

High-frequency data and limit order books

Lectures 1 and 2 - February 25th and March 4th, 2025

Ioane Muni Toke

Université Paris-Saclay
CentraleSupélec, MICS Laboratory, FiQuant Team



CentraleSupélec

U. Paris-Saclay CentraleSupélec cursus Ingénieur 3A Mathématiques et Data Science
IP Paris ENSAE M2 Statistics, Finance and Actuarial Science
U. Paris-Saclay Evry M2 Quantitative Finance

About the course

- ▶ **Expected outline:**
 1. February 25th (today): High-frequency data, LOB data
 2. March 4th: High-frequency data, LOB data
 3. March 11th: Point processes
 4. March 18th: Hawkes processes
 5. March 25th: Hawkes processes in finance
 6. April 1st: LOB models
 7. April 8th: Market impact and Optimal Execution
 8. April 15th: Final exam
- ▶ **Labs:** Always bring a laptop to class ! Needed: working Python configuration with scientific packages and jupyter notebooks. Details in Section Labs.
- ▶ **Final grade:** labs (50%) + final exam (50%).

Table of contents

Timescales in finance and high-frequency data

Limit order books

A bit of history

LOB basic mechanisms

Models vs real markets

High-frequency financial datasets

Different types of datasets

Quality of trades and quotes datasets

Some observations on prices

Prices

Spread and tick size

Some observations on trades

Durations, volumes, trading activity

Trade signs

Imbalance and trading activity

Some observations on LOB

Shape, activity, order lifetimes

Conditional probabilities of the order flow

Labs and Lab 1

General information

Lab 1

Table of contents

Timescales in finance and high-frequency data

Limit order books

A bit of history

LOB basic mechanisms

Models vs real markets

High-frequency financial datasets

Different types of datasets

Quality of trades and quotes datasets

Some observations on prices

Prices

Spread and tick size

Some observations on trades

Durations, volumes, trading activity

Trade signs

Imbalance and trading activity

Some observations on LOB

Shape, activity, order lifetimes

Conditional probabilities of the order flow

Labs and Lab 1

General information

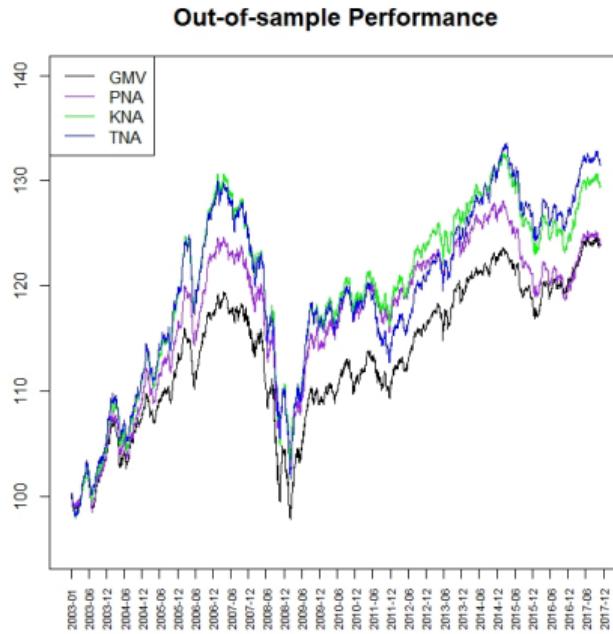
Lab 1

Standard stochastic modeling in finance

- ▶ Price is the fundamental variable described by some stochastic process:
 - ▶ Time series framework (econometric studies, portfolio optimization, etc.)
 - ▶ Continuous-time stochastic modeling for derivatives (Black & Scholes, stochastic volatility models, models with jumps, etc.)
- ▶ Price is unique
- ▶ No focus on price formation
- ▶ No focus on the structure of the exchange (infinite liquidity, no latency, etc.)
- ▶ No focus on strategies and agents' behaviors (horizons, costs/fees, etc.)

Typical timescale of investment strategies

Example graphs from *random* arxiv papers found with keyword search "trading strategy performance"



Source: arxiv:1811.07522 and arxiv:1810.09825

Typical timescale of equity derivatives products

Unusual Options Activity

Thu, Jan 3rd, 2019 [?](#) [Help](#)

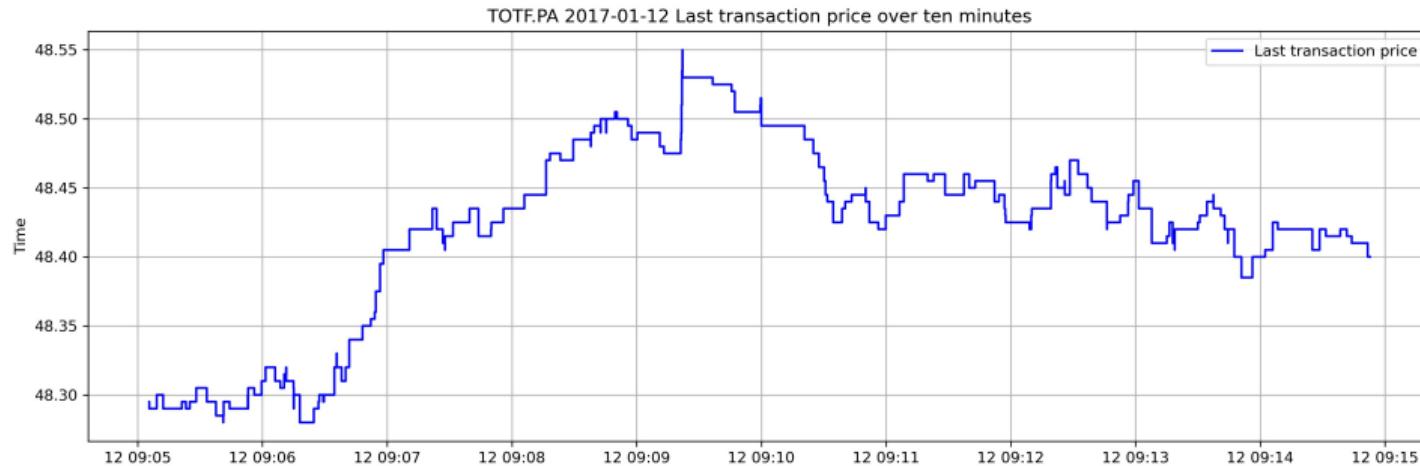
Provides insight on what "smart money" is doing with large volume orders. Option volume that exceeds open interest signals new positioning.

STOCKS	ETFs	INDICES												screen	tutorial	flipcharts	download											
Exclude Expiring After: MM/DD/YYYY <input type="button" value="Calendar"/>																												
Latest price quotes as of Thu, Jan 3rd, 2019.																												
Symbol	Price	Type	Strike	Exp Date	DTE	Bid	Midpoint	Ask	Last	Volume	Open Int	Vol/OI ▾	IV	Last Trade	Links													
+ AR	9.76	Call	15.00	01/17/20	380	0.60	0.70	0.80	0.75	38,618	196	197.03	50.25%	01/02/19														
+ XRX	20.13	Put	15.00	01/17/20	380	0.90	1.05	1.20	1.02	10,007	193	51.85	43.69%	01/02/19														
+ THC	17.77	Call	20.00	01/17/20	380	4.00	4.25	4.50	4.27	19,755	404	48.90	67.96%	01/02/19														
+ AMRN	13.01	Call	13.00	03/15/19	72	2.08	2.12	2.16	2.05	10,490	281	37.33	87.87%	01/02/19														
+ I	22.75	Put	22.00	02/15/19	44	2.60	2.78	2.95	2.66	4,211	113	37.27	99.32%	01/02/19														
+ CPB	32.26	Put	28.00	02/15/19	44	0.40	0.45	0.50	0.50	3,000	101	29.70	47.02%	01/02/19														
+ WTW	39.89	Put	55.00	04/18/19	106	15.90	16.15	16.40	16.40	6,006	253	23.74	66.80%	01/02/19														
+ GE	8.05	Call	9.50	01/18/19	16	0.03	0.04	0.05	0.04	2,385	101	23.61	56.93%	01/02/19														
+ LLY	114.88	Call	115.00	03/15/19	72	4.80	5.00	5.20	4.90	2,378	109	21.82	22.73%	01/02/19														
+ NSC	149.54	Put	125.00	02/15/19	44	1.40	1.53	1.65	1.67	3,650	169	21.60	49.00%	01/02/19														
+ MRVL	16.50	Call	16.50	01/11/19	9	0.51	0.52	0.53	0.51	11,550	643	17.96	48.77%	01/02/19														

Source : barchart.com on January 3rd, 2019

Intraday timescale

- Price of the last observed transaction, TOTF.PA, January 12th 2017, 9:05am to 9:15 am.



The price function $p : t \mapsto p(t)$ is a step function with values in a discrete set.

New data, new paradigm

- ▶ Orders of magnitude for a very liquid CAC40 stock (TOTF.PA, random day in 2017):
 - ▶ roughly 1.5×10^4 transactions per day, i.e. 60 years of close-to-close data
 - ▶ roughly 7×10^5 orders per day, i.e. 28 centuries of close-to-close data
- ▶ Price formation is observable through the order flow
- ▶ Key elements (e.g., volatilities, correlations) may be approached at new time scales
- ▶ Agent strategies may even be partially observable

The study of the microstructure of financial markets (trading mechanisms and price formation) and high-frequency observations at the order scale should provide a better understanding of financial markets.

Table of contents

Timescales in finance and high-frequency data

Limit order books

A bit of history

LOB basic mechanisms

Models vs real markets

High-frequency financial datasets

Different types of datasets

Quality of trades and quotes datasets

Some observations on prices

Prices

Spread and tick size

Some observations on trades

Durations, volumes, trading activity

Trade signs

Imbalance and trading activity

Some observations on LOB

Shape, activity, order lifetimes

Conditional probabilities of the order flow

Labs and Lab 1

General information

Lab 1

Table of contents

Timescales in finance and high-frequency data

Limit order books

A bit of history

LOB basic mechanisms

Models vs real markets

High-frequency financial datasets

Different types of datasets

Quality of trades and quotes datasets

Some observations on prices

Prices

Spread and tick size

Some observations on trades

Durations, volumes, trading activity

Trade signs

Imbalance and trading activity

Some observations on LOB

Shape, activity, order lifetimes

Conditional probabilities of the order flow

Labs and Lab 1

General information

Lab 1

Historical floor markets

- ▶ Consolidation of buyers and sellers
- ▶ Large development in the 19th century
- ▶ Cooperative/association of brokers/members ; members can trade for clients or for themselves ; centralized negotiations (book/board) or bilateral deals
- ▶ Open outcry : shouting, hand signals
- ▶ Progressively complemented/replaced by electronic markets: Toronto (1977), London (1986), Paris (CAC, 1987), NYSE (1995), ...



Source : Wikipedia ("open outcry") / Fortune.com

Orders, cancellations and transactions around 1900

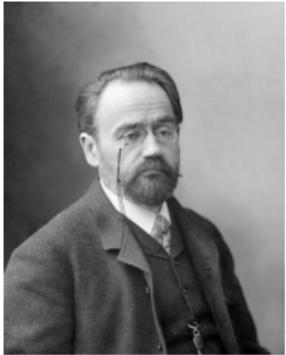
Emile Zola, L'Argent, 1890-1891.



- ▶ "Puis, [...], il vint serrer la main de Mazaud, dont il était le client. Il se pencha, donna un ordre, que celui-ci inscrivit sur une fiche. – Il vend ses Suez, murmura Moser."
- ▶ "sa façon d'écouter impressionnait tellement Moser, que souvent celui-ci, après lui avoir fait une confidence, courait changer un ordre, démonté par son silence."
- ▶ "Saccard entra dans le bureau de la liquidation. C'était là que le liquidateur, [...], dépouillait le carnet que lui remettait l'agent, chaque jour, après la Bourse, puis appliquait aux clients les affaires faites selon les ordres reçus, en s'aidant des fiches, conservées pour savoir les noms ; car le carnet ne porte pas les noms, ne contient que l'indication brève de l'achat ou de la vente : telle valeur, telle quantité, tel cours, de tel agent."

