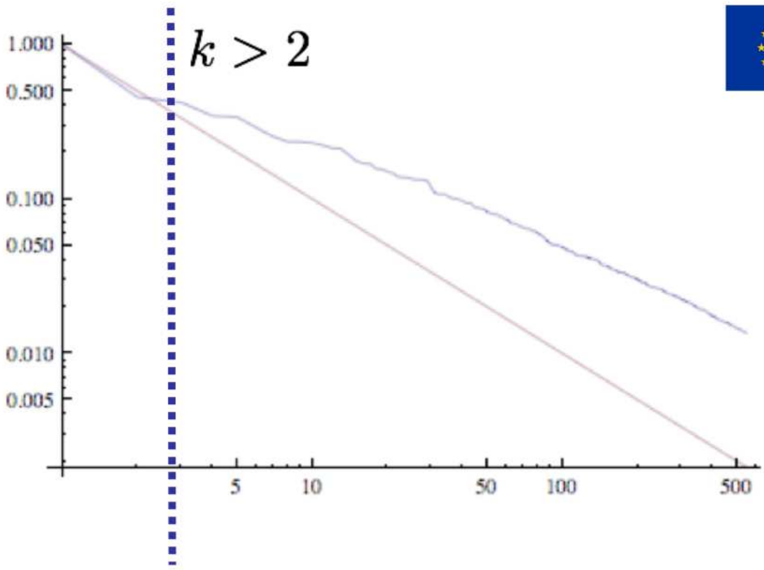


France



Spain





x(k)=1/k approximately holds for each european country, the power law behavior is no more observed once the countries are aggregated in the EU as expected

Replica effect

France, Italy, Spain, etc, are the "right" set

EU and USA have a difference development

Models

Two mainstreams

Simon Model

Multiplicative process repelled from the origin

Gibrat's Law

Simon growth model can be interpreted in the framework of coagulation processes

An argument for Rank-Size Rules 1/k

L. Pietronero, E. Tosatti, V. Tosatti, A. Vespignani, Physica A, 293, pp 297-304 (2001)

Hp: rank-size rule can be interpreted in terms of the exponent of the underlying pdf for city size

$$\{x_1, x_2, \ldots, x_N\} \quad \text{drawn from} \quad f(x) \sim \frac{1}{x^\alpha}$$

$$x_{max} = x(k=1) \qquad x \in [x_m, x_M]$$

The rank k of the x(k) number is given by all the numbers whose value is between x(k) and x_{max}

$$k = N \int_{x(k)}^{x_{max}} f(x)\mathrm{d}x \longrightarrow x(k) \sim k^{1/(1-\alpha)}$$

