CentraleSupélec

# A survey on limit order books

A course given by Frédéric Abergel during the first
**Latin American School and Workshop on Data Analysis and Mathematical Modelling of Social Sciences**

Based on the book
*Limit order books*
by F. Abergel, M. Anane, A. Chakraborti, A. Jedidi and I. Muni Toke

# Why a microscopic description of financial markets ?

- Classical mathematical modelling of financial assets
  - directly describes the price as a stochastic process
  - imposes drastic limitations on *trading strategies* and *agent behaviour*

# Why a microscopic description of financial markets ?

- Classical mathematical modelling of financial assets
  - directly describes the price as a stochastic process
  - imposes drastic limitations on *trading strategies* and *agent behaviour*
- The information contained in high-frequency financial data allows one to
  - relate the price evolution to the *microstructure* of the market
  - explore the strategies of financial agents

# Why a microscopic description of financial markets ?

- Classical mathematical modelling of financial assets
  - directly describes the price as a stochastic process
  - imposes drastic limitations on *trading strategies* and *agent behaviour*
- The information contained in high-frequency financial data allows one to
  - relate the price evolution to the *microstructure* of the market
  - explore the strategies of financial agents

**Motivation**

A description at the **order level** provides a much better understanding of financial markets

A single trading day generates as many data points as **100 years** of close-to-close data

# Why a microscopic description of financial markets ?

New paradigm for high frequency finance

- volatility is observable
- order flow is observable
- agent strategies are (partially) observable

# What is a limit order book ?

- The **limit order book** is the list, at a given time, of all buy and sell limit orders, with their corresponding prices and volumes
- The order book evolves over time according to the arrival of new orders

# What is a limit order book ?

- The **limit order book** is the list, at a given time, of all buy and sell limit orders, with their corresponding prices and volumes
- The order book evolves over time according to the arrival of new orders
- 3 main types of orders:
  - **limit order**: specify a price at which one is willing to buy (sell) a certain number of shares
  - **market order**: immediately buy (sell) a certain number of shares at the best available opposite quote(s)
  - **cancellation order**: cancel an existing limit order

# What is a limit order book ?

- The **limit order book** is the list, at a given time, of all buy and sell limit orders, with their corresponding prices and volumes
- The order book evolves over time according to the arrival of new orders
- 3 main types of orders:
  - **limit order**: specify a price at which one is willing to buy (sell) a certain number of shares
  - **market order**: immediately buy (sell) a certain number of shares at the best available opposite quote(s)
  - **cancellation order**: cancel an existing limit order

**Price dynamics**

The price dynamics becomes a by-product of the order book dynamics

# Limit order book evolution

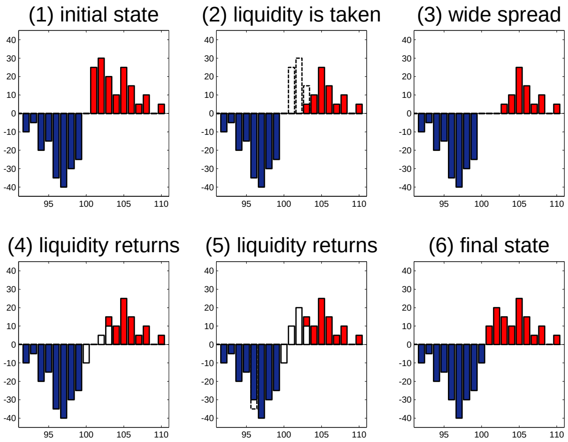



Figure: Dynamics of the order book

Basic questions asked since the first studies on limit order books:

- What will the next event be ?
- When will it happen ?
- Where will it take place ?
- What size will it have ?

# Some basic questions

Basic questions asked since the first studies on limit order books:

- What will the next event be ?
- When will it happen ?
- Where will it take place ?
- What size will it have ?

And, of course,

- What will be its impact on the price ?

# Plan I

1. Stylized facts of limit order books
   - A word on data
   - Arrival times of orders
   - Volume of orders
   - Placement of orders
   - Cancellation of orders
   - Intraday seasonality

2. Dependency properties of inter-arrival times
   - Empirical evidence of market making
   - The fine structure of inter-event durations: using lagged correlation matrices

# Plan I

1. Stylized facts of limit order books
   - A word on data
   - Arrival times of orders
   - Volume of orders
   - Placement of orders
   - Cancellation of orders
   - Intraday seasonality

2. Dependency properties of inter-arrival times
   - Empirical evidence of market making
   - The fine structure of inter-event durations: using lagged correlation matrices

| Timestamp | Side | Level | Price | Quantity |
|---|---|---|---|---|
| 33480.158 | B | 1 | 121.1 | 480 |
| 33480.476 | B | 2 | 121.05 | 1636 |
| 33481.517 | B | 5 | 120.9 | 1318 |
| *33483.218* | *B* | *1* | *121.1* | *420* |
| 33484.254 | B | 1 | 121.1 | 556 |
| 33486.832 | A | 1 | 121.15 | 187 |
| 33489.014 | B | 2 | 121.05 | 1397 |
| *33490.473* | *B* | *1* | *121.1* | *342* |
| *33490.473* | *B* | *1* | *121.1* | *304* |
| *33490.473* | *B* | *1* | *121.1* | *256* |
| 33490.473 | A | 1 | 121.15 | 237 |

Table: Tick by tick data file sample

# Trades

| Timestamp | Last | Last quantity |
|---|---|---|
| 33483.097 | 121.1 | 60 |
| 33490.380 | 121.1 | 214 |
| 33490.380 | 121.1 | 38 |
| 33490.380 | 121.1 | 48 |

Table: Trades data file sample.

# Trades and order book data processing

For each stock and each trading day:

1. Parse the tick by tick data file to compute order book state variations;
2. Parse the trades file and for each trade:
    1. Compare the trade price and volume to likely market orders whose timestamps are in $[t^{Tr} - \Delta t, t^{Tr} + \Delta t]$, where $t^{Tr}$ is the trade timestamp and $\Delta t$ is a predefined time window;
    2. Match the trade to the first likely market order with the same price and volume and label the corresponding event as a market order;
    3. Remaining negative variations are labeled as cancellations.

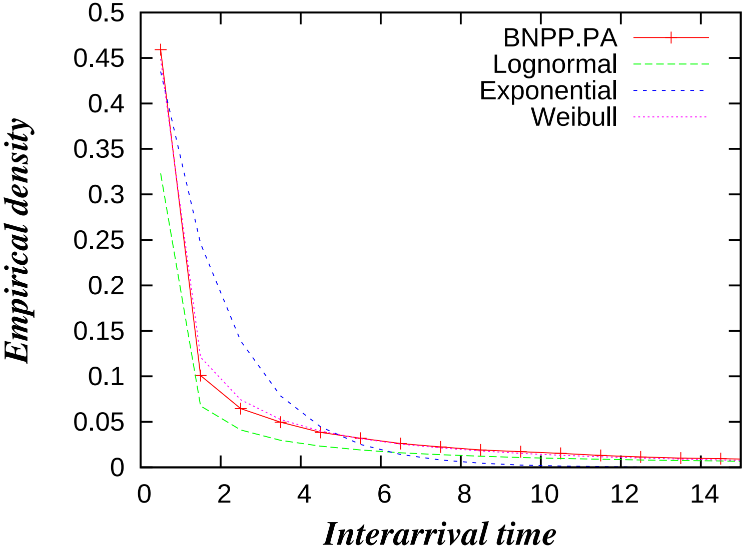
Doing so, we have an average matching rate of around 85% for CAC 40 stocks. As a byproduct, one gets the sign of each matched trade.

# Plan I

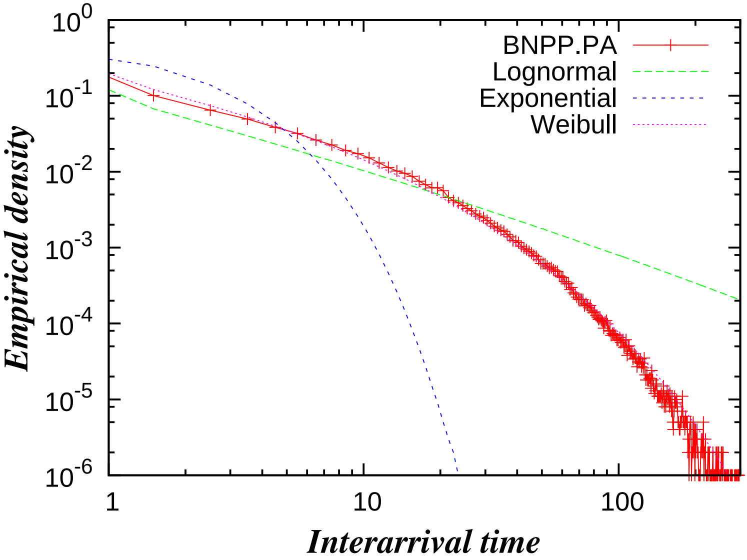
1. Stylized facts of limit order books
   - A word on data
   - Arrival times of orders
   - Volume of orders
   - Placement of orders
   - Cancellation of orders
   - Intraday seasonality

2. Dependency properties of inter-arrival times
   - Empirical evidence of market making
   - The fine structure of inter-event durations: using lagged correlation matrices