Orders, cancellations and transactions around 1900

Emile Zola, L'Argent, 1890-1891 (Money, translation 1900).



- ▶ "after which, [...], he came forward to shake hands with Mazaud, whose customer he was. Leaning over, **he gave an order to the broker, which the latter inscribed upon a fiche**. 'He is selling his Suez,' murmured Moser." ;
- ▶ "and his way of listening so impressed Moser, that the latter, after making him his confidant, was frequently so disconcerted by his silence that **he ran off to countermand an order**."
- ▶ "Saccard entered the account office. It was there that the head accountant [...] went through **the memorandum-book, handed him by the broker every afternoon after the Bourse**, and entered to the various customers the sales and purchases which had been effected according to their orders. In doing this, he referred to the numerous fiches in order to ascertain the customers' names, for these did not appear in the memorandum-book, which contained only brief notes of the transactions: **such a stock, such an amount bought or sold, at such a rate, from such a broker**."

Order-driven vs. Dealer markets

Order-driven markets

- ▶ Market participants can trade with one another.
- ▶ Bid and offers from all market participants are displayed in a limit order book.
- ▶ Very frequent in modern equity markets.
- ▶ Some more or less public data available to academics.

Dealer/Quote-driven markets

- ▶ A dealer acts as counterpart to his clients.
- ▶ Only dealers' bid and offer are available ; use of requests for quotes (RFQ) to get bid/ask prices. Interdealer markets to manage risks and positions.
- ▶ Foreign exchange (FX) and corporate bonds markets are often dealers markets.
- ▶ Private data: less studies available

In this course: analysis of high-frequency data from electronic order-driven markets. But remember that all presentations simplify trading mechanisms, and all exchanges are more or less hybrid (because of their history, evolving rules and regulations, etc.)

Table of contents

Timescales in finance and high-frequency data

Limit order books

A bit of history

LOB basic mechanisms

Models vs real markets

High-frequency financial datasets

Different types of datasets

Quality of trades and quotes datasets

Some observations on prices

Prices

Spread and tick size

Some observations on trades

Durations, volumes, trading activity

Trade signs

Imbalance and trading activity

Some observations on LOB

Shape, activity, order lifetimes

Conditional probabilities of the order flow

Labs and Lab 1

General information

Lab 1

Limit order books in electronic order-driven markets

VOD.L VODAFONE GROUP PLC ORD USD 0.11 3/7						
Last	AT	162.95				15:42
Size		2,517	Mid	162.95	Period	SMMP
Time		15:42	Mid Change	-0.10 (-0.06%)		
Change		-0.10 (-0.06%)	LAT	162.95	Uncross Price	163.25
Trades		10,457	LAT Change	-0.10 (-0.06%)	Uncross Vol	2,234,586
Cum Vol		55,950,283	LAT High	163.7	VWAP	162.3293
LAT Vol		53,810,923	LAT Low	160.75	A-VWAP	162.33515
P Close		163.05	ISIN	GB00B16GWD56	Market Cap	85,894,857,429
Open		163.25	NSIN	B16GWD5	P.E.	9.91
High		163.7	Cur		Yield	4.82
Low		160.75	NMS	35,000	Dividend	7.86
52w Hi		164.4	Segment	SET0	Div-EPS Cur	GBP
52w Lo		125	Bid Indicator	-	Ex Div	02 Jun 10
Order Book VOD.L						
4	71006	162.9 -162.95	79959	9		
261	6,825,863	155,42523 -167,88226	7,839,197	432		
Cumul	Maker	Size	Bid	Ask	Size	Maker
4		71006	162.90	162.95	79959	9
10		110436	162.85	163.00	165547	11
11		194292	162.80	163.05	95435	15
14		165796	162.75	163.10	246286	18
16		319872	162.70	163.15	237244	14
10		224002	162.65	163.20	229145	13
7		163907	162.60	163.25	304053	13
4		108296	162.55	163.30	266717	13
3		90365	162.50	163.35	169815	8
25		165282	162.45	163.40	177534	7
1		30702	162.40	163.45	173809	5

- ▶ *Left:* snapshot of the Vodafone limit order book, London Stock Exchange, 2010
- ▶ **Order:** message sent to the exchange to buy or sell a security

The limit order book is the central structure aggregating all displayed buy and sell intentions (orders) by market participants.

- ▶ The financial exchange (e.g., Euronext in Paris) is in charge of maintaining the limit order book and matching orders resulting in transactions.

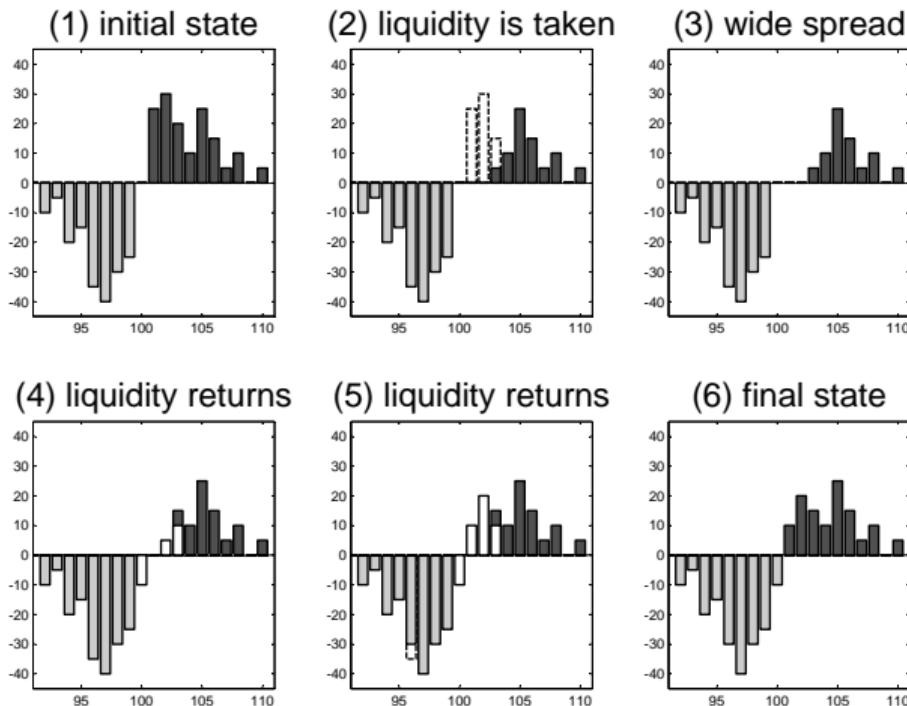
Schematic representation and basic interactions



- ▶ Double priority in price and time.

- ▶ **Limit order:** specifies a quantity and a maximum purchase price or minimum selling price ; price-time priority.
- ▶ **Market order:** specifies a quantity to be executed immediately at the best price(s) obtainable when the order reaches the order book ; has priority over all other orders.
- ▶ **Cancellation** of any order that has not been executed or canceled.

Schematic LOB dynamics



Source : [1]

Table of contents

Timescales in finance and high-frequency data

Limit order books

A bit of history

LOB basic mechanisms

Models vs real markets

High-frequency financial datasets

Different types of datasets

Quality of trades and quotes datasets

Some observations on prices

Prices

Spread and tick size

Some observations on trades

Durations, volumes, trading activity

Trade signs

Imbalance and trading activity

Some observations on LOB

Shape, activity, order lifetimes

Conditional probabilities of the order flow

Labs and Lab 1

General information

Lab 1

Real markets I - Type of orders

- ▶ Limit, market and cancellations for modeling purposes
- ▶ Multiple types of orders, variations across exchanges
- ▶ From the Euronext Trading Manual:
 - ▶ Order types: Pure Market orders, Limit orders, Market-to-limit orders, Stop-orders, Stop limit orders, Pegged orders
 - ▶ Validity parameters: Day Validity, Good Till Date validity, Good Till Cancelled validity, Good Till Time validity, Immediate Or Cancel validity, Fill Or Kill validity, Valid For Auction, Valid For Closing
 - ▶ Execution parameters: Minimum-quantity orders (in continuous trading only), All-or-none orders (Only applicable for Euronext Bond Match)
 - ▶ Transparency parameters: Iceberg orders
- ▶ Availability of order types and parameters depend on the trading platform (within the exchange).

Real markets II - Trading day

- ▶ Trading cycles are exchange- and asset-dependent.
- ▶ (Euronext) “Continuous” trading:
 - ▶ Pre-opening phase (order accumulation period)
 - ▶ Opening auction
 - ▶ Main trading session (continuous double auction)
 - ▶ Pre-closing phase (order accumulation period)
 - ▶ Closing auction
 - ▶ Trading-at-last phase, After hours trading
- ▶ (Euronext) Auction trading (once or twice a day):
 - ▶ Call phase (order accumulation period)
 - ▶ Auction
 - ▶ Trading-at-last phase and After hours trading
- ▶ Auctions will not be addressed in this course ([in labs: data from the continuous trading phase](#)), although volume traded on auctions is reported to have steadily increased in the past few years. Auctions are an active research subject.

Real markets III - Regulations, fragmented markets

- ▶ 2004: Regulation National Market System (Reg NMS) in the USA
- ▶ 2007: Markets in Financial Instruments Directive (MiFID) in Europe
- ▶ 2018: MiFID II in Europe
- ▶ Regulations impose organisational rules to access different trading venues (NBBO, EBBO)
- ▶ Since 2013, retail limit order book in Euronext Paris.
- ▶ Issues linked to fragmented markets will not be addressed in this course (in labs: data from a unique exchange).

LOB and microstructure modeling

Why do limit order books matter ?

- ▶ All trades have a market impact.
- ▶ (Most) traders observe the order book before trading.
- ▶ Uninformed trading means potential loss to wiser/faster traders.

Broad aims and challenges of modeling

- ▶ Build models able to reproduce/explain empirical facts of the time series of prices and events.
- ▶ Link micro and macro timescales, i.e. microstructure models and standard price models ; better reproduce stylized facts.
- ▶ Import LOB observations (trading signals) into trading strategies, optimal execution strategies, etc.

Table of contents

Timescales in finance and high-frequency data

Limit order books

A bit of history

LOB basic mechanisms

Models vs real markets

High-frequency financial datasets

Different types of datasets

Quality of trades and quotes datasets

Some observations on prices

Prices

Spread and tick size

Some observations on trades

Durations, volumes, trading activity

Trade signs

Imbalance and trading activity

Some observations on LOB

Shape, activity, order lifetimes

Conditional probabilities of the order flow

Labs and Lab 1

General information

Lab 1

Table of contents

Timescales in finance and high-frequency data

Limit order books

A bit of history

LOB basic mechanisms

Models vs real markets

High-frequency financial datasets

Different types of datasets

Quality of trades and quotes datasets

Some observations on prices

Prices

Spread and tick size

Some observations on trades

Durations, volumes, trading activity

Trade signs

Imbalance and trading activity

Some observations on LOB

Shape, activity, order lifetimes

Conditional probabilities of the order flow

Labs and Lab 1

General information

Lab 1

Databases on electronic markets

- ▶ Many types of databases can be found:
 - ▶ Daily prices (e.g., OHLC data available on publicly available financial news services)
 - ▶ Intraday transaction prices (trades data)
 - ▶ Best bid and offer (quotes data)
 - ▶ More or less detailed LOB data (see below)
 - ▶ On some markets : ids of market participants (brokers) may be available
- ▶ A recent typology for LOB data:
 - ▶ L1 : best bid and offers (best quotes)
 - ▶ L2 : aggregated view of (all the / 10 best / 5 best) limits of the order book
 - ▶ L3 : databases with order granularity (order id and modifications)
- ▶ In general, databases can be built by market participants (stream recording), or bought directly to the exchanges, or bought to third-party vendors.