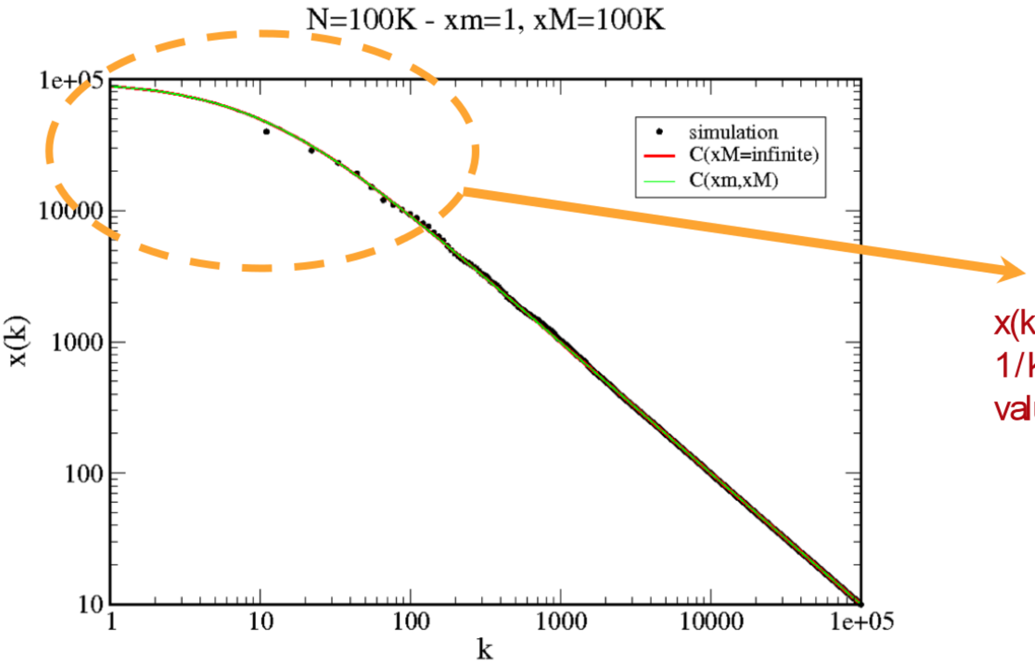
$$\alpha = 2 \longrightarrow 1/k$$

but....

An argument for Rank-Size Rules 1/k

x_{max} cannot be neglected:

$$x(k) = \frac{C}{\frac{k}{N} + \frac{C}{x_{max}}} \qquad C = \frac{x_M\, x_m}{x_M - x_m}$$



x(k) fails to reproduce the 1/k behavior for small values of k (i.e. big cities)

Backward problem

Direct problem: given the underlying pdf we search for the rank-size rule

$$f(x) \rightleftarrows x(k)$$

Backward problem: given the rank-size rule 1/k we want to find the pdf for x

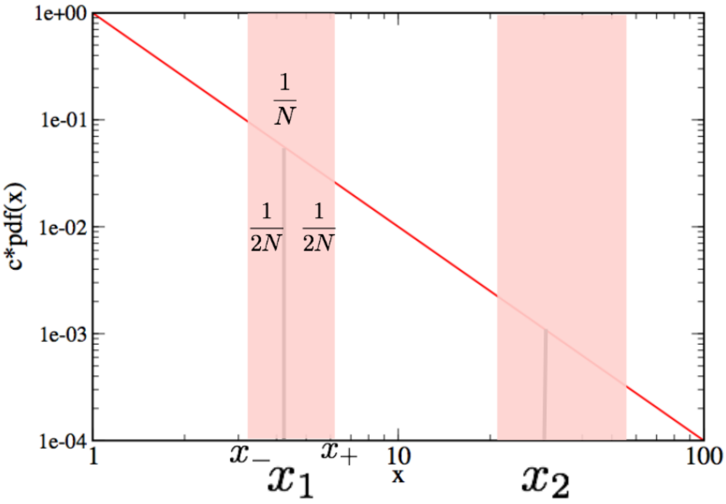
$$k - 1 = NC \int_{x(k)}^{\hat{x}_M} f(x)\mathrm{d}x$$

The direct problem is meaningful from a mathematical point of view since the range of x is fixed by the pdf range.

The backward problem fails since x(k) has not a priori an inferior bound

The pdf must depends on the set size and must know the values of the previous draws → conditioned draws

Model for Rank-Size Rules: conditioned draws



Big cities have a strong screening effect

Unconditioned and conditioned draws nearly coincide for small cities

0. $f(x) = \frac{C}{x^2}$ $\quad x \in [x_m, x_M]$

1. x_1 is drawn from f(x)

2. $$\frac{1}{2N} = C \int_{x_1}^{x_+} \frac{1}{x^2}\, dx$$
$$\frac{1}{2N} = C \int_{x_-}^{x_1} \frac{1}{x^2}\, dx$$

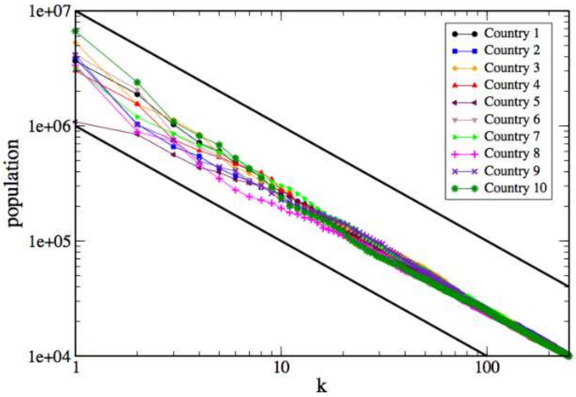
3. $$f_2(x) = C_2 / x^2$$
$$x \in [x_m, x_-] \cup [x_+, x_M]$$