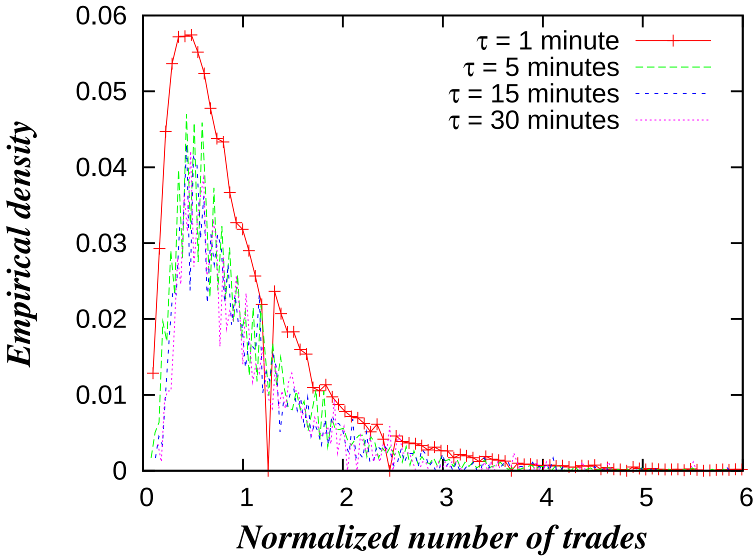
# Arrival time of orders



# Arrival time of orders



# Distribution of the number of trades



# Plan I

1. Stylized facts of limit order books
   - A word on data
   - Arrival times of orders
   - Volume of orders
   - Placement of orders
   - Cancellation of orders
   - Intraday seasonality

2. Dependency properties of inter-arrival times
   - Empirical evidence of market making
   - The fine structure of inter-event durations: using lagged correlation matrices

# Volume of incoming orders

The unconditional distribution of order sizes is very complex to characterize:

- Gopikrishnan et al. (2000) and Maslov and Mills (2001) observe a power law decay with an exponent $1 + \mu \approx 2.3 – 2.7$ for market orders and $1 + \mu \approx 2.0$ for limit orders;
- Challet and Stinchcombe (2001) emphasizes a clustering property.

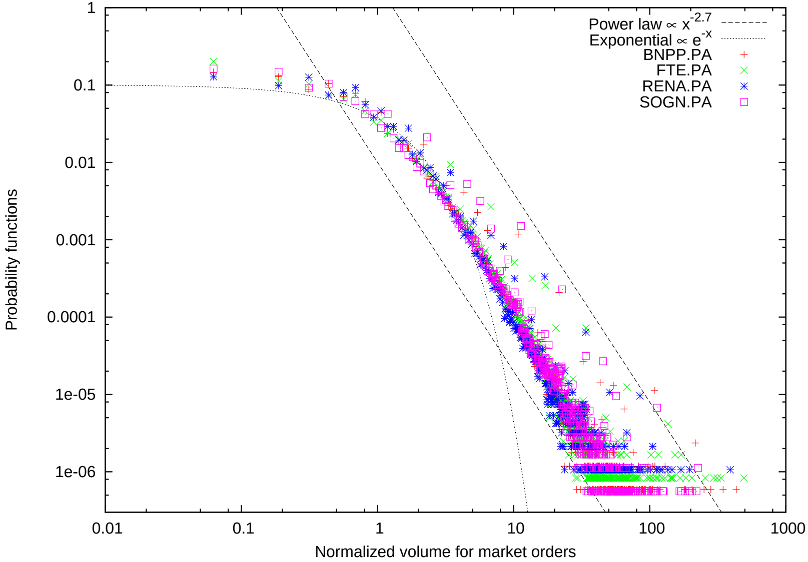



Figure: Distribution of volumes of market orders. Quantities are normalized by their mean.

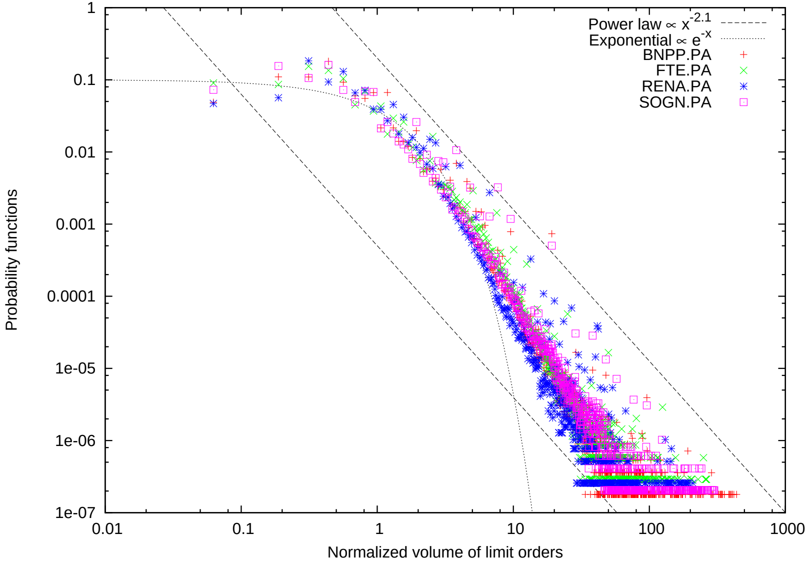



Figure: Distribution of normalized volumes of limit orders. Quantities are normalized by their mean.

# Plan I

1. Stylized facts of limit order books
   - A word on data
   - Arrival times of orders
   - Volume of orders
   - Placement of orders
   - Cancellation of orders
   - Intraday seasonality

2. Dependency properties of inter-arrival times
   - Empirical evidence of market making
   - The fine structure of inter-event durations: using lagged correlation matrices

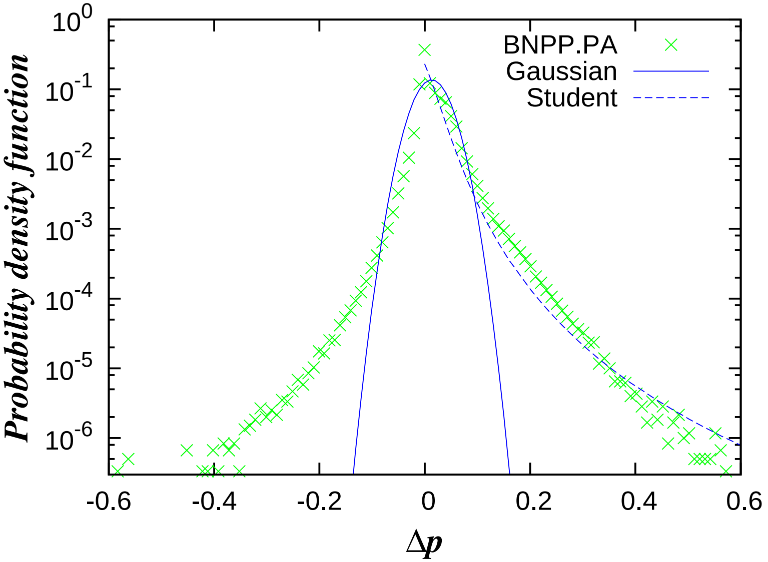



Figure: Placement of limit orders using the same best quote reference in semilog scale

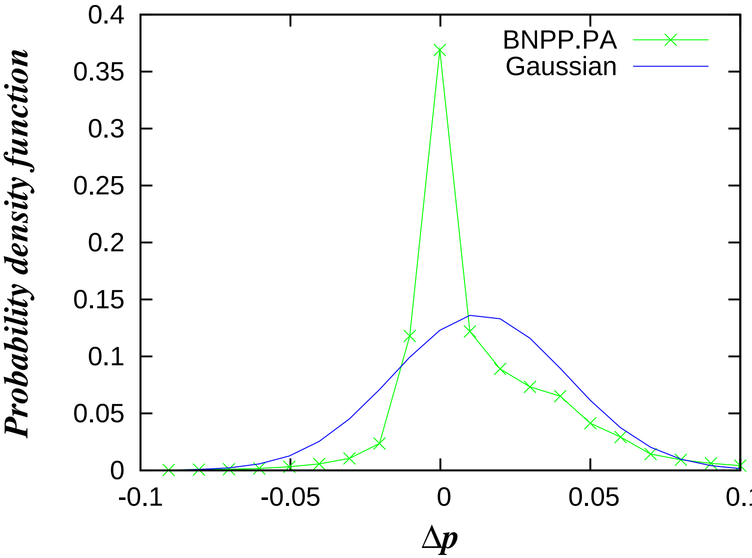



Figure: Placement of limit orders using the same best quote reference in linear scale

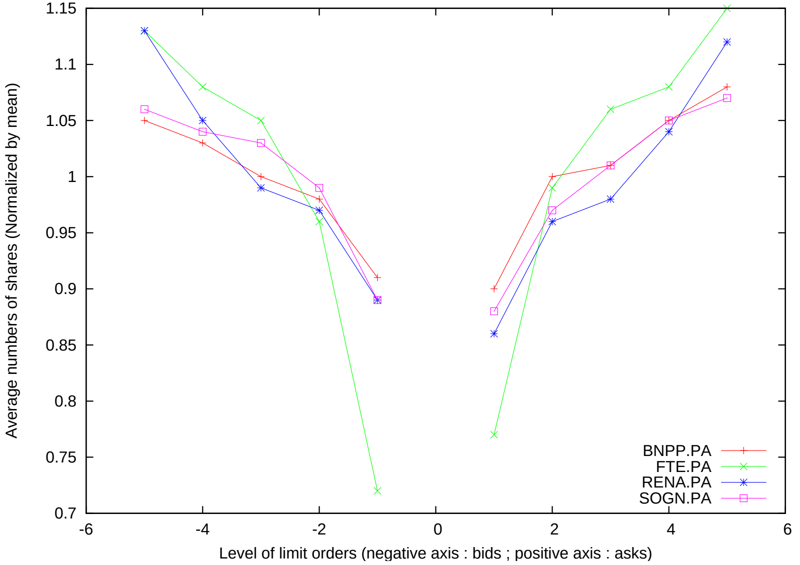



Figure: Average quantity offered in the limit order book.