Example of trades data

ts	price	volume
2017-01-02 09:00:15.576	48.470	100
2017-01-02 09:00:15.576	48.465	113
2017-01-02 09:00:15.576	48.465	100
2017-01-02 09:00:29.442	48.500	216
2017-01-02 09:00:38.125	48.500	177
2017-01-02 09:00:38.125	48.500	200
2017-01-02 09:00:44.755	48.510	297
2017-01-02 09:00:45.815	48.575	339
2017-01-02 09:00:50.105	48.590	210
2017-01-02 09:00:56.174	48.580	11
2017-01-02 09:01:04.474	48.530	91
2017-01-02 09:01:04.474	48.530	85
2017-01-02 09:01:04.474	48.530	84
2017-01-02 09:01:06.474	48.530	67
2017-01-02 09:01:06.484	48.530	31

- ▶ Example of a trades dataset for a stock traded in Euronext Paris.
- ▶ Millisecond precision of the timestamps.
- ▶ To be used in labs.

Example of L2 quotes data

ts	bp1	bq1	ap1	aq1	bp2	bq2	ap2	aq2	bp3	...	ap8	aq8	bp9	bq9	ap9	aq9	bp10	bq10	ap10	aq10
2017-01-12 09:02:23.013	48.335	384.0	48.35	86	48.33	1457	48.365	472	48.325	...	48.395	1178	48.295	728	48.4	710	48.290	430	48.41	681
2017-01-12 09:02:23.013	48.335	504.0	48.35	86	48.33	1457	48.365	472	48.325	...	48.395	1178	48.295	728	48.4	710	48.290	430	48.41	681
2017-01-12 09:02:23.013	48.335	734.0	48.35	86	48.33	1457	48.365	472	48.325	...	48.395	1178	48.290	430	48.4	710	48.285	741	48.41	681
2017-01-12 09:02:23.013	48.335	734.0	48.35	86	48.33	1457	48.365	472	48.325	...	48.395	1178	48.290	430	48.4	710	48.285	741	48.41	681
2017-01-12 09:02:23.013	48.335	734.0	48.35	86	48.33	1457	48.365	472	48.325	...	48.395	1178	48.290	430	48.4	710	48.285	741	48.41	681
2017-01-12 09:02:23.013	48.335	734.0	48.35	86	48.33	1457	48.365	472	48.325	...	48.395	1178	48.290	430	48.4	710	48.285	741	48.41	681
2017-01-12 09:02:23.093	48.335	734.0	48.35	86	48.33	1457	48.365	472	48.325	...	48.395	1178	48.295	728	48.4	710	48.290	430	48.41	681
2017-01-12 09:02:23.093	48.335	734.0	48.35	86	48.33	1457	48.365	472	48.325	...	48.395	1178	48.295	728	48.4	710	48.290	430	48.41	681

- ▶ Example of L2 dataset for a stock traded in Euronext Paris.
- ▶ Here, consecutive snapshots of the 10 best limits of the LOB.
- ▶ Millisecond precision of the timestamps.
- ▶ To be used in labs.

Another example of L2 quotes data

timestamp	side	level	price	quantity
36003.97	A	3	27.585	697
36003.97	A	3	27.585	177
36004.067	A	3	27.59	391
36004.067	A	4	27.595	311
36004.067	A	5	27.6	427
36004.067	A	6	27.605	1688
36004.067	A	7	27.61	586
36004.067	A	8	27.615	677
36004.067	A	9	27.62	1999
36004.067	A	10	27.625	568
36004.613	A	6	27.605	2315
36005.314	B	1	27.55	6829

- ▶ Example of L2 dataset for a stock traded in Euronext Paris.
- ▶ Here, one line for each changes.
- ▶ Millisecond precision of the timestamps (in seconds).

Example of L3 order data

o_id_fd	o_t_va	o_m_va	o_bs	o_price	o_q_ini	o_q_rem	o_state	o_validity	o_origin	o_account	o_member
	09:09:40	726516	B	45.320	276	276.0	4	0.0	A	6.0	MIX
	09:09:40	727164	B	45.320	272	272.0	4	0.0	A	6.0	MIX
	09:09:40	729410	S	45.340	173	173.0	4	0.0		3.0	MIX
	09:09:40	729418	S	45.350	171	171.0	4	0.0		3.0	MIX
	09:09:40	729937	S	45.340	122	0.0	2	0.0	A	6.0	MIX
	09:09:40	731937	S	45.340	570	570.0	5	0.0		3.0	HFT
	09:09:40	732723	B	45.320	313	313.0	4	0.0		6.0	MIX
	09:09:40	733221	S	45.365	339	339.0	4	0.0		6.0	MIX
	09:09:40	733236	S	45.375	317	317.0	4	0.0		6.0	MIX
	09:09:40	736283	S	45.345	122	122.0	4	0.0	A	6.0	MIX
	09:09:40	738185	S	45.340	114	0.0	2	0.0	A	6.0	MIX
	09:09:40	738496	S	45.340	151	151.0	5	0.0		6.0	MIX
	09:09:40	742776	B	45.305	200	200.0	4	0.0	A	2.0	HFT
	09:09:40	742808	B	45.310	200	200.0	4	0.0	A	2.0	HFT
	09:09:40	743020	B	45.310	200	200.0	4	0.0	A	2.0	HFT

- ▶ Example of L3 dataset for a stock traded in Euronext Paris.
- ▶ Microsecond precision of the timestamps.
- ▶ Vendor-formatted post-trading (added information)
- ▶ Partial view, many more fields available

Table of contents

Timescales in finance and high-frequency data

Limit order books

A bit of history

LOB basic mechanisms

Models vs real markets

High-frequency financial datasets

Different types of datasets

Quality of trades and quotes datasets

Some observations on prices

Prices

Spread and tick size

Some observations on trades

Durations, volumes, trading activity

Trade signs

Imbalance and trading activity

Some observations on LOB

Shape, activity, order lifetimes

Conditional probabilities of the order flow

Labs and Lab 1

General information

Lab 1

Order flow reconstruction from L2 data

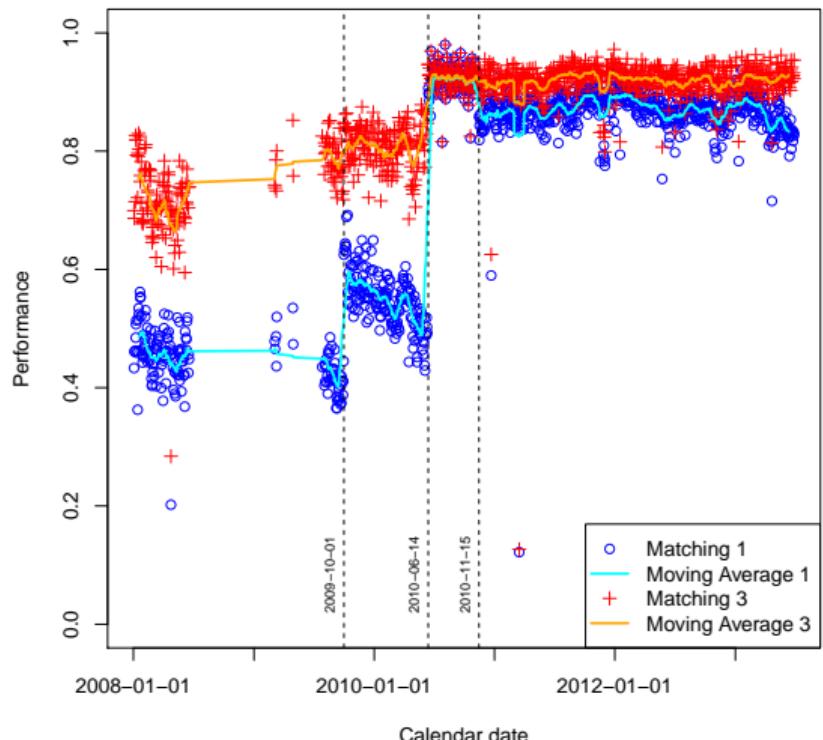
- ▶ L2 data can be used to reconstruct an approximation of the order flows in terms of limit orders, market orders and cancellations.

General reconstruction algorithm

1. Compute differences between L2 snapshots.
 2. Increases in quantities are limit orders.
 3. Decreases in quantities *outside the best quotes* are cancellations.
 4. Decreases in quantities at the best quotes that match an event in the trades files are market orders.
 5. Decreases in quantities at the best quotes that do *not* match an event in the trades files are cancellations.
- ▶ Many difficulties : shifts when one limit is removed/added ; the trades and quotes files may not be synchronized ; multiple transactions in the trades files can be updated in one step in the quotes files ; etc.

Order flow reconstruction on Euronext Paris

Evolution of the matching performance through time BNPP.PA



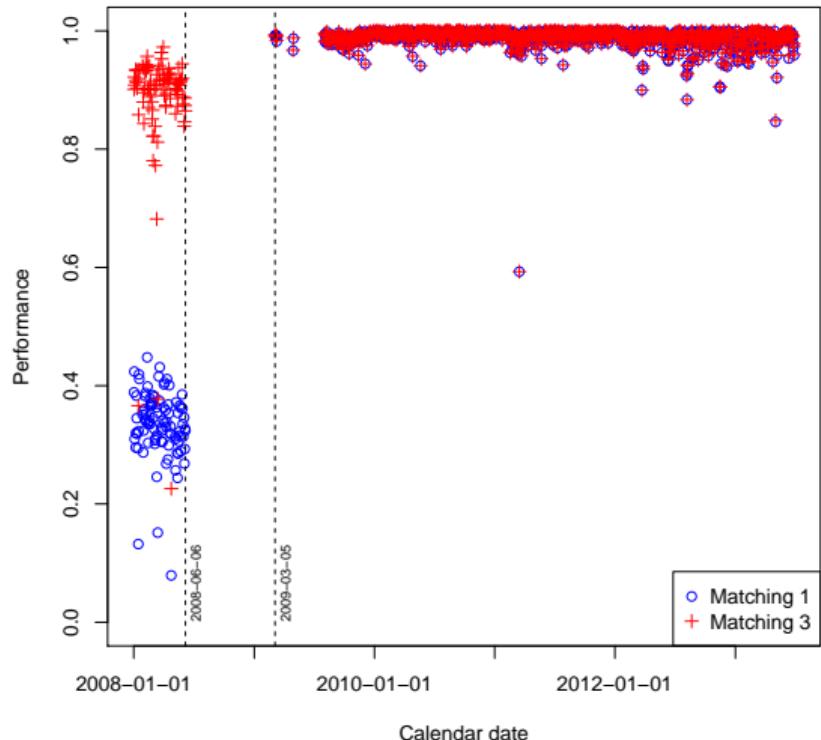
Important technical changes can be tracked through the performance of a matching algorithm

- ▶ October 1st, 2009: Change of primary datafeed on Euronext (Universal Trading Platform - Market Data)
- ▶ June 14th, 2010: Improvement of the best bid and offer quotes dissemination on UTP

Source : [8]

Order flow reconstruction on the London Stock Exchange

Evolution of the matching performance through time PSON.L



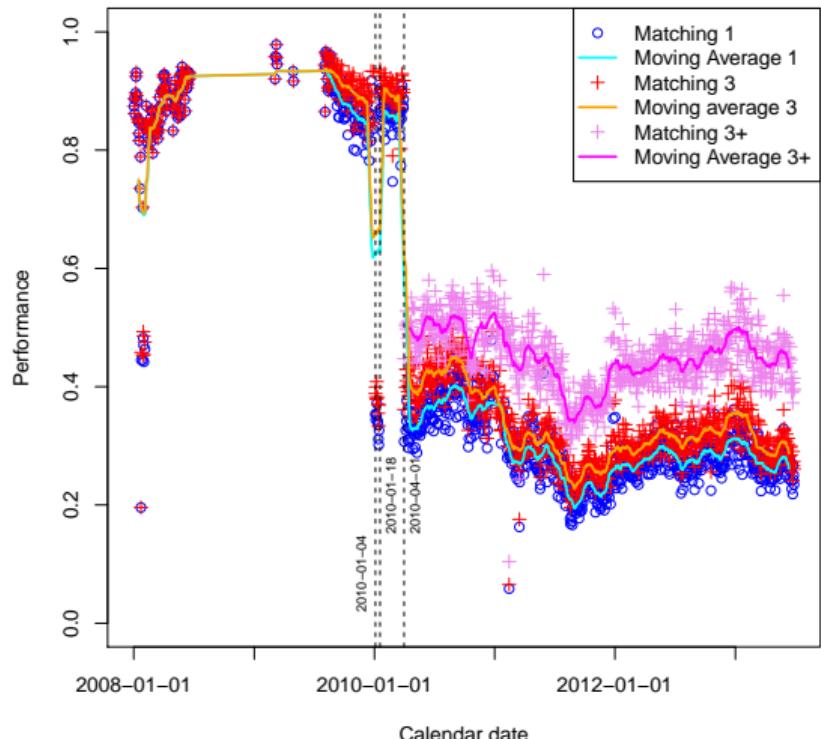
Important technical changes can be tracked through the performance of a matching algorithm

- December 2008 - January 2009: Implementation of the “Market Data Optimisation 2009” project, designed in particular to improve data feeds.