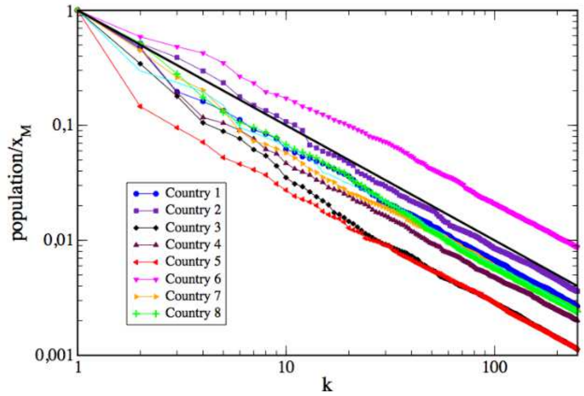
4. x_2 is drawn from f_2(x)

⋮

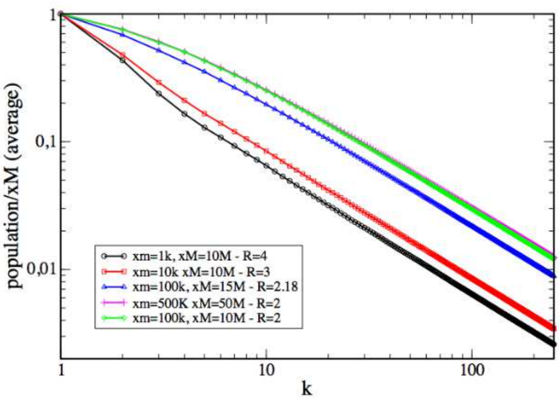
N. x_N is drawn from f_N(x)