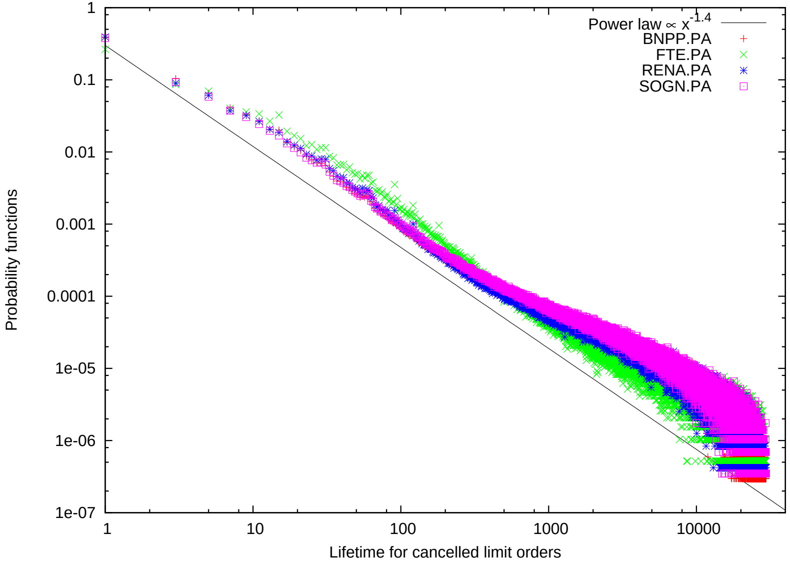



Figure: Distribution of estimated lifetime of cancelled limit orders.

# Plan I

1. Stylized facts of limit order books
   - A word on data
   - Arrival times of orders
   - Volume of orders
   - Placement of orders
   - Cancellation of orders
   - Intraday seasonality

2. Dependency properties of inter-arrival times
   - Empirical evidence of market making
   - The fine structure of inter-event durations: using lagged correlation matrices

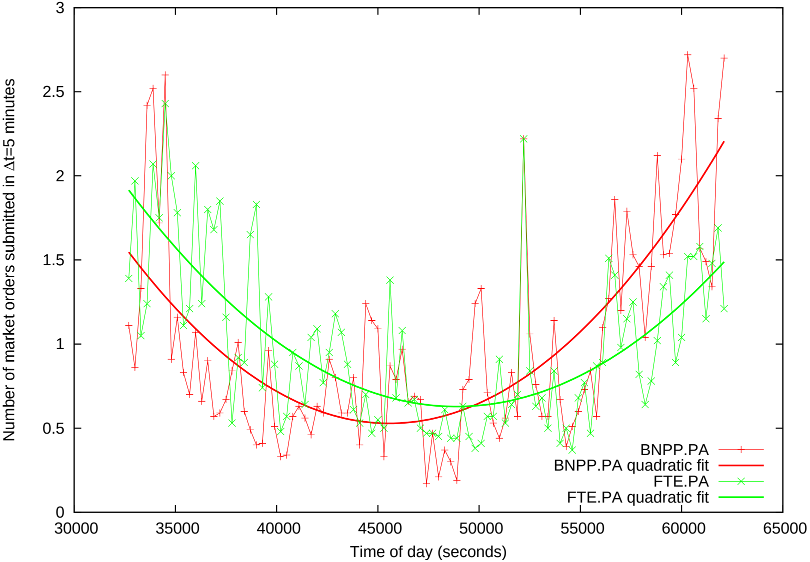



Figure: Normalized average number of market orders in a 5-minute interval.

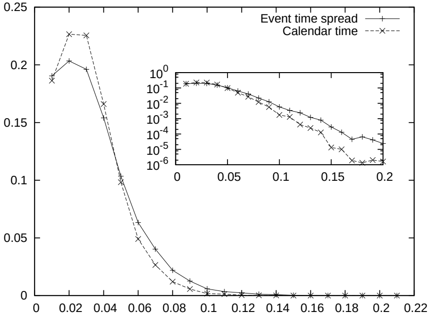
# Plan I

# Re-introducing physical time

Most early order book models use the clock known as *event time*. Under time-homogeneity and independance assumptions, such a time treatment is equivalent to the assumption that order flows are homogeneous Poisson processes.
However, it is clear that *physical* time has to be taken into account for the modelling of a realistic order book model : the Poisson hypothesis for the arrival times of orders of different kinds does not stand under careful scrutiny.

# Re-introducing physical time

The figure below plots examples of the empirical distribution function of the observed spread in event time (i.e. spread is measured each time an event happens in the order book), and in physical (calendar) time (i.e. measures are weighted by the time interval during which the order book is idle).




The frequencies of the most probable values of the time-weighted

# Plan I

1. Stylized facts of limit order books
   - A word on data
   - Arrival times of orders
   - Volume of orders
   - Placement of orders
   - Cancellation of orders
   - Intraday seasonality

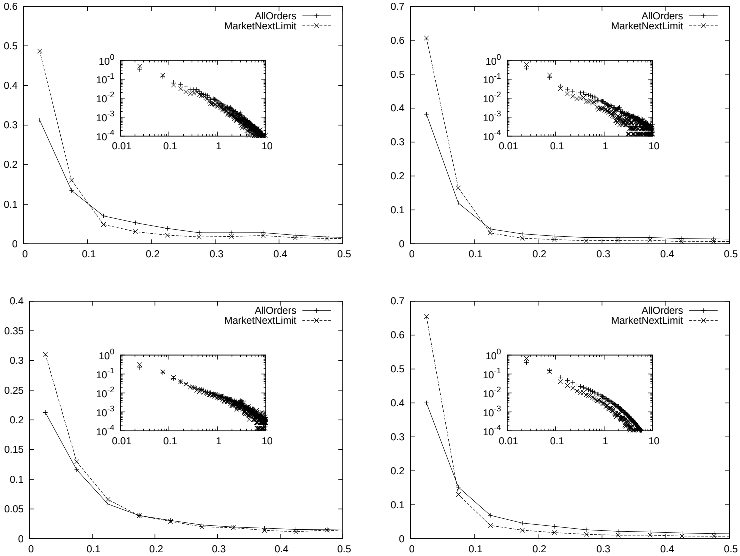
2. Dependency properties of inter-arrival times
   - Empirical evidence of market making
   - The fine structure of inter-event durations: using lagged correlation matrices

# Empirical evidence of market making

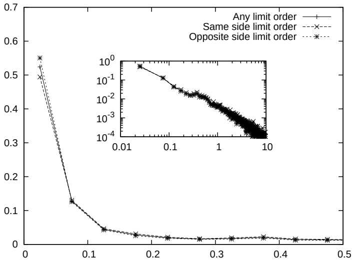
We compute for several assets:

- the empirical probability distribution of the time intervals of the counting process of all orders (limit orders and market orders mixed), i.e. the time step between any order book event (other than cancellation)
- and the empirical probability distribution of the time intervals between a market order and the immediately following limit order.

If an independent Poisson assumption held, then these empirical distributions should be identical.


AllOrders
MarketNextLimit
0.6
0.5
0.4
0.3
0.2
0.1
0
0
0.1
0.2
0.3
0.4
0.5
$10^{0}$
$10^{-1}$
$10^{-2}$
$10^{-3}$
$10^{-4}$
0.01
0.1
1
10
AllOrders
MarketNextLimit
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
0
0.1
0.2
0.3
0.4
0.5
$10^{0}$
$10^{-1}$
$10^{-2}$
$10^{-3}$
$10^{-4}$
0.01
0.1
1
10
AllOrders
MarketNextLimit
0.4
0.35
0.3
0.25
0.2
0.15
0.1
0.05
0
0
0.1
0.2
0.3
0.4
0.5
$10^{0}$
$10^{-1}$
$10^{-2}$
$10^{-3}$
$10^{-4}$
0.01
0.1
1
10
AllOrders
MarketNextLimit
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
0
0.1
0.2
0.3
0.4
0.5
$10^{0}$
$10^{-1}$
$10^{-2}$
$10^{-3}$
$10^{-4}$
0.01
0.1
1
10

No systematic link between the sign of the market order and the sign of the following limit order.