Source : [8]

Order flow reconstruction on Deutsche Börse

Evolution of the matching performance through time BMWG.DE



Important technical changes can be tracked through the performance of a matching algorithm

- ▶ April 1st, 2010: Access to un-netted market data restricted to the proprietary “Xetra Enhanced Broadcast Solution”.

Source : [8]

In brief

- ▶ Trades and quotes data is standard for empirical statistics of financial markets.
- ▶ Many different databases exist: exchange feeds, data providers, privately recorded databases (by desks/brokers/funds), "cleaned" /"repackaged" versions of previous databases, etc.
- ▶ A database, even from a well-known data provider, is not necessarily a consistent object. Long datasets should be carefully dealt with, because of technical changes (especially around 2008-2010) and of regulatory changes (e.g., MiFID II in 2018).
- ▶ Choices of data cleaning procedures and order flow reconstruction might influence modeling performances.
- ▶ In general, data quality improves with time, but technical choices and regularoty changes can break previous codes and analyses.

Table of contents

Timescales in finance and high-frequency data

Limit order books

A bit of history

LOB basic mechanisms

Models vs real markets

High-frequency financial datasets

Different types of datasets

Quality of trades and quotes datasets

Some observations on prices

Prices

Spread and tick size

Some observations on trades

Durations, volumes, trading activity

Trade signs

Imbalance and trading activity

Some observations on LOB

Shape, activity, order lifetimes

Conditional probabilities of the order flow

Labs and Lab 1

General information

Lab 1

Table of contents

Timescales in finance and high-frequency data

Limit order books

A bit of history

LOB basic mechanisms

Models vs real markets

High-frequency financial datasets

Different types of datasets

Quality of trades and quotes datasets

Some observations on prices

Prices

Spread and tick size

Some observations on trades

Durations, volumes, trading activity

Trade signs

Imbalance and trading activity

Some observations on LOB

Shape, activity, order lifetimes

Conditional probabilities of the order flow

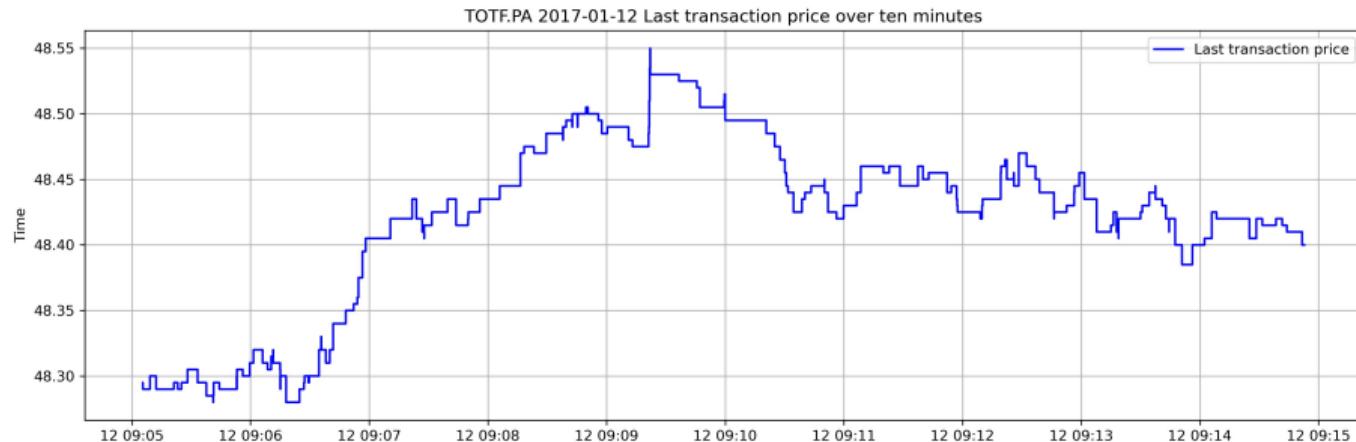
Labs and Lab 1

General information

Lab 1

Last transaction price

The last transaction price is the simplest definition of the price of a financial asset.



- The price function $p : t \mapsto p(t)$ is a step function with values in a discrete set.
- Fully defined with the trades data.

Multiple price definitions

Trades and quotes files define multiple *intraday* prices at time t :

- ▶ trade price : price of the last transaction
- ▶ bid price $b(t)$: price of best available buy limit order (bid, demand)
- ▶ ask price $a(t)$: price of best available sell limit order (ask, supply)
- ▶ mid price : $0.5 \times (b(t) + a(t))$
- ▶ weighted mid-price : see below.
- ▶ ...



Prices of the LOB queues

- ▶ Evolution of the 10 best bid and ask queues for the LOB of AXAF.PA on January 2nd, 2017.

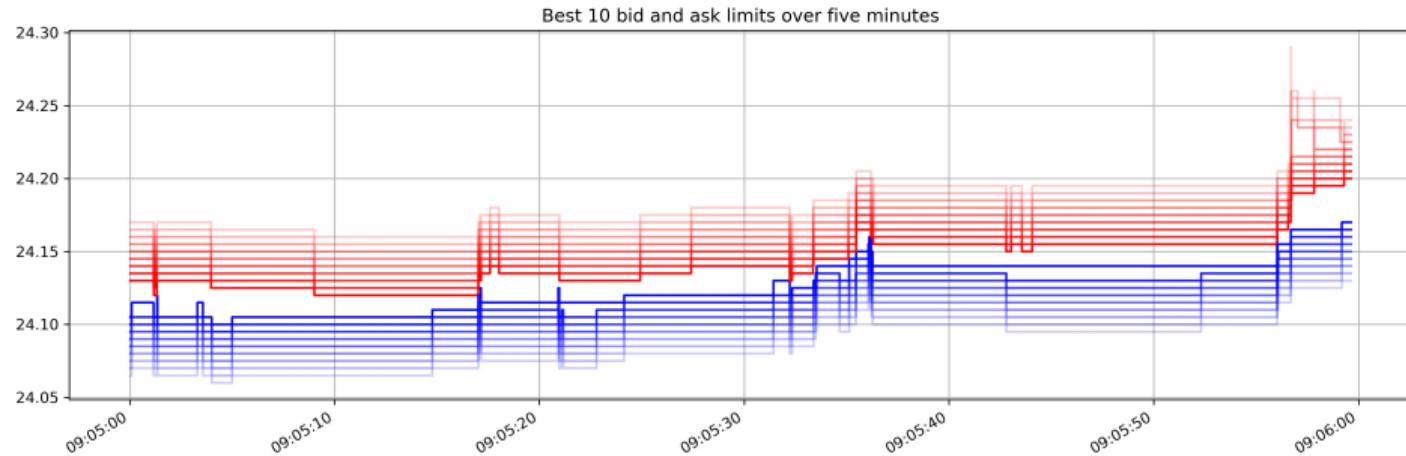


Table of contents

Timescales in finance and high-frequency data

Limit order books

A bit of history

LOB basic mechanisms

Models vs real markets

High-frequency financial datasets

Different types of datasets

Quality of trades and quotes datasets

Some observations on prices

Prices

Spread and tick size

Some observations on trades

Durations, volumes, trading activity

Trade signs

Imbalance and trading activity

Some observations on LOB

Shape, activity, order lifetimes

Conditional probabilities of the order flow

Labs and Lab 1

General information

Lab 1

Ticksizes

The ticksize is the minimum increment of the price of an asset. All prices are integer multiples of the ticksize.

- ▶ The ticksize must not be too large in order to allow sufficient granularity of prices, but must not be too small in order to limit the competition on price priority.
- ▶ Euronext ticksizes until December 31st, 2017 for liquid stocks

Price from	To	Tick size
€0	€9.999	€0.001
€10.000	€49.995	€0.005
€50.000	€99.99	€0.01
€100.00		€0.05

- ▶ The ticksize is one of the most defining element of the microstructure of a financial asset.

Euronext spread definition (MIFID II, ESMA regulation)

Starting January 1st, 2018, the regulator introduced new tick size regimes.

- ▶ Spread depends on price and liquidity.
- ▶ An average spread between 1.5 and 2 ticks is targeted by the regulator. See next slide.

Price ranges	Liquidity bands					
	0 ≤ Average daily number of transactions < 10	10 ≤ Average daily number of transactions < 80	80 ≤ Average daily number of transactions < 600	600 ≤ Average daily number of transactions < 200	200 ≤ Average daily number of transactions < 9000	9000 ≤ Average daily number of transactions
0 ≤ price < 0.1	0.0005	0.0002	0.0001	0.0001	0.0001	0.0001
0.1 ≤ price < 0.2	0.001	0.0005	0.0002	0.0001	0.0001	0.0001
0.2 ≤ price < 0.5	0.002	0.001	0.0005	0.0002	0.0001	0.0001
0.5 ≤ price < 1	0.005	0.002	0.001	0.0005	0.0002	0.0001
1 ≤ price < 2	0.01	0.005	0.002	0.001	0.0005	0.0002
2 ≤ price < 5	0.02	0.01	0.005	0.002	0.001	0.0005
5 ≤ price < 10	0.05	0.02	0.01	0.005	0.002	0.001
10 ≤ price < 20	0.1	0.05	0.02	0.01	0.005	0.002
20 ≤ price < 50	0.2	0.1	0.05	0.02	0.01	0.005
50 ≤ price < 100	0.5	0.2	0.1	0.05	0.02	0.01
100 ≤ price < 200	1	0.5	0.2	0.1	0.05	0.02
200 ≤ price < 500	2	1	0.5	0.2	0.1	0.05
500 ≤ price < 1000	5	2	1	0.5	0.2	0.1
1000 ≤ price < 2000	10	5	2	1	0.5	0.2
2000 ≤ price < 5000	20	10	5	2	1	0.5
5000 ≤ price < 10000	50	20	10	5	2	1
10000 ≤ price < 20000	100	50	20	10	5	2
20000 ≤ price < 50000	200	100	50	20	10	5
50000 ≤ price	500	200	100	50	20	10

Spread and ticksize

The bid-ask spread is the distance between the best bid price and the best ask price.

- ▶ Spread is fundamental in microstructure : it is a proxy of the cost of transaction.
- ▶ Spread is constrained by the ticksize.
- ▶ Large-tick stocks and small-tick stocks exhibit different properties.