Model for Rank-Size Rules: some results



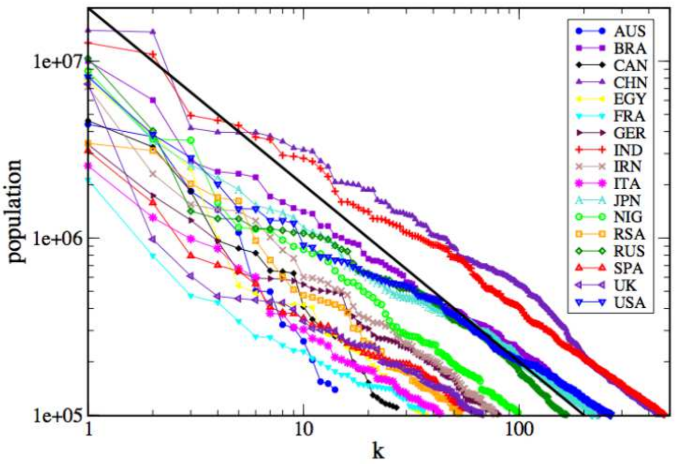


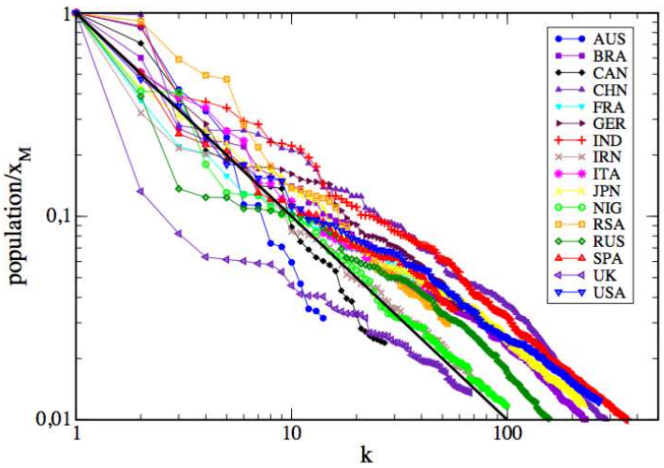
$$R = \log \frac{x_M}{x_m}$$



Cities around the world

Data from Mathematica online database

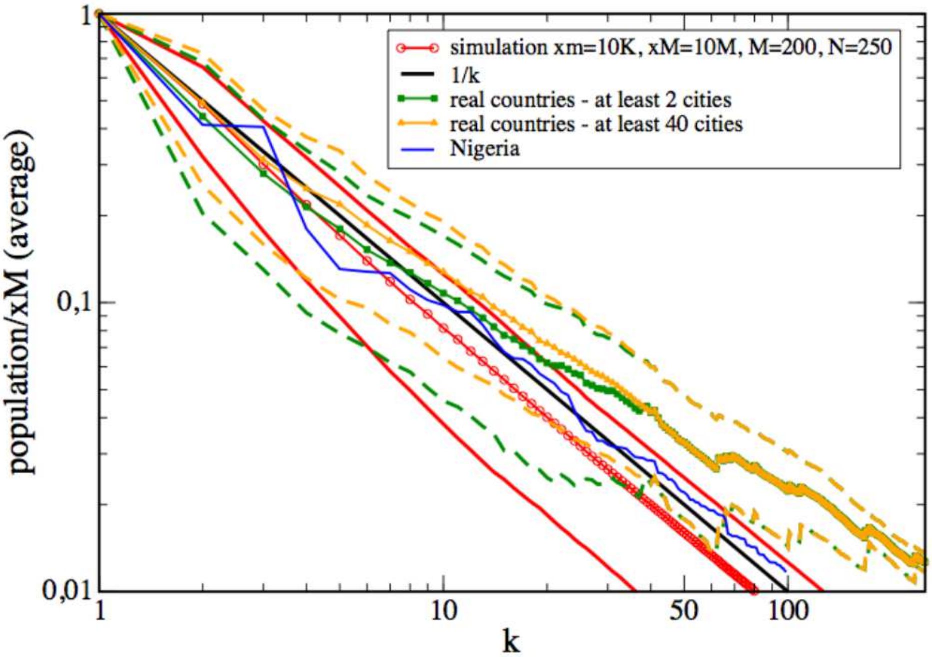




Nigeria!

We perform an average over all the countries in order to compare real countries with the results of our model

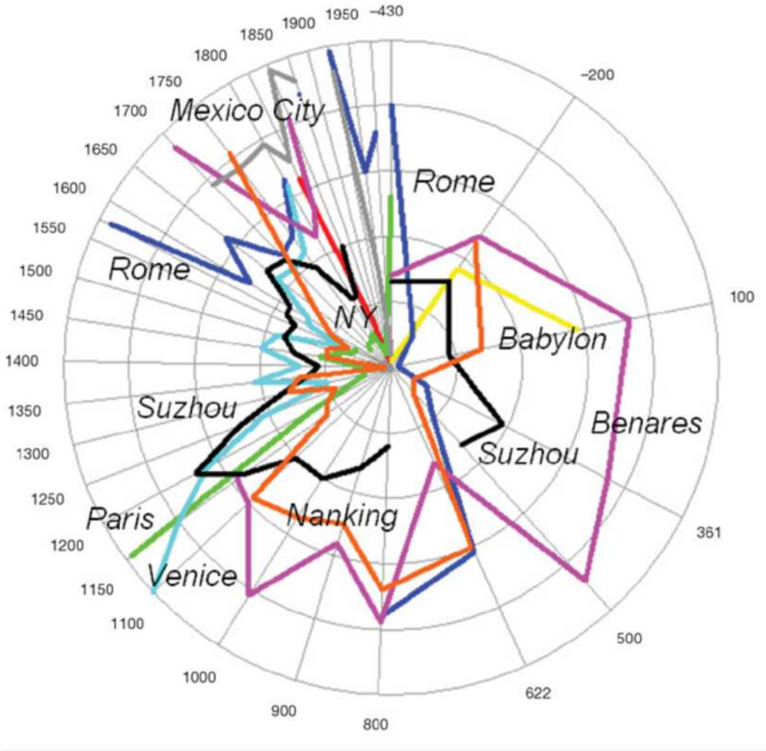
A model for Rank-Size Rules: some results



Rank-size Laws vs time

M. Batty, Nature 444, pp 592-596

Today rank-size rules for cities approximately hold? Is this still true in the past?