# Plan I

1. Stylized facts of limit order books
   - A word on data
   - Arrival times of orders
   - Volume of orders
   - Placement of orders
   - Cancellation of orders
   - Intraday seasonality

# A coarser description of the order flow

In Abergel et al. (2016), order book events are clustered according to a coarser-grain description in order to identify some dependency structures

| Notation | Definition |
|---|---|
| $M^0_{buy}$, $M^0_{sell}$ | buy/sell market order that does not change the mid price |
| $M^1_{buy}$, $M^1_{sell}$ | buy/sell market order that changes the mid price |
| $L^0_{buy}$, $L^0_{sell}$ | buy/sell limit order that does not change the mid price |
| $L^1_{buy}$, $L^1_{sell}$ | buy/sell limit order that changes the mid price |
| $C^0_{buy}$, $C^0_{sell}$ | buy/sell cancellation that does not change the mid price |
| $C^1_{buy}$, $C^1_{sell}$ | buy/sell cancellation that changes the mid price |

# Lagged correlation matrices of point processes

Some insight can be gained by studying the covariance matrix of inter-arrival times.
Given a duration $h$ and a lag $\tau$, the *lagged covariance matrix* $C_\tau^h = \left(C_\tau^h(i,j)\right)_{1\leq i,j\leq M}$ of the process is defined by:

$$C_\tau^h(i,j) = \frac{1}{h} Cov(N^i(t+h+\tau) - N^i(t+\tau), N^j(t+h) - N^j(t)). \quad (2.1)$$

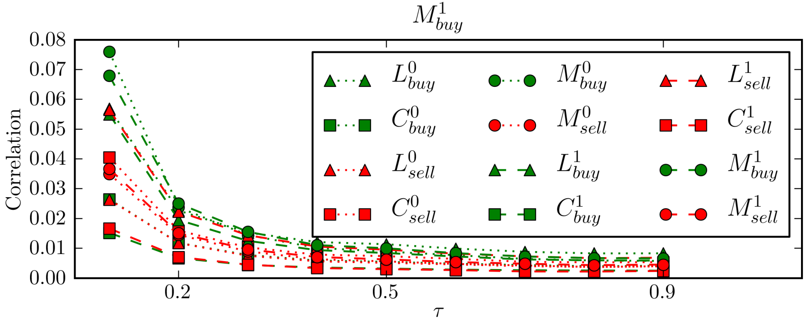
In order to avoid side effects caused by the wide variability of the frequencies across event type, it is actually more robust to rely on the lagged linear correlation matrix

$$\rho_\tau^h(i,j) = Correlation(N^i(t+h+\tau) - N^i(t+\tau), N^j(t+h) - N^j(t)). \quad (2.2)$$

In the case under scrutiny, the components of the process $N$ correspond to the 12 types of event

The time step $h$ is chosen as 0.1 second, and $\tau \in \{0.1, 0.2, \ldots 0.9\}$.

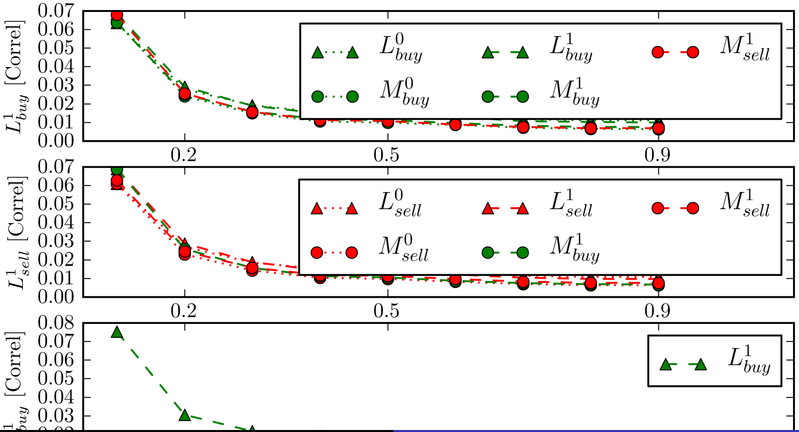
The figure below details the impact of the different types of orders on the arrival of $M^1_{buy}$ orders.




We see that the intensity of aggressive market orders $M^1_{buy}$ is primarily correlated with previous market orders on the same side

# Using lagged correlation matrices

We provide in the figure below the same results computed for the six types of aggressive events. In order to plot only the most relevant information, an arbitrary threshold of 6% is chosen: events for which the highest correlation coefficient is lower than 6% are discarded.



# Plan I

3. Order-driven market modelling
   - Early models in Econophysics
   - An empirical zero-intelligence model
   - Mathematical theory of zero-intelligence models
     - Stability and long-time dynamics
     - Large-scale limit of the price process

4. Advanced models
   - Modelling interactions between agents
   - Enhancing zero-intelligence model
     - Stability and long-time dynamics

3 Order-driven market modelling
- Early models in Econophysics
- An empirical zero-intelligence model
- Mathematical theory of zero-intelligence models
  - Stability and long-time dynamics
  - Large-scale limit of the price process

4 Advanced models
- Modelling interactions between agents
- Enhancing zero-intelligence model
  - Stability and long-time dynamics

# An elementary model: the Bak, Paczuski and Shubik model

A model inspired from reaction-diffusion, where half of the agents are asking for one share of stock with price

$$p_b^j(0) \in \{0, \bar{p}/2\}, \qquad j = 1, \ldots, N/2,$$

and the other half are selling one share of stock with price:

$$p_s^j(0) \in \{\bar{p}/2, \bar{p}\}, \qquad j = 1, \ldots, N/2.$$

At each time step $t$, agents revise their offer by exactly one tick, with equal probability to go up or down. Therefore, at time $t$, each seller (resp. buyer) agent chooses his new price as:

$$p_s^j(t+1) = p_s^j(t) \pm 1 \qquad \text{(resp. } p_b^j(t+1) = p_b^j(t) \pm 1 \text{ ).}$$

In Maslov (2000), at each time step, a trader is chosen to perform an action: this trader can submit a limit order with probability $q_l$, or a market order with probability $1 - q_l$.

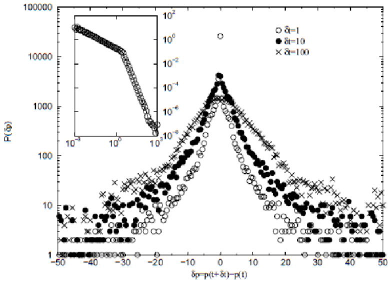



Figure: Empirical probability density functions of the price increments in the Maslov model. In inset, log-log plot of the positive increments

# Plan I

3. Order-driven market modelling
   - Early models in Econophysics
   - An empirical zero-intelligence model
   - Mathematical theory of zero-intelligence models
     - Stability and long-time dynamics
     - Large-scale limit of the price process

4. Advanced models
   - Modelling interactions between agents
   - Enhancing zero-intelligence model
     - Stability and long-time dynamics

# The zero-intelligence model of Mike and Farmer (2008)

Mike and Farmer (2008) is the first model that proposes an advanced calibration on market data of order placement and cancellation of orders.

At each time step, one trading order is simulated: an ask (resp. bid) trading order is randomly placed at $n(t) = a(t) + \delta a$ (resp. $n(t) = b(t) + \delta b$) according to a Student distribution calibrated on market data. If an ask (resp. bid) order satisfies $\delta a < -s(t) = b(t) - a(t)$ (resp. $\delta b > s(t) = a(t) - b(t)$), then it is a market order and a transaction occurs at price $a(t)$ (resp. $b(t)$.

# Introducing cancellations

Mike and Farmer (2008) proposes an empirical distribution for cancellation based on three components:

- the position in the order book, measured as the ratio $y(t) = \frac{\Delta(t)}{\Delta(0)}$ where $\Delta(t)$ is the distance of the order from the opposite best quote at time $t$,
- the order book imbalance, measured by the indicator $N_{imb}(t) = \frac{N_a(t)}{N_a(t)+N_b(t)}$ (resp. $N_{imb}(t) = \frac{N_b(t)}{N_a(t)+N_b(t)}$) for ask (resp. bid) orders, where $N_a(t)$ and $N_b(t)$ are the number of orders at ask and bid in the book at time $t$,
- the total number $N_t(t) = N_a(t) + N_b(t)$ of orders in the book.

The probability $P(C|y(t), N_{imb}(t), N_t(t))$ to cancel an ask order at time $t$ is

$$P(C|y(t), N_{imb}(t), N_t(t)) = A(1 - e^{-y(t)})(N_{imb}(t) + B)\frac{1}{N_t(t)}.$$