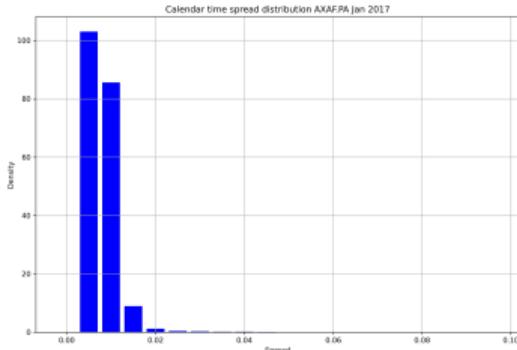
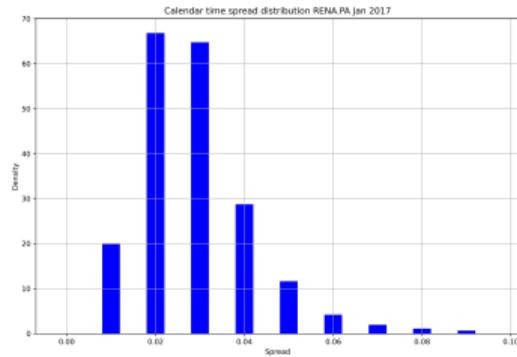


Table of contents

Timescales in finance and high-frequency data

Limit order books

A bit of history

LOB basic mechanisms

Models vs real markets

High-frequency financial datasets

Different types of datasets

Quality of trades and quotes datasets

Some observations on prices

Prices

Spread and tick size

Some observations on trades

Durations, volumes, trading activity

Trade signs

Imbalance and trading activity

Some observations on LOB

Shape, activity, order lifetimes

Conditional probabilities of the order flow

Labs and Lab 1

General information

Lab 1

Table of contents

Timescales in finance and high-frequency data

Limit order books

A bit of history

LOB basic mechanisms

Models vs real markets

High-frequency financial datasets

Different types of datasets

Quality of trades and quotes datasets

Some observations on prices

Prices

Spread and tick size

Some observations on trades

Durations, volumes, trading activity

Trade signs

Imbalance and trading activity

Some observations on LOB

Shape, activity, order lifetimes

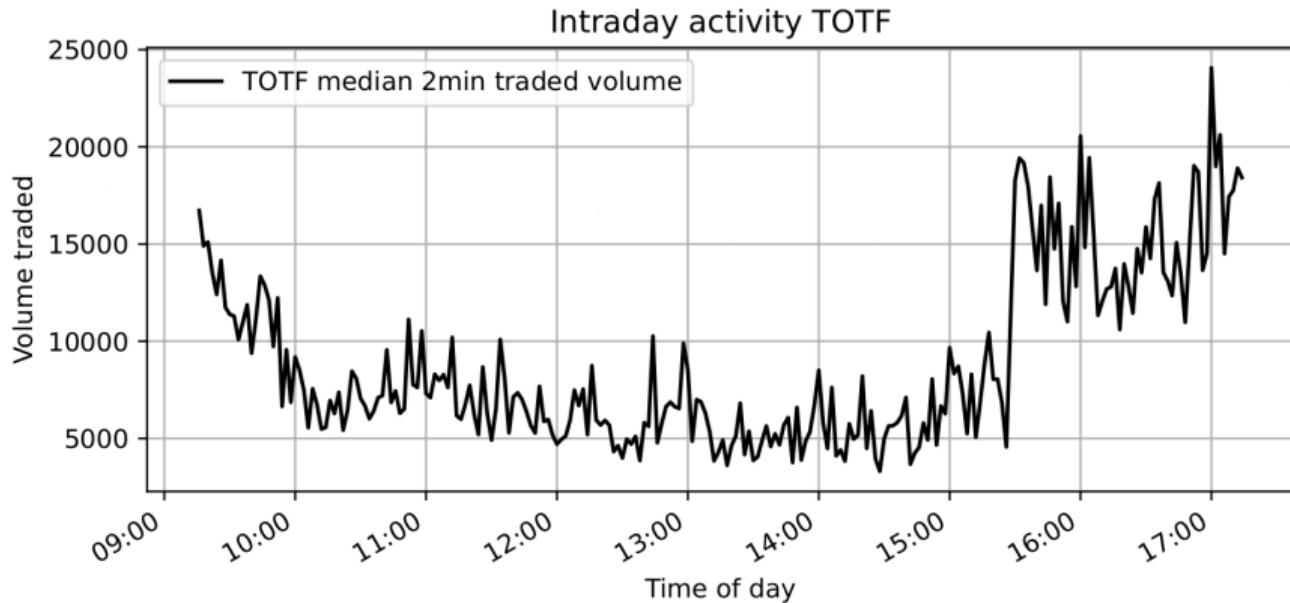
Conditional probabilities of the order flow

Labs and Lab 1

General information

Lab 1

Intraday trading activity



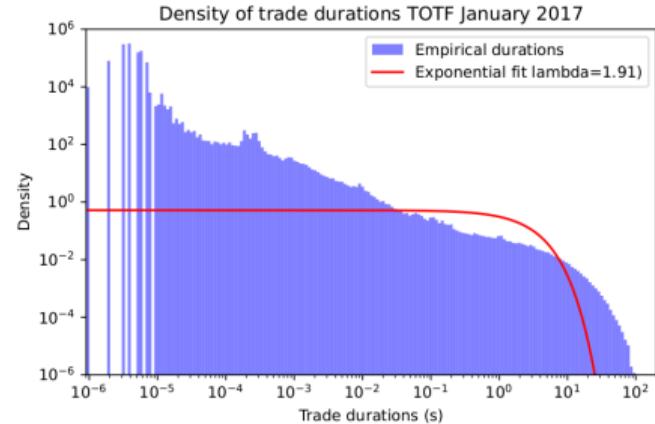
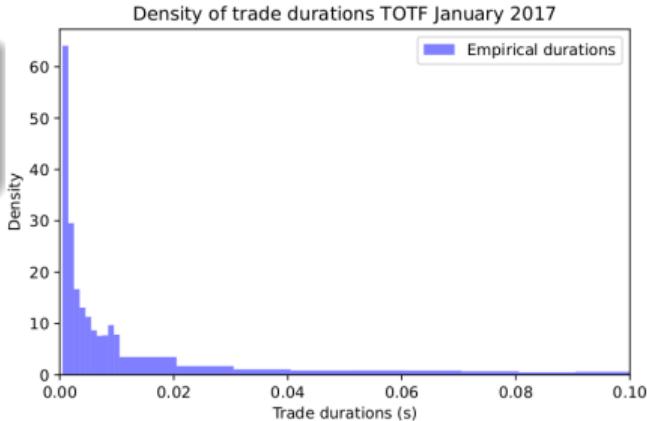
Intraday trading activity often has a typical U-shaped pattern.

- ▶ In Europe, opening of American market is often visible on trading activity.
- ▶ Post-opening auction and pre-closing auction are the busiest phases. Volume traded at auctions dramatically increased in recent years.

Trade durations

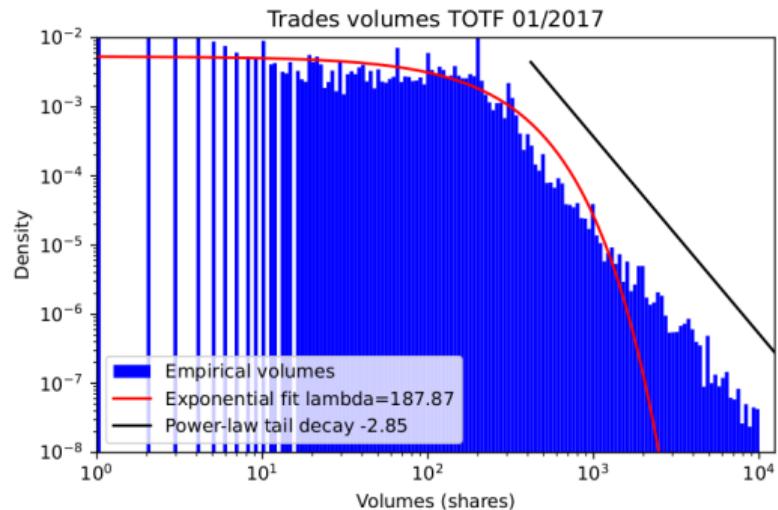
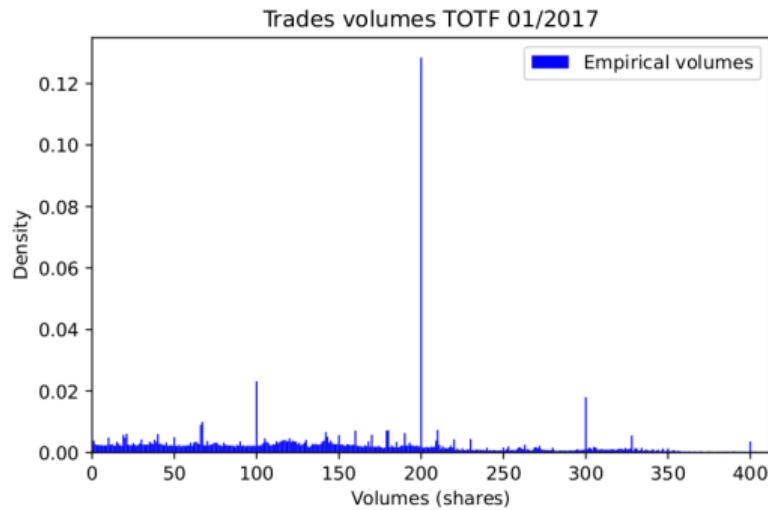
Trade durations are not exponentially distributed.

- ▶ High peak of short durations. Tail heavier than exponential.
- ▶ If data is precise up to the microsecond, knowledge of the data recording process is important : speed of the matching engine, of the recorder, colocation, etc.
- ▶ The order of magnitude of reaction time (round-trip) of high-frequency agents on equity markets is often estimated to be in a few 10s or 100s microseconds.
- ▶ *More later with discussions on Poisson modeling, Hawkes modeling, subordination, etc.*



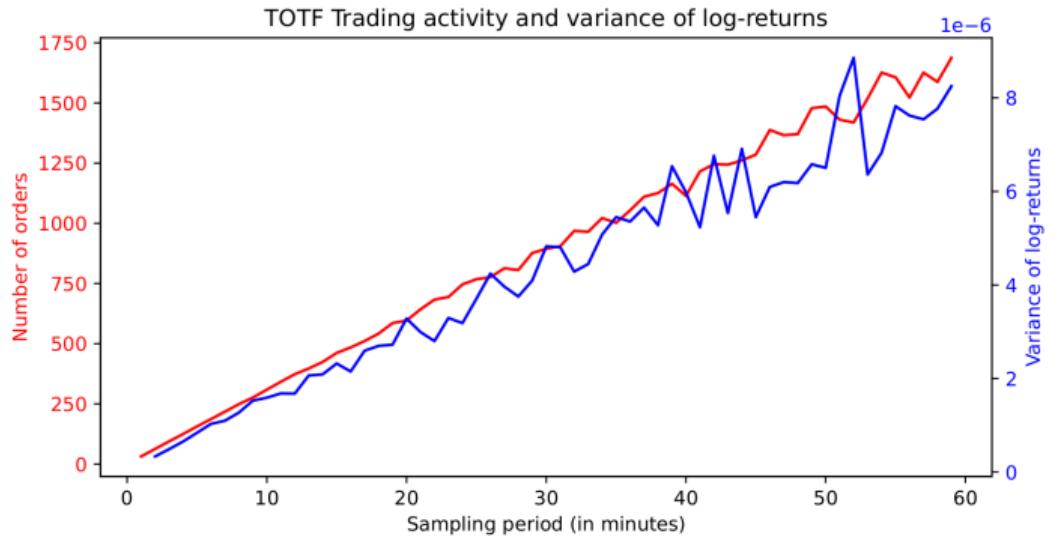
Trade volumes

Distributions of transactions volumes exhibit often exhibit peaks at round numbers of shares, and a heavy tail.



- ▶ On some exchanges, one may also observe peaks for some number of shares corresponding to a round volume in currency (e.g., 10,000 USD).

Trading activity and volatility



The variance of the stock price is linked to the trading activity .

- ▶ Such observations suggest that *event time* may be an interesting clock to measure time on financial markets. See, e.g., [11] and the notion of subordination of stochastic processes. You'll be able to test this assertion in labs.

Table of contents

Timescales in finance and high-frequency data

Limit order books

A bit of history

LOB basic mechanisms

Models vs real markets

High-frequency financial datasets

Different types of datasets

Quality of trades and quotes datasets

Some observations on prices

Prices

Spread and tick size

Some observations on trades

Durations, volumes, trading activity

Trade signs

Imbalance and trading activity

Some observations on LOB

Shape, activity, order lifetimes

Conditional probabilities of the order flow

Labs and Lab 1

General information

Lab 1

Trade sign

The sign of a trade designates the side of *aggressor*, i.e. the order triggering the transaction.

- ▶ +1 denotes an aggressive buyer, typically a buy market order removing liquidity on the ask side of the LOB.
- ▶ -1 denotes an aggressive seller, typically sell market order removing liquidity on the bid side of the LOB.
- ▶ Trade sign is an important information in microstructure theory: it is often assumed that the aggressive buyer/seller is trades with some knowledge/information, hence its *impatience*.