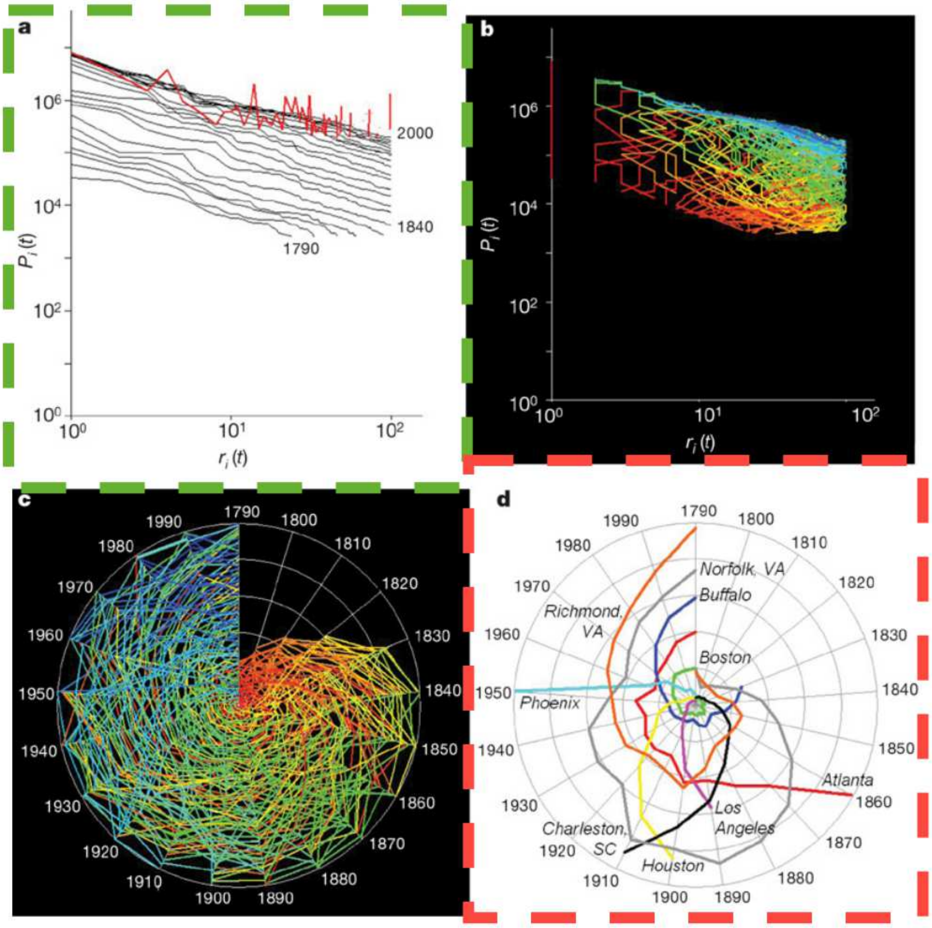
Rank-size Laws vs time: US urban system 1790-2000

M. Batty, Nature 444, pp 592-596

The rank-size rule appear to be stable in time ...