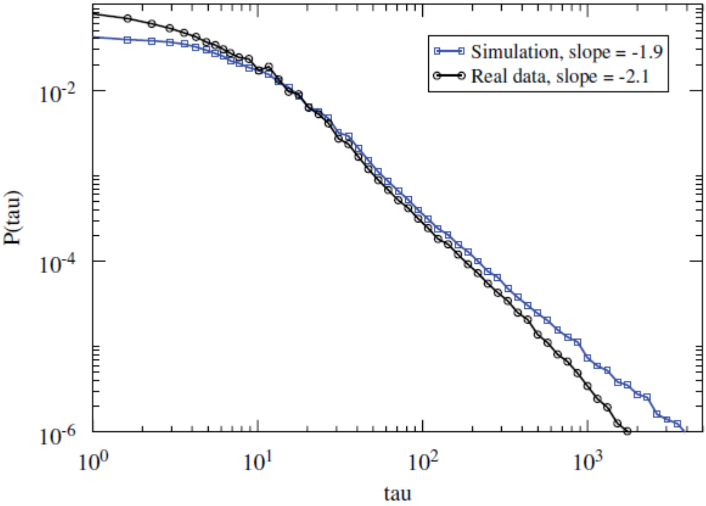
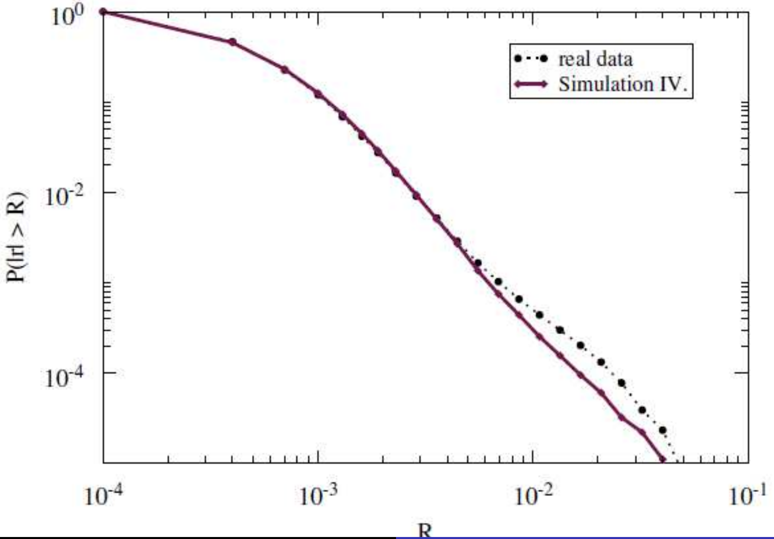



Figure: Lifetime of orders for simulated data in the Mike and Farmer model, compared to the empirical data used for fitting

# Distribution of returns

The distribution of returns exhibit fat tails in agreement with empirical data



# Plan I

3. Order-driven market modelling
   - Early models in Econophysics
   - An empirical zero-intelligence model
   - Mathematical theory of zero-intelligence models
     - Stability and long-time dynamics
     - Large-scale limit of the price process

4. Advanced models
   - Modelling interactions between agents
   - Enhancing zero-intelligence model
     - Stability and long-time dynamics

# Some general mathematical modelling assumptions

A set of reasonable assumptions:

- the limit order book is described as a point process;
- several types of events can happen;
- two events cannot occur simultaneously (simple point process).

The main questions of interest are

- the model-generated *shape*;
- the stationarity and ergodicity of the order book;
- the dynamics of the induced price...

# Some general mathematical modelling assumptions

A set of reasonable assumptions:

- the limit order book is described as a point process;
- several types of events can happen;
- two events cannot occur simultaneously (simple point process).

The main questions of interest are

- the model-generated *shape*;
- the stationarity and ergodicity of the order book;
- the dynamics of the induced price...
- ... particularly, its behaviour at larger time scales.

$\Rightarrow$ Linking microstructural properties and continuous-time finance

# Some notations

- The order book is represented by a finite-size vector of quantities

$$\mathbf{X}(t) := (\mathbf{a}(t); \mathbf{b}(t)) := (a_1(t), \dots, a_K(t); b_1(t), \dots, b_K(t));$$

- $\mathbf{a}(t)$: ask side of the order book
- $\mathbf{b}(t)$: bid side of the order book
- $\Delta P$: tick size
- $q$: unit volume
- $P = \frac{P^A + P^B}{2}$: mid-price
- $A(p), B(p)$: cumulative number of sell (buy) orders up to price level $p$

# Notations

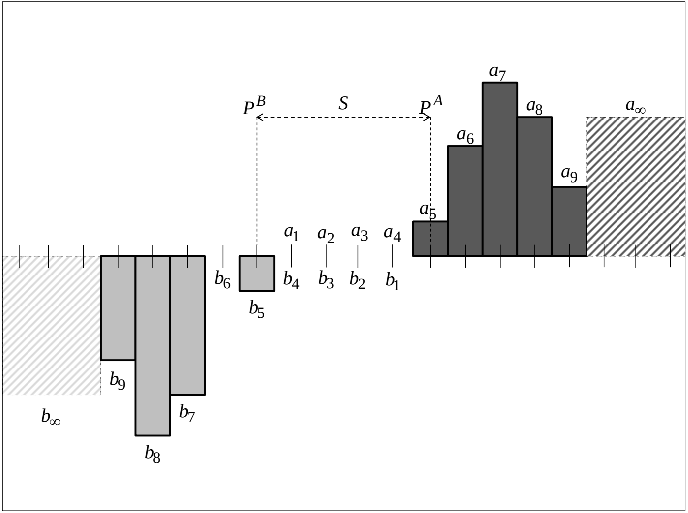



Figure: Order book notations

# Zero-intelligence model

Inspired by Smith et al. (2003), Abergel and Jedidi (2013) analyzes an elementary LOB model where the events affecting the order book are described by **independent Poisson processes**:

- $M^{\pm}$: arrival of new **market order**, with intensity $\dfrac{\lambda^{M^{\pm}}}{q}$;
- $L_i^{\pm}$: arrival of a **limit order** at level $i$, with intensity $\dfrac{\lambda_i^{L^{\pm}} \Delta P}{q}$;
- $C_i^{\pm}$: **cancellation** of a limit order at level $i$, with intensity $\lambda_i^{C^+} \dfrac{a_i}{q}$ and $\lambda_i^{C^-} \dfrac{|b_i|}{q}$
  $\Rightarrow$ Cancellation rate is **proportional** to the outstanding quantity at each level.

Under these assumptions, $(\mathbf{X}(t))_{t \geq 0}$ is a **Markov process** with state space $\mathcal{S} \subset \mathbb{Z}^{2K}$.

# Order book dynamics

$$\begin{cases} da_i(t) = -\left(q - \sum_{k=1}^{i-1} a_k\right)_+ dM^+(t) + q dL_i^+(t) - q dC_i^+(t) \\ \qquad + (J^{M^-}(\mathbf{a}) - \mathbf{a})_i dM^-(t) + \sum_{i=1}^{K} (J^{L_i^-}(\mathbf{a}) - \mathbf{a})_i dL_i^-(t) \\ \qquad + \sum_{i=1}^{K} (J^{C_i^-}(\mathbf{a}) - \mathbf{a})_i dC_i^-(t), \\ db_i(t) = \text{similar expression}, \end{cases}$$

where $J^{M^\pm}$, $J^{L_i^\pm}$, and $J^{C_i^\pm}$ are shift operators corresponding to the effect of order arrivals on the reference frame

# The shift operator

For instance, the shift operator corresponding to the arrival of a sell market order is

$$J^{M^-}(\mathbf{a}) = \left( \underbrace{0, 0, \ldots, 0}_{\text{k times}}, a_1, a_2, \ldots, a_{K-k} \right),$$

with

$$k := \inf\{p : \sum_{j=1}^{p} |b_j| > q\} - \inf\{p : |b_p| > 0\}.$$

# Infinitesimal generator

$$
\begin{aligned}
\mathcal{L}f(\mathbf{a};\mathbf{b}) = {} & \lambda^{M^+}\left(f\left([a_i-(q-A(i-1))_+]_+; J^{M^+}(\mathbf{b})\right)-f\right) \\
& + \sum_{i=1}^{K} \lambda_i^{L^+}\left(f\left(a_i+q; J^{L_i^+}(\mathbf{b})\right)-f\right) \\
& + \sum_{i=1}^{K} \lambda_i^{C^+} a_i\left(f\left(a_i-q; J^{C_i^+}(\mathbf{b})\right)-f\right) \\
& + \lambda^{M^-}\left(f\left(J^{M^-}(\mathbf{a}); [b_i+(q-B(i-1))_+]_-\right)-f\right) \\
& + \sum_{i=1}^{K} \lambda_i^{L^-}\left(f\left(J^{L_i^-}(\mathbf{a}); b_i-q\right)-f\right) \\
& + \sum_{i=1}^{K} \lambda_i^{C^-} |b_i|\left(f\left(J^{C_i^-}(\mathbf{a}); b_i+q\right)-f\right)
\end{aligned}
$$

# Stability of the order book

## Stationary order book distribution

Abergel and Jedidi (2013) If $\underline{\lambda_C} = \min_{1 \le i \le N}\{\lambda_i^{C^\pm}\} > 0$, then $(\mathbf{X}(t))_{t \ge 0} = (\mathbf{a}(t); \mathbf{b}(t))_{t \ge 0}$ is an **ergodic** Markov process. In particular $(\mathbf{X}(t))$ has a **stationary distribution** $\pi$. Moreover, the rate of convergence of the order book to its stationary state is **exponential**.

The proof relies on the use of a Lyapunov function
The proportional cancellation rate helps a lot... and so do the boundary conditions!