Trade signature

- ▶ Trade sign is often available in trades databases.
- ▶ *If trade signs are not available in the database*, one may compute sign proxies with trades data or trades and quotes data:
 - ▶ **Tick test** : a trade is signed $+1$ if its price is strictly greater than the previous one, -1 if it is strictly lower, and has the same sign as the previous trade if their prices are equal. No use of quotes data.
 - ▶ **Quote test**: if quotes data is available, transaction price can be compared to the best quotes to determine the sign of the trade ($+1$ if the order is executed at the best ask price, -1 if the order is executed at the best bid price). Difficulty : need to adjust for non-synchronized trades and quotes files. [7] suggests 5 seconds (!) on 1991 data. [8] finds a typical time shift of 100 ms in 2013 data. Recent high-quality data should be fully synchronized.

Autocorrelation of trade signs and order book events

Time series of trade signs exhibit long-memory.

- ▶ Rules out a random trading assumption.
- ▶ Evidence of order splitting, but not only.
- ▶ This can be generalized by attributing to any order book event a sign representing the expected impact on the price.
- ▶ One may also look at the side (bid/ask) impacted by any order book event.

More generally, time series of order book events signs/sides exhibit long-memory.

Source : [2, 6]

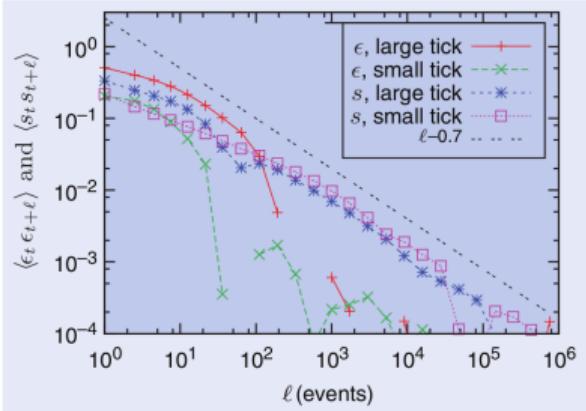
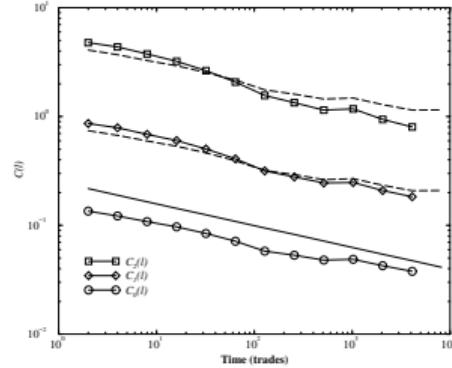


Table of contents

Timescales in finance and high-frequency data

Limit order books

A bit of history

LOB basic mechanisms

Models vs real markets

High-frequency financial datasets

Different types of datasets

Quality of trades and quotes datasets

Some observations on prices

Prices

Spread and tick size

Some observations on trades

Durations, volumes, trading activity

Trade signs

Imbalance and trading activity

Some observations on LOB

Shape, activity, order lifetimes

Conditional probabilities of the order flow

Labs and Lab 1

General information

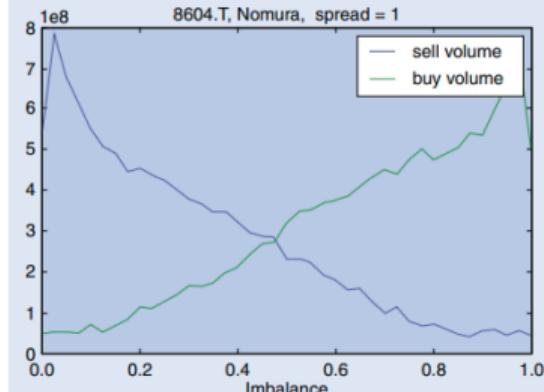
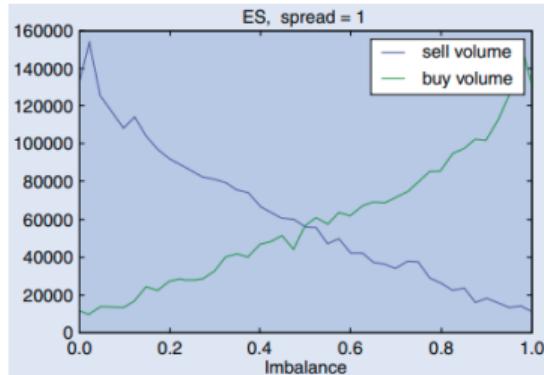
Lab 1

Imbalance and trading activity

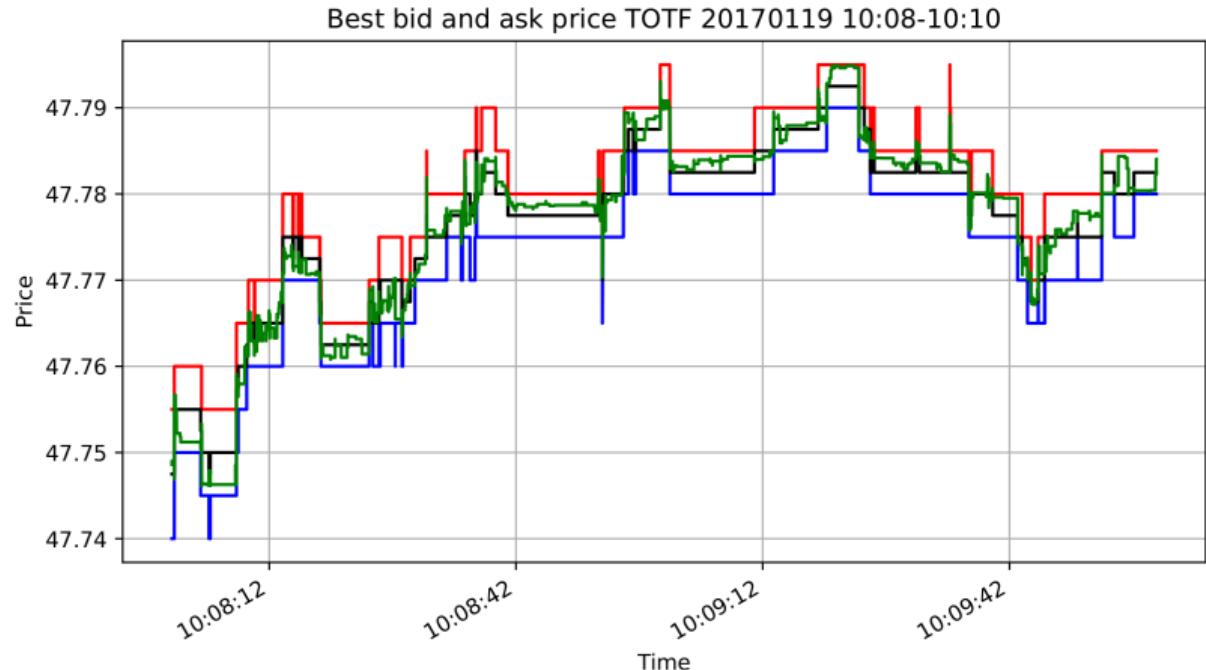
Imbalance $i(t) = \frac{q^B(t)-q^A(t)}{q^B(t)+q^A(t)}$ measures the (dis)equilibrium between the volumes available at the bid and ask best queues.

- ▶ Imbalance is a well-known variable with a strong influence on trading activity.
- ▶ Imbalance is indeed a signal used by market participants: strong imbalance suggests an imminent price movement
- ▶ The weighted mid-price $\frac{a(t)q^B(t)+b(t)q^A(t)}{q^B(t)+q^A(t)}$ "improves" the mid-price by taking into account the volumes at the best quotes.

Source: [12]



Weighted mid-price



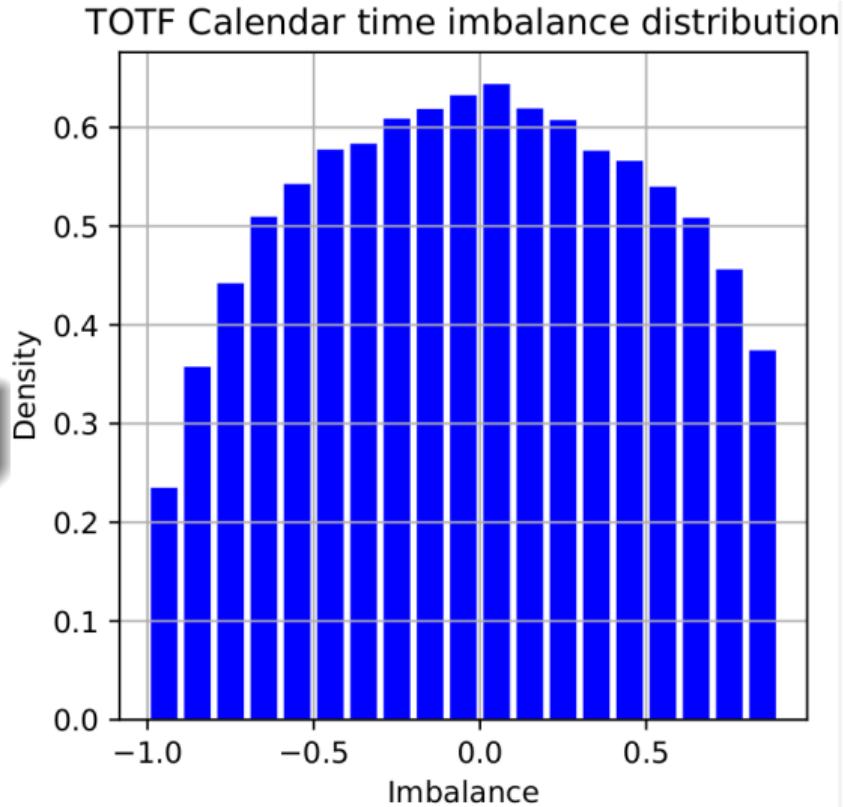
- ▶ Best ask price (red),
best bid price (blue),
mid price (black),
weighted mid-price
(green).

Imbalance distribution

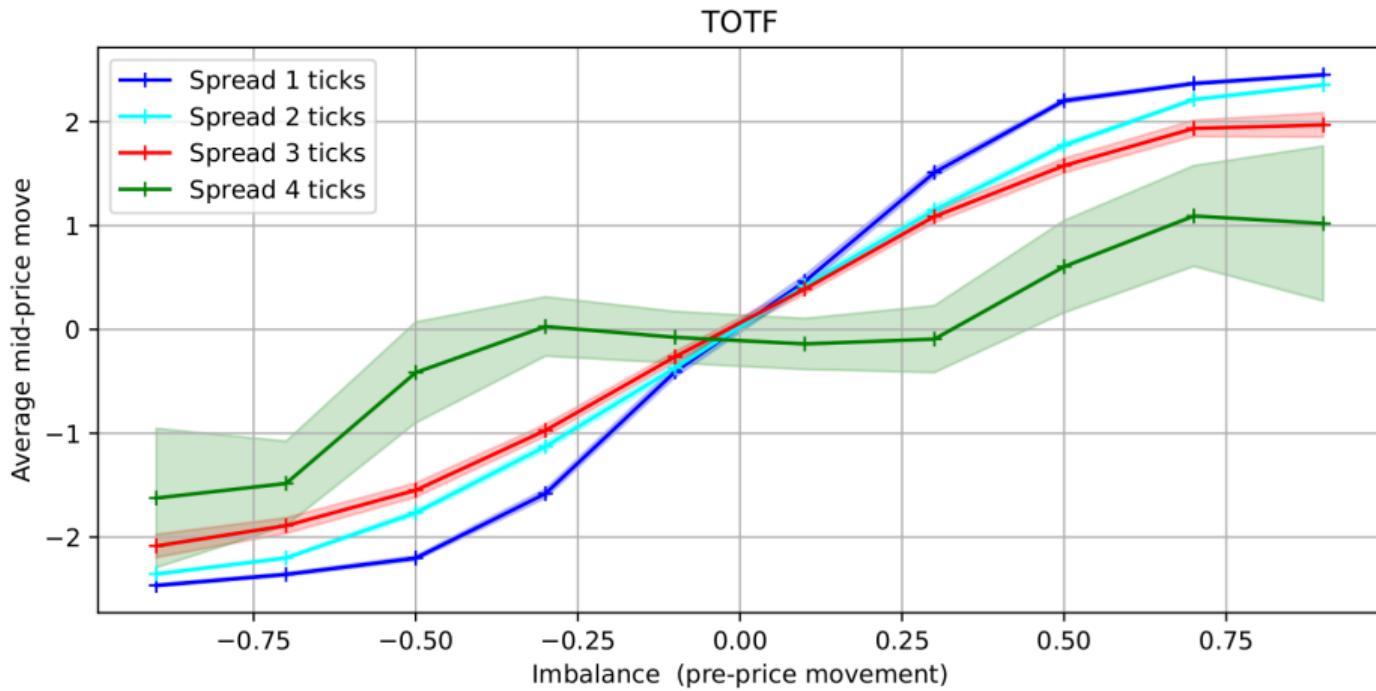
- ▶ If queues are i.i.d. exponential rvs, imbalance is uniformly distributed.