... but the rank of each city is not constant

The time evolution is not trivial

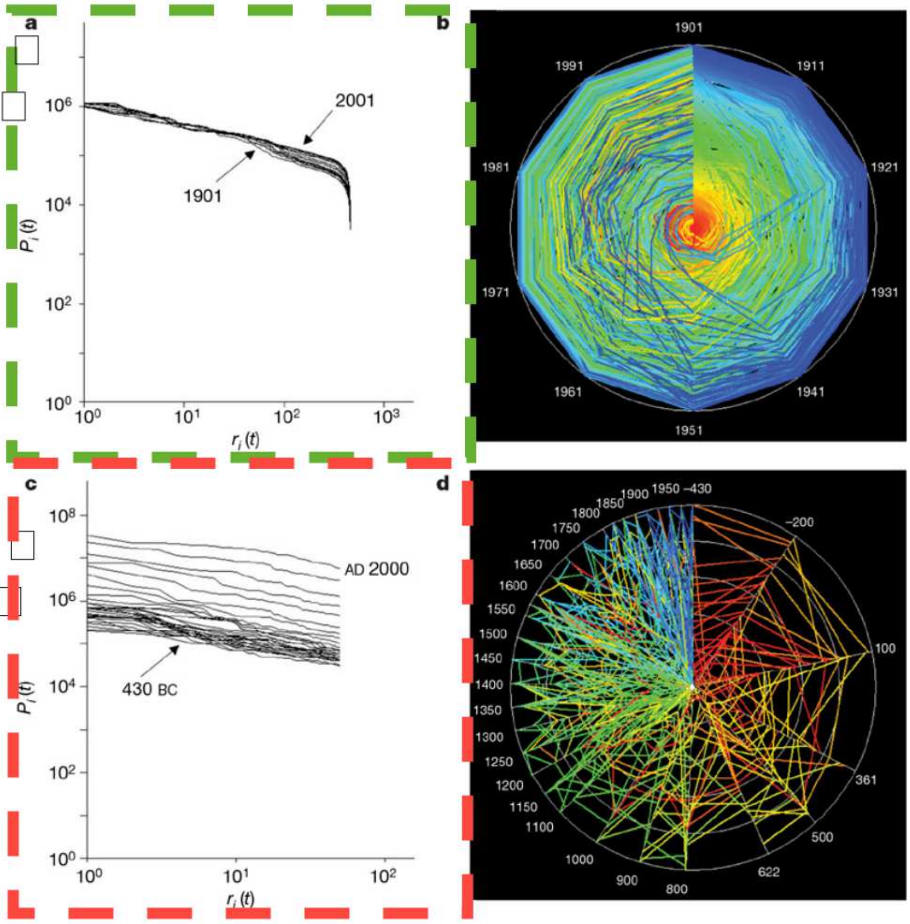


Rank-size Laws vs time: UK and world urban system

M. Batty, Nature 444, pp 592-596

UK
1901-2001

WORLD
430B.C. -2000 A.D.



Widespread occurence of the inverse square law

G. Caldarelli, C. C. Cartozo , P. De Rios, V. Servedio, Phys. Rev. E, 69, (2004)

- Super-critical tree/network
- Random division of the unit interval
- Phenomena organized in hierarchical subgroup
  - taxonomy trees
  - drainage basins of rivers

- Simon and Yule processes
- Aggregation processes - Smoluchowski's coagulation equation
  - cluster-cluster aggregation
  - DLA cluster cluster aggregation
- ?Matter density?

All these phenomena are characterized by aggregation growth or by successive fragmentations

Microscopic origin of the inverse square law

Smoluchowski's coagulation equation = mean field equation for cluster-cluster aggregation

$$\dot{c}_k = \frac{1}{2} \sum_{i+j=k} K_{ij} c_i c_j - c_k \sum_{j=1}^{\infty} K_{kj} c_j$$

If there are not any sinks and sources for particles, the system does not have a stationary states

We can add sinks and sources → the stationary state exists

W. H. White, Journal of Colloid and Interface Science, 87, (1982)

$$0 = \frac{1}{2} \sum_{i+j=k} K_{ij} c_i c_j - c_k \sum_{j=1}^{\infty} K_{kj} c_j - \beta_k c_k + \alpha_k$$

$$K_{ij} \propto i^{\beta} j^{\beta}$$
$$\alpha_k = \alpha_1 \delta_{1,k}$$
$$\beta_k = 0, \ \forall k$$

$$c_k \propto k^{-\frac{3}{2}-\beta}$$

Simon's model is very similar: monodisperse source and no sinks

Conclusion and perspectives

- Role of the set/subset selection
  - cities
  - other phenomena
- Role of the definition of the border of a city
- Analysis of rank-size plot versus time: dynamics
- Microscopic origin of the inverse square law
  - aggregation/fragmentation processes
  - Smoluchoswki coagulation equation
  - Is $1/x^2$ an attractor for aggregation processes? how strong? for which regime?
  - aggregation/fragmentation processes in 2D embedded in the country “topology” (distance, radius of influence of a city, mountain, plain, etc)