# Martingale approach

## A powerful tool

The ergodicity of the order book allows for a a direct approach using martingale convergence theorems *à la* Ethier and Kurtz (2005)and ergodic theorems :

- the evolution of the price is $dP_t = \sum_{i=1}^{K'} F_i(X_t) dN_t^i$;
- the rescaled, centered price is $\tilde{P}_t^n \equiv \frac{P_{nt} - \int_0^{nt} \sum_{i=1}^{K'} F_i(X_t)\lambda_i dt}{\sqrt{n}}$
- its predictable quadratic variation is $< \tilde{P}^n, \tilde{P}^n >_t = \frac{\int_0^{nt} \sum_{i=1}^{K'} (F_i(X_t))^2 \lambda_i dt}{n}$
- ergodicity ensures the convergence of $\frac{\int_0^{nt} \sum_{i=1}^{K'} (F_i(X_t))^2 \lambda_i dt}{nt}$ as $n \to \infty$

# Plan I

3. Order-driven market modelling
   - Early models in Econophysics
   - An empirical zero-intelligence model
   - Mathematical theory of zero-intelligence models
     - Stability and long-time dynamics
     - Large-scale limit of the price process

4. **Advanced models**
   - Modelling interactions between agents
   - Enhancing zero-intelligence model
     - Stability and long-time dynamics

# Plan I

3 Order-driven market modelling
- Early models in Econophysics
- An empirical zero-intelligence model
- Mathematical theory of zero-intelligence models
  - Stability and long-time dynamics
  - Large-scale limit of the price process

4 Advanced models
- Modelling interactions between agents
- Enhancing zero-intelligence model
  - Stability and long-time dynamics

# Herding behaviour

In Cont and Bouchaud (2000), $N$ agents trade a given stock with price $p(t)$. At each time step, agents choose to buy or sell one unit of stock, i.e. their demand is $\phi_i(t) = \pm 1, i = 1, \ldots, N$ with probability $a$ or are idle with probability $1 - 2a$. The price change is assumed to be linearly linked with the excess demand $D(t) = \sum_{i=1}^{N} \phi_i(t)$ with a factor $\lambda$ measuring the liquidity of the market :

$$p(t+1) = p(t) + \frac{1}{\lambda} \sum_{i=1}^{N} \phi_i(t).$$

$\lambda$ can also be interpreted as a market depth, i.e. the excess demand needed to move the price by one unit.

# Herding behaviour

At each time step, agents $i$ and $j$ can be linked with probability $p_{ij} = p = \frac{c}{N}$, $c$ being a parameter measuring the degree of clustering among agents.

Denoting by $n_c(t)$ the number of cluster at a given time step $t$, $W_k$ the size of the $k$-th cluster, $k = 1, \ldots, n_c(t)$ and $\phi_k = \pm 1$ its investement decision, the price variation is:

$$\Delta p(t) = \frac{1}{\lambda} \sum_{k=1}^{n_c(t)} W_k \phi_k .$$

This modelling is a direct application to the field of finance of the random graph framework as studied in Erdos and Renyi (1960). Kirman (1983) previously suggested it in economics.

# Fundamentalists and trend followers

Lux and Marchesi (2000) proposed a model in line with agent-based models in behavioural finance, but where trading rules are kept simple enough. The market has $N$ agents that can be part of two distinct groups of traders: $n_f$ traders are "fundamentalists", who share an exogenous idea $p_f$ of the value of the current price $p$; and $n_c$ traders are "chartists" (or trend followers), who make assumptions on the price evolution based on the observed trend (moving average).

# Fundamentals and trend followers

The total number of agents is constant, so that $n_f + n_c = N$ at any time. At each time step, the price can be moved up or down with a fixed jump size of $\pm 0.01$ (a tick). The probability to go up or down is directly linked to the excess demand $ED$ through a coefficient $\beta$.

Fundamentalists are expected to stabilize the market, while chartists should destabilize it.
In addition, non-trivial features are expected from herding behaviour and transitions between groups of traders: the $n_c$ chartists can change their view on the market, based on a clustering process modelled by an opinion index $x = \frac{n_+ - n_-}{n_c}$ representing the weight of the majority. The probabilities $\pi_+$ and $\pi_-$ to switch from one group to another are formally written :

$$\pi_\pm = v \frac{n_c}{N} e^{\pm U}, \qquad U = \alpha_1 x + \alpha_2 p / v.$$

Transitions between fundamentalists and chartists are also allowed, see Lux and Marchesi (2000) for details).

# Volatility clustering and threshold behaviour

Cont (2007) proposes a model where, at each period, an agent $i \in \{1, \ldots, N\}$ can issue a buy or a sell order: $\phi_i(t) = \pm 1$. Information is represented by a series of i.i.d Gaussian random variables. $(\epsilon_t)$. This public information $\epsilon_t$ is a forecast for the value $r_{t+1}$ of the return of the stock. Each agent $i \in \{1, \ldots, N\}$ decides whether to follow this information according to a threshold $\theta_i > 0$ representing its sensibility to the public information:

$$\phi_i(t) = \begin{cases} 1 & \text{if} \quad \epsilon_i(t) > \theta_i(t) \\ 0 & \text{if} \quad |\epsilon_i(t)| < \theta_i(t) \\ -1 & \text{if} \quad \epsilon_i(t) < -\theta_i(t) \end{cases}$$

Then, once every choice is made, the price evolves according to the excess demand $D(t) = \sum_{i=1}^{N} \phi_i(t)$, in a way similar to Cont and Bouchaud (2000).

# The limitations of zero-intelligence models

Zero-intelligence models fail to capture the dependencies between various types of orders:

- clustering of market orders;
- interplay between liquidity taking and providing;
- leverage effect.

# Enhancing zero-intelligence models

Several avenues are of interest :
For instance, Huang et al. (2015) propose a **Markovian** order book model in which the intensities of arrival of orders are state-dependent.
The order book is represented as a collection of queues indexed by their distance in ticks to a reference price.
Also, the use of **Hawkes** processes for modelling limit order books is a very active and fruitful direction of research

Hawkes processes provide an *ad hoc* tool to describe the mutual excitations of the arrivals of different types of orders.
In $D$ dimensions, the process $N_s^j$ has a stochastic intensity $\lambda_t^j$ such that

$$\lambda_t^j = \lambda_0^j + \sum_{p=1}^{D} \int_0^t \phi_{jp}(t-s)dN_s^p. \tag{4.1}$$

A typical choice is the exponential kernel

$$\phi_{jp}(s) = \alpha_{jp} \exp(-\beta_{jp} s) \tag{4.2}$$

leading to markovian processes. A classical result states that the process is stationary *iff* the spectral radius of the matrix

$$[\frac{\alpha_{jp}}{\beta_{jp}}] \tag{4.3}$$

is $< 1$, see Brémaud and Massoulié (1996).

# Large-time behaviour

Results similar to those obtained in the zero-intelligence case hold

**Large-time behaviour for Hawkes processes**

- Under the usual stationarity conditions for the intensities, there exists a Lyapunov function $V = \sum |a_i| + \sum |b_i| + \sum U_k \lambda_k$ and the LOB converges exponentially to its stationary distribution
- The rescaled, centered price converges to a Wiener process

The proofs are essentially the same as in the Poisson arrival case, see Abergel and Jedidi (2015), thanks to the fact that the proportional cancellation rate remains bounded away from zero

The question of modelling the interactions between agents of different types is quite fascinating. It has important consequences on many aspects of the understanding of limit order books, be it from an empirical, theoretical or practical point of view.
It is clear that much more work is still to be done, in view in particular of the fierce *competition* between agents following different strategies.
Such a game-theoretic approach to limit order book modelling is still in its infancy.

5 Simulation of limit order books
- Zero-intelligence models
- Simulation with Hawkes processes

# Plan I

5 Simulation of limit order books

- Zero-intelligence models
- Simulation with Hawkes processes

# The base algorithm

First specify the

- Model parameters: arrival rates $\lambda^M, \{\lambda_i^L\}_{i\in\{1,\dots K\}}, \{\lambda_i^C\}_{i\in\{1,\dots K\}}$, order book size $K$, reservoirs $a_\infty$, $b_\infty$, volume distribution parameters $(v^M, s^M), (v^L, s^L), (v^C, s^C)$;
- Simulation parameters (number of time steps) $N$;
- Initialization $t \leftarrow 0$, $\mathbf{X}(0) \leftarrow \mathbf{X}_{\text{init}}$.

For time step $n = 1, \dots, N$,

- Compute the best bid $p^B$ and best ask $p^A$.
- Compute $\Lambda^C(\mathbf{b}) = \sum_{i=1}^{K} \lambda_i^C |b_i|$, i.e. the weighted sum of shares at price levels from $p^A - K$ to $p^A - 1$.
- Compute $\Lambda^C(\mathbf{a}) = \sum_{i=1}^{K} \lambda_i^C a_i$.

# The base algorithm

- Draw the waiting time $\tau$ for the next event from an exponential distribution with parameter

$$\Lambda(\mathbf{a}, \mathbf{b}) = 2(\lambda^M + \Lambda^L) + \Lambda^C(\mathbf{a}) + \Lambda^C(\mathbf{b}).$$

- Draw a new event according to the probability vector

$$\left(\lambda^M, \lambda^M, \Lambda^L, \Lambda^L, \Lambda^C(\mathbf{a}), \Lambda^C(\mathbf{b})\right) / \Lambda(\mathbf{a}, \mathbf{b}).$$

- Depending on the event type, draw the order volume from a lognormal distribution with parameters $(v^M, s^M), (v^L, s^L)$ or $(v^C, s^C)$.