In liquid equity markets, unimodal distribution.

- ▶ Not the same in very small tick stocks, especially in cryptocurrency markets.



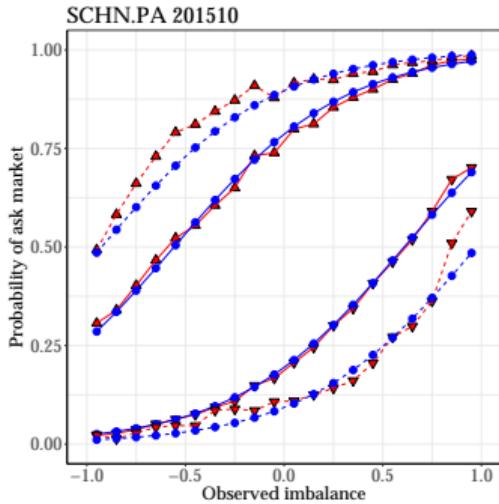
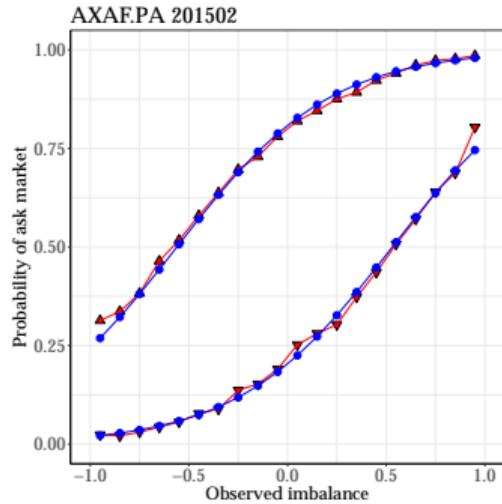
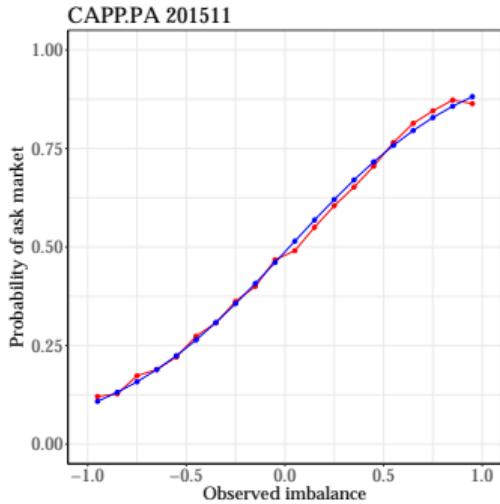
Imbalance before a mid-price moves



Imbalance is a well-known short term predictor of the price moves (y-axis in 10^{-3}€).

LOB state and history influence on trades

One can try to predict trade signs using imbalance, past trades, and spread.



Source : [9]

Table of contents

Timescales in finance and high-frequency data

Limit order books

A bit of history

LOB basic mechanisms

Models vs real markets

High-frequency financial datasets

Different types of datasets

Quality of trades and quotes datasets

Some observations on prices

Prices

Spread and tick size

Some observations on trades

Durations, volumes, trading activity

Trade signs

Imbalance and trading activity

Some observations on LOB

Shape, activity, order lifetimes

Conditional probabilities of the order flow

Labs and Lab 1

General information

Lab 1

Table of contents

Timescales in finance and high-frequency data

Limit order books

A bit of history

LOB basic mechanisms

Models vs real markets

High-frequency financial datasets

Different types of datasets

Quality of trades and quotes datasets

Some observations on prices

Prices

Spread and tick size

Some observations on trades

Durations, volumes, trading activity

Trade signs

Imbalance and trading activity

Some observations on LOB

Shape, activity, order lifetimes

Conditional probabilities of the order flow

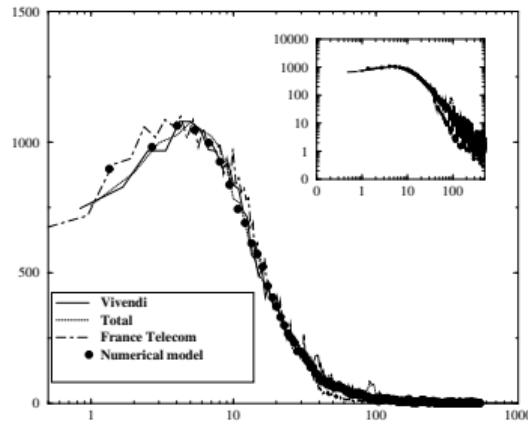
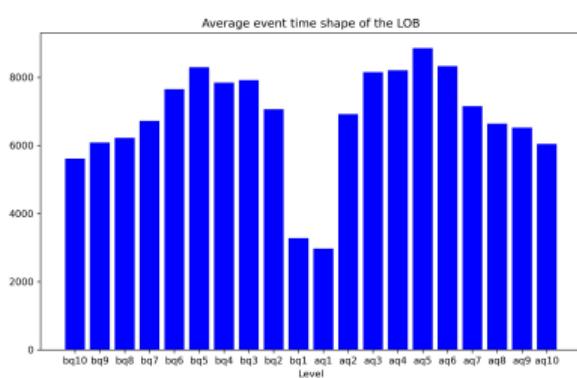
Labs and Lab 1

General information

Lab 1

Average shape of the LOB

LOB usually exhibits a humped-shape with a maximum away from the best prices.

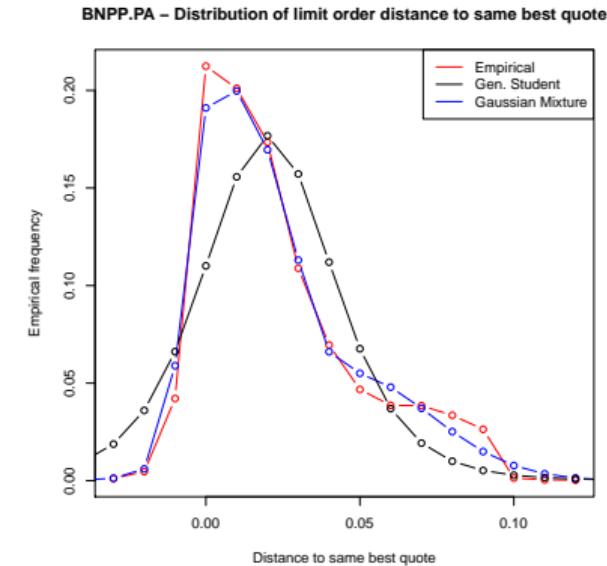
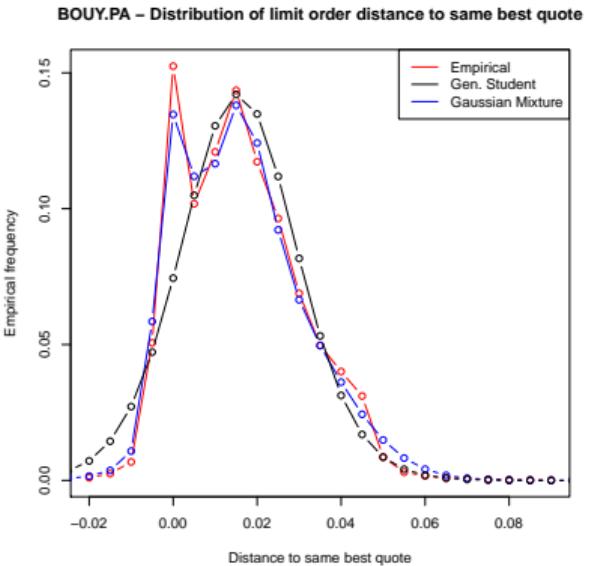


- Limits far away from the best limits reflects the interaction between limit orders and cancellations. Market orders must be taken into account to understand the behaviour around the best limits.

Source of right panel: [3]

Trading activity in the book

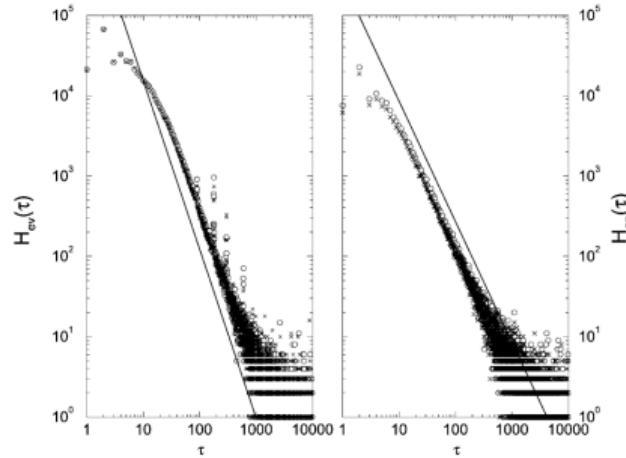
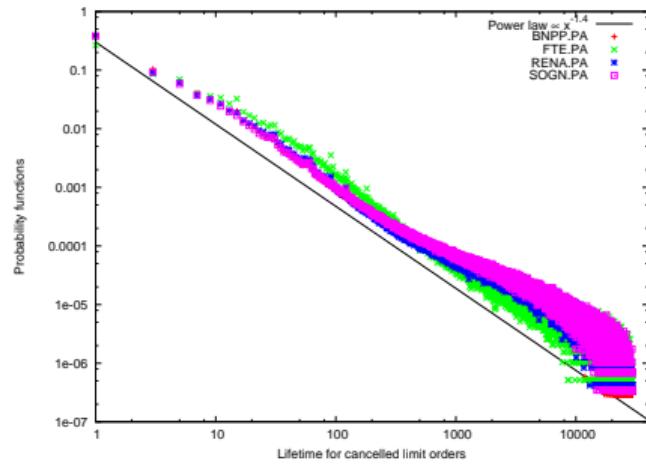
Most liquidity is provided at the best quotes.



Source : [10]

Lifetime of limit orders

Distribution of lifetime of executed or canceled limit orders exhibit heavy tails.



- Poisson/exponential assumptions will be useful for modeling but are not satisfied in practice.

Source : [5, 4]

Table of contents

Timescales in finance and high-frequency data

Limit order books

A bit of history

LOB basic mechanisms

Models vs real markets

High-frequency financial datasets

Different types of datasets

Quality of trades and quotes datasets

Some observations on prices

Prices

Spread and tick size

Some observations on trades

Durations, volumes, trading activity

Trade signs

Imbalance and trading activity

Some observations on LOB

Shape, activity, order lifetimes

Conditional probabilities of the order flow

Labs and Lab 1

General information

Lab 1

Some conditional probabilities of the order flow I

- ▶ A useful description of the order flow

Notation	Definition
M_{buy}^0, M_{sell}^0	buy/sell market order that does not change the mid price
M_{buy}^1, M_{sell}^1	buy/sell market order that changes the mid price
L_{buy}^0, L_{sell}^0	buy/sell limit order that does not change the mid price
L_{buy}^1, L_{sell}^1	buy/sell limit order that changes the mid price
C_{buy}^0, C_{sell}^0	buy/sell cancellation that does not change the mid price
C_{buy}^1, C_{sell}^1	buy/sell cancellation that changes the mid price

Some conditional probabilities of the order flow II

	L_{buy}^0	L_{sell}^0	C_{buy}^0	C_{sell}^0	M_{buy}^0	M_{sell}^0	L_{buy}^1	L_{sell}^1	C_{buy}^1	C_{sell}^1	M_{buy}^1	M_{sell}^1
L_{buy}^0	41.37	9.64	16.00	22.40	2.90	1.58	2.35	1.12	0.02	1.08	1.39	0.16
L_{sell}^0	9.61	41.79	21.95	16.12	1.61	2.96	1.02	2.29	1.05	0.02	0.15	1.44
C_{buy}^0	17.91	25.88	40.67	5.98	1.39	1.74	1.20	2.34	1.49	0.37	0.56	0.47
C_{sell}^0	25.18	17.98	6.04	41.30	1.79	1.42	2.08	1.27	0.37	1.49	0.51	0.60
M_{buy}^0	22.17	5.33	4.75	9.94	34.64	0.70	7.68	0.65	0.55	1.31	11.86	0.42
M_{sell}^0	5.60	21.14	10.61	5.01	0.72	34.32	0.53	7.19	1.48	1.10	0.42	11.88
L_{buy}^1	32.39	8.06	0.21	25.27	4.84	5.58	1.42	1.57	5.80	1.77	2.44	10.65
L_{sell}^1	7.65	29.94	26.04	0.22	5.63	5.62	1.39	1.36	1.42	5.39	12.37	2.96
C_{buy}^1	25.02	19.09	35.70	4.96	0.96	0.67	8.34	3.59	0.72	0.35	0.48	0.12
C_{sell}^1	21.48	23.28	5.42	34.70	0.76	1.16	3.20	7.88	0.63	0.75	0.18	0.57
M_{buy}^1	28.27	9.60	7.38	28.12	3.11	1.02	11.52	7.98	0.90	0.87	0.67	0.55
M_{sell}^1	11.83	23.05	33.36	7.24	1.04	3.13	6.79	9.34	1.05	1.81	0.66	0.70
O	22.82	22.93	19.80	20.03	2.99	3.00	2.07	2.12	0.85	0.88	1.27	1.26

Source: [1]

Some conditional probabilities of the order flow III

Possible interpretation

- ▶ Consensus reinforcement ($L_{buy}^0 \rightarrow L_{buy}^0, L_{sell}^0 \rightarrow L_{sell}^0$)
- ▶ Consensus breach ($C_{buy}^0 \rightarrow C_{buy}^0, C_{buy}^1$)
- ▶ Momentum effect and/or order splitting ($M_{buy}^0 \rightarrow M_{buy}^0$ and M_{buy}^1)
- ▶ Market taking ($L_{buy}^1 \rightarrow M_{sell}^1$)
- ▶ Move for liquidity ($C_{buy}^0 \rightarrow M_{sell}^1$)
- ▶ Market making ($M_{buy}^1 \rightarrow L_{buy}^1$ and $M_{sell}^1 \rightarrow L_{sell}^1$)
- ▶ Market making ($C_{buy}^1 \rightarrow L_{buy}^1$; but asymmetry)
- ▶ Market manipulation (?) ($L_{buy}^1 \rightarrow C_{buy}^1$)

But be careful with possible overinterpretation...

- ▶ Latency
- ▶ Mechanical reasons (order matching procedures, pegged orders, iceberg orders, . . .)

Table of contents

Timescales in finance and high-frequency data

Limit order books

A bit of history

LOB basic mechanisms

Models vs real markets

High-frequency financial datasets

Different types of datasets

Quality of trades and quotes datasets

Some observations on prices

Prices

Spread and tick size

Some observations on trades

Durations, volumes, trading activity

Trade signs

Imbalance and trading activity

Some observations on LOB

Shape, activity, order lifetimes

Conditional probabilities of the order flow

Labs and Lab 1

General information

Lab 1

Table of contents

Timescales in finance and high-frequency data

Limit order books

A bit of history

LOB basic mechanisms

Models vs real markets

High-frequency financial datasets

Different types of datasets

Quality of trades and quotes datasets

Some observations on prices

Prices

Spread and tick size

Some observations on trades

Durations, volumes, trading activity

Trade signs

Imbalance and trading activity

Some observations on LOB

Shape, activity, order lifetimes

Conditional probabilities of the order flow

Labs and Lab 1

General information

Lab 1

Labs - General information

- ▶ You should work by **groups of three** students.
- ▶ You're expected to work in **Python** with standard scientific and data libraries (see list below).
- ▶ The expected output of each lab is a single **Python notebook** with code, graphs and comments/analyses **Notebooks must be fully executable** from top to bottom in one run. Provided data should be in a directory 'Data' (See details below).
- ▶ For the lab *i*, your notebook **MUST** be named **DHF-LAB*i*-NAME1-NAME2-NAME3-Partj.ipynb**, where NAME1 and NAME2 and NAME3 are the names of the students of the group in capital letters (and without space).
- ▶ For each lab, a dedicated space will be opened on the **Edunao** page of the course. **Upload your notebook only**, nothing else (no data, no saved graphs, no pdf report).
- ▶ Each lab will have a specific deadline, please check it.

Labs - About the data provided

- ▶ Data used in the labs of this course is not public. This is research data provided via the FiQuant research team (MICS, CentraleSupélec, Université Paris-Saclay).
- ▶ The use of such data is restricted to teaching and research purposes.
- ▶ Data will be available on Edunao until the end of the course. You must not keep any copy of this data after the course has ended.
- ▶ **You must not use this data for any purpose other than the labs of this course. Keep a local copy only. You must not share this data. This data must not be uploaded to any github, gitlab, or any other platform.**

Labs - Dos. . .

- ▶ Keep the file structure of the data. The directory Data should be at the root of your Jupyter server, with your notebook.
- ▶ Decompress the global Data archive, but do not decompress individual data files. pandas can directly read .csv.gz files.
- ▶ Convert timestamps to datetime objects with pd.to_datetime(...). Matplotlib deals well with datetime and this is necessary for efficient sampling of the data when needed.
- ▶ Depending on your command of Python/pandas and the computing/memory resources of your computer, you may decide to develop/debug on a small subsample of the available data (e.g. a few days first), then use the full sample for final graphs and analyses.
- ▶ All graphs should have a legend, a title, labeled x-axis and y-axis with units if applicable. All graphs should be analyzed and commented.

Labs - ... and don'ts

- ▶ **Don't** leave any system command (pip install ...) in your notebooks.
- ▶ **Don't** use plotly (html/pdf conversion problems). Please use matplotlib instead.
- ▶ **Don't** use horizontal scrolling in your code cells. Please break your lines of code instead (in your IDE, add a vertical ruler at 80 characters).
- ▶ **Don't** plot log of (physical) quantities. Please use log scales instead.
- ▶ **Don't** leave plot randomly in the notebook without context. Please indicate question numbers and comment on your results (and if they are not what you were expecting, say it !).

Labs - Grading process

- ▶ The grading process of the labs is the following:
 1. Notebooks are converted to python scripts
 2. Python scripts are parsed by a plagiarism detection software.
 3. Python scripts are parsed for forbidden instructions (e.g., pip install...). Notebooks with forbidden instructions will not be graded.
 4. Notebooks are cleared of all outputs, fully reexecuted with clean kernels, then exported to html, and finally converted to pdf. Notebooks with execution problems will not be graded.
- ▶ Because of this grading pipeline:
 - ▶ Avoid complex latex formulas in your markdown comments.
 - ▶ Make sure your notebooks are fully executable: clear all outputs, kill and restart the kernel and finally run all cells in one pass before submitting.
 - ▶ Make sure your graphs are correctly exported to html.
- ▶ Notebooks will be executed in a Python environment containing **ONLY** : python 3.13.2 and its standard library, jupyter, numpy, scipy, pandas, matplotlib, powerlaw, statsmodels, tqdm, Hawkes.

Table of contents

Timescales in finance and high-frequency data

Limit order books

A bit of history

LOB basic mechanisms

Models vs real markets

High-frequency financial datasets

Different types of datasets

Quality of trades and quotes datasets

Some observations on prices

Prices

Spread and tick size

Some observations on trades

Durations, volumes, trading activity

Trade signs

Imbalance and trading activity

Some observations on LOB

Shape, activity, order lifetimes

Conditional probabilities of the order flow

Labs and Lab 1

General information

Lab 1

Lab 1 - Questions I

Labs are made of open questions. These questions are only guides. For each question, you are expected to perform some data analysis or numerical experiment and show the results using plots. Each analysis/experiment must be clearly explained, and all results must be commented.

1. **Dataset.** Perform general statistical analyses on the datasets. Are the results expected ?
Perform some sanity checks on the data. Is the dataset free of visible problems ?
2. **Prices and tick sizes.** Plot the best bid price, best ask price, mid price and weighted mid-price on a few minutes of your sample. Comment. Measure the tick sizes on your datasets. Are the measured values of the tick sizes in agreement with the official definition ?
3. **Trades.** Are the empirical distributions of the trade durations and trade sizes on your datasets in agreement with known stylized facts ?

Lab 1 - Questions II

4. **Spread.** Are you dealing with large-tick stocks or small-tick stocks ? Compute the empirical distribution of the spread *in event time* (each observed value counts for one observation) and *in calendar time* (each observed value is weighted by the length of the time interval on which it is observed). What do you observe ?
5. **High-frequency log-returns** Analyze the log-returns in your dataset for various sampling frequencies in calendar time. Is the Gaussian assumption satisfied ? Is it better in event time ?
6. **Imbalance** Is the imbalance a good predictor of mid-price moves or trade signs ?
7. **Autocorrelations.** Sign the trades and analyze the autocorrelation function of this timeseries. Sign the LOB events depending the sign of their price pressure and analyze the autocorrelation function of this timeseries. Sign the LOB events depending the impacted side of the limit order book and analyze the autocorrelation function of this timeseries.

References I

- [1] F. Abergel, M. Anane, A. Chakraborti, A. Jedidi, and I. Muni Toke. *Limit order books*. Cambridge University Press, 2016.
- [2] J.-P. Bouchaud, Y. Gefen, M. Potters, and M. Wyart. “Fluctuations and response in financial markets: the subtle nature of ‘random’ price changes”. In: *Quantitative finance* 4.2 (2004), pp. 176–190.
- [3] J.-P. Bouchaud, M. Mézard, and M. Potters. “Statistical properties of stock order books: empirical results and models”. In: *Quantitative Finance* 2.4 (2002), pp. 251–256.
- [4] A. Chakraborti, I. Muni Toke, M. Patriarca, and F. Abergel. “Econophysics Review: I. Empirical facts”. In: *Quantitative Finance* 11.7 (2011), pp. 991–1012.
- [5] D. Challet and R. Stinchcombe. “Analyzing and modeling 1+ 1d markets”. In: *Physica A: Statistical Mechanics and its Applications* 300.1-2 (2001), pp. 285–299.

References II

- [6] Z. Eisler, J.-P. Bouchaud, and J. Kockelkoren. "The price impact of order book events: market orders, limit orders and cancellations". In: *Quantitative Finance* 12.9 (2012), pp. 1395–1419.
- [7] C. M. C. Lee and M. J. Ready. "Inferring Trade Direction from Intraday Data". In: *The Journal of Finance* 46.2 (1991), pp. 733–746.
- [8] I. Muni Toke. "Reconstruction of Order Flows using Aggregated Data". In: *Market Microstructure and Liquidity* 02.02 (2016), p. 1650007.
- [9] I. Muni Toke and N. Yoshida. "Analyzing order flows in limit order books with ratios of Cox-type intensities". In: *Quantitative Finance* 20.1 (2020), pp. 81–98.
- [10] I. Muni Toke and N. Yoshida. "Modelling intensities of order flows in a limit order book". In: *Quantitative Finance* 17 (5 2017), pp. 683–701.

References III

- [11] A. C. Silva and V. M. Yakovenko. "Stochastic volatility of financial markets as the fluctuating rate of trading: An empirical study". In: *Physica A: Statistical Mechanics and its Applications* 382.1 (2007), pp. 278–285.
- [12] S. Stoikov. "The micro-price: a high-frequency estimator of future prices". In: *Quantitative Finance* 18.12 (2018), pp. 1959–1966.