- If the selected event is a limit order, select the relative price level from $\{1, 2, \ldots, K\}$ according to the probability vector

$$\left(\lambda_1^L, \ldots, \lambda_K^L\right) / \Lambda^L.$$

- If the selected event is a cancellation, select the relative price level at which to cancel an order from $\{1, 2, \ldots, K\}$ according to the probability vector

$$\left(\lambda_1^C a_1, \ldots, \lambda_K^C a_K\right) / \Lambda^C(\mathbf{a}).$$

  (or $\lambda^C(\mathbf{b})/\Lambda^C(\mathbf{b})$ for the bid side.)
- Update the order book state
- Enforce the boundary conditions:

$$\begin{aligned} a_i &= a_\infty, i \geq K+1, \\ b_i &= b_\infty, i \geq K+1. \end{aligned}$$

- Increment the event time $n$ by 1 and the physical time $t$ by $\tau$.

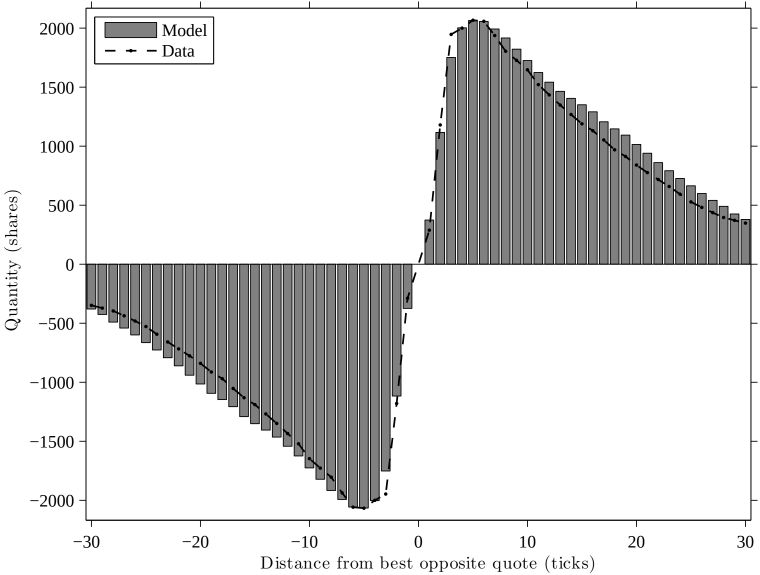



Figure: Average depth profile

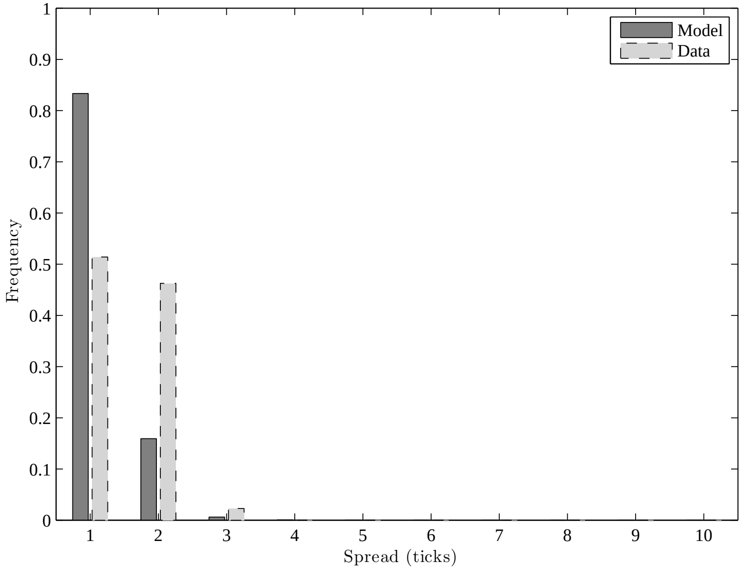



Figure: Probability distribution of the spread

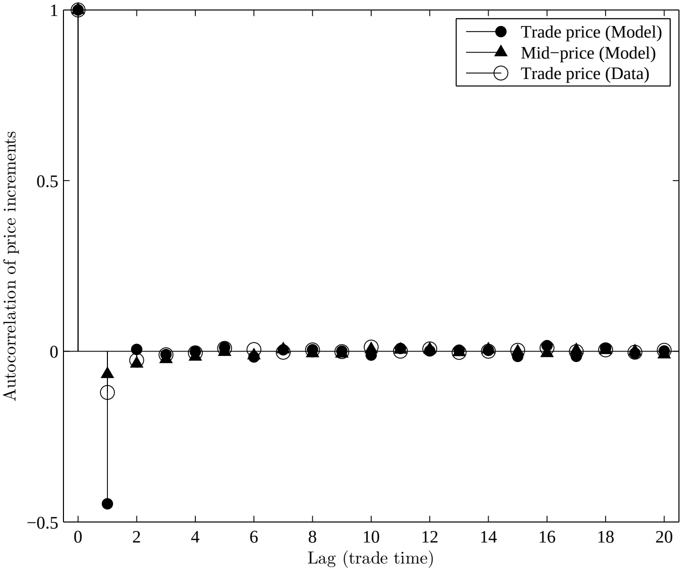



Figure: Autocorrelation of price increments

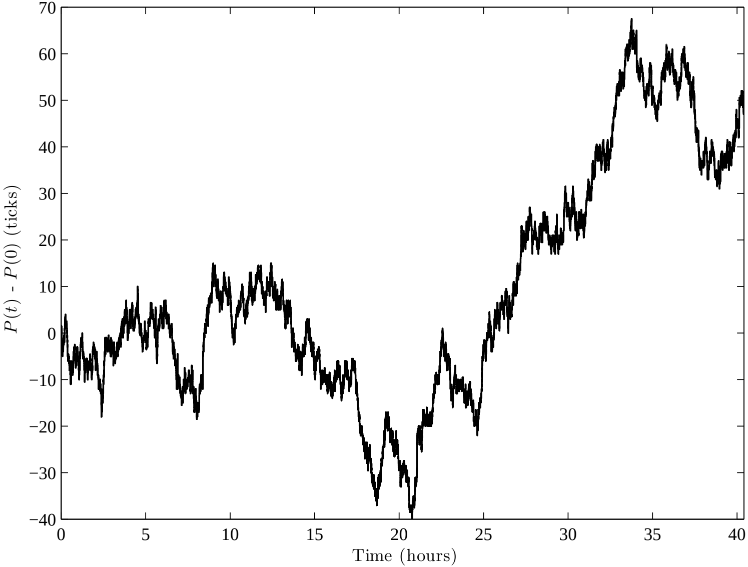



Figure: Price sample path

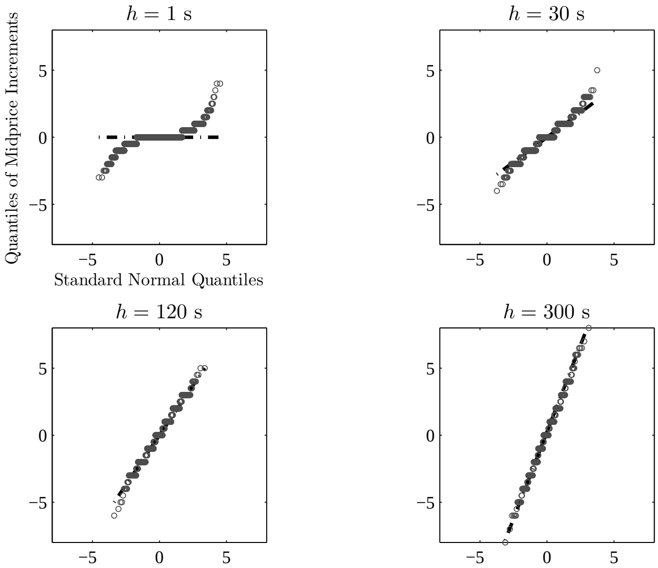



Figure: Q-Q plot of mid-price increments

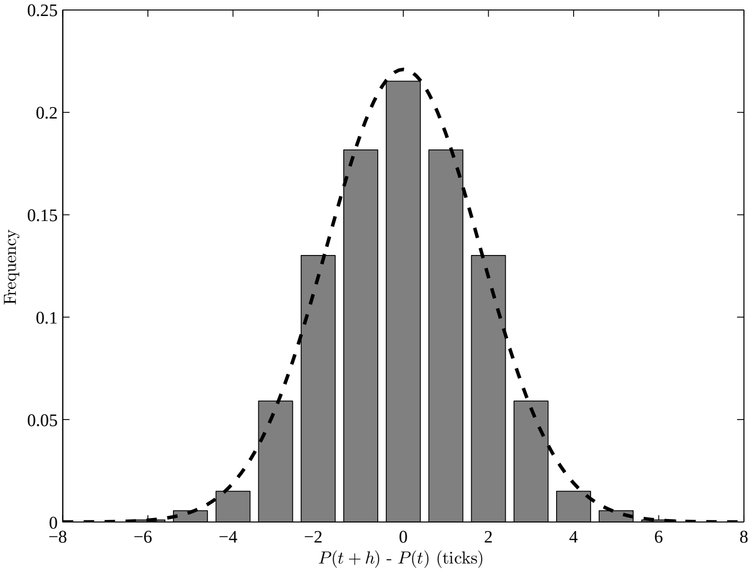
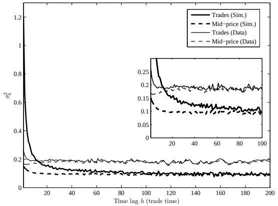
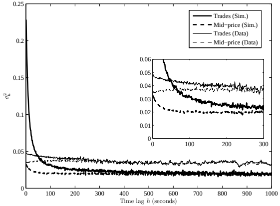



Figure: Probability distribution of price increments




(a) Trade time.




(b) Calendar time.

Figure: Signature plot: $\sigma_h^2 := \mathbb{V}\left[P(t+h) - P(t)\right]/h$. $y$ axis unit is tick$^2$ per trade for panel $(a)$ and tick$^2$.second$^{-1}$ for panel $(b)$

# Diffusive behaviour

Such a simple model already captures some empirically observed properties of the price dynamics

Long time diffusive behaviour

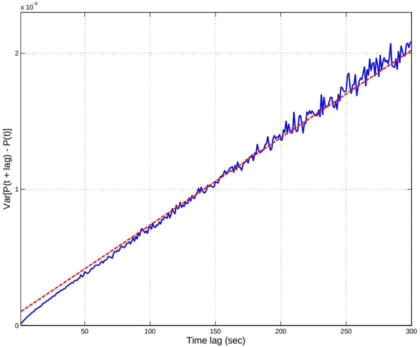



Figure: Diffusive behaviour of the price at low frequencies

# Parameters estimation

$T$ is the trading period of interest each day ($T = 4.5$ hours—[9 : 30–14 : 00]—in our case). Then

$$\widehat{\lambda^M} := \frac{\#\text{trades}}{2T},$$

and

$$\widehat{\lambda_i^L} := \frac{1}{2T}.$$

$$\left(\#\text{buy limit orders arriving } i \text{ tick away from the best opposite quote} \right.$$
$$\left. + \#\text{sell lim. orders etc.}\right).$$

For cancellations, we need to normalize the count by the average number of shares $\langle \mathbf{X}_i \rangle$ at distance $i$ from the best opposite quote:

$$\widehat{\lambda_i^C} := \frac{1}{\langle \mathbf{X}_i \rangle} \frac{1}{2T}.$$

$$\left(\#\text{cancellation orders in the bid side arriving } i \text{ tick away from the best opposite} \right.$$
$$\left. + \ \#\text{cancellation orders in the ask side etc.}\right),$$

We then average $\widehat{\lambda^M}$, $\widehat{\lambda_i^L}$ and $\widehat{\lambda_i^L}$ across 23 trading days to get the final estimates. As for the volumes, we estimate by maximum likelihood the parameters $(\widehat{v}, \widehat{s})$ of a lognormal distribution separately for each order type.

5. Simulation of limit order books
    - Zero-intelligence models
    - Simulation with Hawkes processes

# A Hawkes-process-based "toy" model

The flow of limit and market orders are now modelled by Hawkes processes $N^L$ and $N^M$, see Muni Toke (2011), with stochastic intensities $\lambda$ and $\mu$ defined as follows:

$$\begin{cases} \mu^M(t) & = & \mu_0^M & + \int_0^t \alpha_{MM} e^{-\beta_{MM}(t-s)} dN_s^M & & , \\ \lambda^L(t) & = & \lambda_0^L & + \int_0^t \alpha_{LM} e^{-\beta_{LM}(t-s)} dN_s^M & + \int_0^t \alpha_{LL} e^{-\beta_{LL}(t-s)} dN_s^L & \end{cases}$$

# Computation of the log-likelihood function

- The **log-likelihood** of a simple point process N with intensity $\lambda$ is:

$$\ln \mathcal{L}((N_t)_{t\in[0,T]}) = \int_0^T (1-\lambda(s))ds + \int_0^T \ln \lambda(s) dN(s),$$

which for a Hawkes process can be explicitly written as:

$$\ln \mathcal{L}(\{t_i\}_{i=1\dots n}) = t_n - \Lambda(0,t_n) + \sum_{i=1}^{n} \ln \left[ \lambda_0(t_i) + \sum_{j=1}^{P} \sum_{k=1}^{i-1} \alpha_j e^{-\beta_j (t_i - t_k)} \right].$$

$\Rightarrow$ Can be computed recursively, see **?**Muni Toke (2011).

Fitting and simulation

| Model | $\mu_0$ | $\alpha_{MM}$ | $\beta_{MM}$ | $\lambda_0$ | $\alpha_{LM}$ | $\beta_{LM}$ | $\alpha_{LL}$ | $\beta_{LL}$ |
|---|---|---|---|---|---|---|---|---|
| HP | 0.22 | - | - | 1.69 | - | - | - | - |
| LM | 0.22 | - | - | 0.79 | 5.8 | 1.8 | - | - |
| MM | 0.09 | 1.7 | 6.0 | 1.69 | - | - | - | - |
| MM LL | 0.09 | 1.7 | 6.0 | 0.60 | - | - | 1.7 | 6.0 |
| MM LM | 0.12 | 1.7 | 6.0 | 0.82 | 5.8 | 1.8 | - | - |
| MM LL LM | 0.12 | 1.7 | 5.8 | 0.02 | 5.8 | 1.8 | 1.7 | 6.0 |

Common parameters: $m_1^P = 2.7, \nu_1^P = 2.0, s_1^P = 0.9$
${}_1^V = 275, m_2^V = 380$
$\lambda^C = 1.35, \delta = 0.015$

Table: Estimated values of parameters used for simulations.

# Impact on arrival times

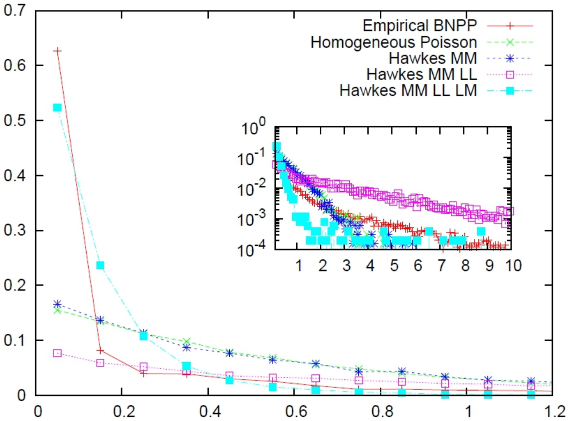



Figure: Empirical density function of the distribution of the interval times between a market order and the following limit order for three simulations, namely HP, MM+LL, MM+LL+LM, compared to empirical measures. In inset, same data using a semi-log scale

# References I


Abergel, F., Anane, M., Chakraborti, A., Jedidi, A., and Muni Toke, I. (2016). *Limit order books*. Cambridge University Press.

Abergel, F. and Jedidi, A. (2013). A mathematical approach to order book modelling. *International Journal of Theoretical and Applied Finance*, 16:1350025.

Abergel, F. and Jedidi, A. (2015). Long-time behavior of a Hawkes process-based limit order book. *SIAM Journal on Financial Mathematics*, 6:1026–1043.

Brémaud, P. and Massoulié, L. (1996). Stability of nonlinear Hawkes processes. *Ann. Probab.*, 24:1563–1588.

Challet, D. and Stinchcombe, R. (2001). Analyzing and modeling 1+ 1d markets. *Physica A: Statistical Mechanics and its Applications*, 300:285–299.

# References II

Cont, R. (2007). Volatility clustering in financial markets: Empirical facts and Agent-Based models. In Teyssiere, G. and Kirman, A. P., editors, *Long Memory in Economics*, pages 289–309. Springer, Heidelberg.

Cont, R. and Bouchaud, J.-P. (2000). Herd behavior and aggregate fluctuations in financial markets. *Macroeconomic Dynamics*, 4:170–196.

Erdos, P. and Renyi, A. (1960). On the evolution of random graphs. *Publications of the Mathematical Institute of the Hungarian Academy of Sciences*, 5:17–61.

Ethier, S. N. and Kurtz, T. G. (2005). *Markov Processes: Characterization and Convergence*. Wiley, Hoboken.

Gopikrishnan, P., Plerou, V., Gabaix, X., and Stanley, H. E. (2000). Statistical properties of share volume traded in financial markets. *Physical Review E*, 62:4493–4496.

Huang, W., Lehalle, C.-A., and Rosenbaum, M. (2015). Simulating and analyzing order book data: The queue-reactive model. *Journal of the American Statistical Association*, 110:107–122.

Kirman, A. P. (1983). On mistaken beliefs and resultant equilibria. In Frydman, R. and Phelps, E., editors, *Individual forecasting and aggregate outcomes: Rational expectations examined*, pages 147–168. Cambridge University Press, Cambridge.

Lux, T. and Marchesi, M. (2000). Volatility clustering in financial markets. *International Journal of Theoretical and Applied Finance*, 3:675–702.

Maslov, S. (2000). Simple model of a limit order-driven market. *Physica A: Statistical Mechanics and its Applications*, 278:571–578.

Maslov, S. and Mills, M. (2001). Price fluctuations from the order book perspective – empirical facts and a simple model. *Physica A: Statistical Mechanics and its Applications*, 299:234–246.

Mike, S. and Farmer, J. D. (2008). An empirical behavioral model of liquidity and volatility. *Journal of Economic Dynamics and Control*, 32:200–234.

Muni Toke, I. (2011). Market making in an order book model and its impact on the spread. In Abergel, F., Chakrabarti, B. K., Chakraborti, A., and Mitra, M., editors, *Econophysics of Order-driven Markets*, pages 49–64. Springer, Milan.

Smith, E., Farmer, J. D., Gillemot, L., and Krishnamurthy, S. (2003). Statistical theory of the continuous double auction. *Quantitative Finance*, 3:481–514.