Reconstruction of an Order Book using High-Frequency Data

Ioane Muni Toke

Ecole Centrale Paris
Applied Maths Department
Chair of Quantitative Finance

Market Microstructure: Confronting Many Viewpoints
Paris, December 12th 2010

Description of common high-frequency data

Two streams: "Trades" and "Quotes"

Date	Time stamp	Price	Quantity	Flag
2010-03-09	36000.309	56.15	20	
2010-03-09	36001.143	56.17	80	
2010-03-09	36001.143	56.17	420	
2010-03-09	36001.436	56.14	117	
2010-03-09	36001.651	56.14	150	
2010-03-09	36001.651	56.14	231	
2010-03-09	36001.651	56.14	1000	
2010-03-09	36003.523	56.14	150	
2010-03-09	36003.523	56.14	150	
2010-03-09	36003.698	56.14	150	
2010-03-09	36003.698	56.14	189	
2010-03-09	36003.698	56.14	231	
2010-03-09	36003.75	56.14	102	

Date	Time stamp	Side	Level	Price	Quantity
2010-03-09	36000.335	B	3	56.12	100
2010-03-09	36000.335	B	3	56.12	200
2010-03-09	36000.335	A	2	56.19	700
2010-03-09	36000.335	B	3	56.12	100
2010-03-09	36000.335	A	2	56.19	600
2010-03-09	36000.335	A	2	56.19	400
2010-03-09	36000.335	A	2	56.19	500
2010-03-09	36000.349	A	2	56.19	400
2010-03-09	36000.352	B	3	56.12	200
2010-03-09	36000.454	B	2	56.13	133
2010-03-09	36001.034	B	2	56.13	425
2010-03-09	36001.149	A	1	56.17	1361
2010-03-09	36001.149	A	1	56.17	941

- Pending orders aggregated per level
- Limited number of levels
- No broker id, no trade id, no order type

Simple trades and quotes matching

Date	Time stamp	Price	Quantity
2010-03-09	36000.309	56.15	20
2010-03-09	36001.143	56.17	80
2010-03-09	36001.143	56.17	420
2010-03-09	36001.436	56.14	117
2010-03-09	36001.651	56.14	150
2010-03-09	36001.651	56.14	231
2010-03-09	36001.651	56.14	1000
2010-03-09	36003.523	56.14	150
2010-03-09	36003.523	56.14	150
2010-03-09	36003.698	56.14	150
2010-03-09	36003.698	56.14	189
2010-03-09	36003.698	56.14	231
2010-03-09	36003.750	56.14	102

Date	Time stamp	Side	Level	Price	Quantity
2010-03-09	36000.335	B	3	56.12	100
2010-03-09	36000.335	B	3	56.12	200
2010-03-09	36000.335	A	2	56.19	700
2010-03-09	36000.335	B	3	56.12	100
2010-03-09	36000.335	A	2	56.19	600
2010-03-09	36000.335	A	2	56.19	400
2010-03-09	36000.335	A	2	56.19	500
2010-03-09	36000.349	A	2	56.19	400
2010-03-09	36000.352	B	3	56.12	200
2010-03-09	36000.454	B	2	56.13	133
2010-03-09	36001.034	B	2	56.13	425
2010-03-09	36001.149	A	1	56.17	1361
2010-03-09	36001.149	A	1	56.17	941

- Trade (56.17, 420) at time 36001.143 is perfectly match
- Time lag : Trade reported 6 milliseconds before

Matching trades that change the best limit

Date	Time stamp	Price	Quantity
2010-03-09	36000.309	56.15	20
2010-03-09	36001.143	56.17	80
2010-03-09	36001.143	56.17	420
2010-03-09	36001.436	56.14	117
2010-03-09	36001.651	56.14	150
2010-03-09	36001.651	56.14	231
2010-03-09	36001.651	56.14	1000
2010-03-09	36003.523	56.14	150
2010-03-09	36003.523	56.14	150
2010-03-09	36003.698	56.14	150
2010-03-09	36003.698	56.14	189
2010-03-09	36003.698	56.14	231
2010-03-09	36003.750	56.14	102

Date	Time stamp	Side	Level	Price	Quantity
2010-03-09	36003.519	A	1	56.14	300
...					
2010-03-09	36003.519	B	3	56.1	1356
2010-03-09	36003.519	A	1	56.17	1441
2010-03-09	36003.519	A	2	56.19	249
2010-03-09	36003.519	B	3	56.1	1056
2010-03-09	36003.519	A	3	56.2	1182
2010-03-09	36003.519	A	4	56.21	518
2010-03-09	36003.519	A	5	56.22	651
2010-03-09	36003.519	A	6	56.23	536
2010-03-09	36003.519	A	7	56.24	552
2010-03-09	36003.519	A	8	56.25	602
2010-03-09	36003.519	A	9	56.26	454
2010-03-09	36003.519	A	10	56.27	344
2010-03-09	36003.528	A	1	56.17	1241
2010-03-09	36003.528	A	1	56.14	300
2010-03-09	36003.528	A	2	56.17	1241

- Trades (56.14, 150) at time 36006.523 (twice) are perfectly matched
- Time lag : Trades reported 4 milliseconds later
- Next trade is not on the same side...

Rebuilding the history of treated orders

- ► An algorithm to match trades and quotes:

Parameter ΔT as a "minimum confidence lag"

Idea If time lag between 2 events is smaller than ΔT, Then they can be switched

Principle Reorder the events if the new combination is "better"

Criteria Maximize the number of perfectly matching trades
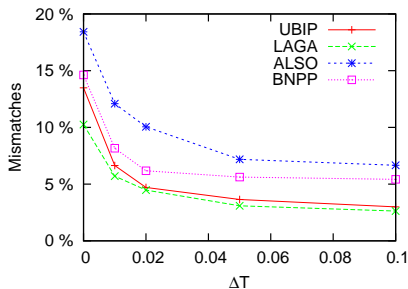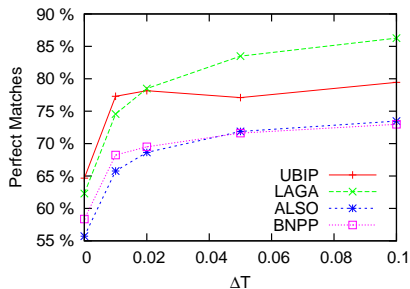Maximize the overall quality of the match
. . .

Output 1 Matched trades are market orders

Output 2 Once trades are matched, remaining quotes variations are limit orders and cancellation orders

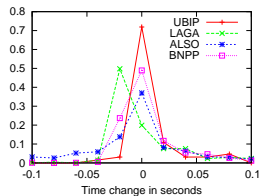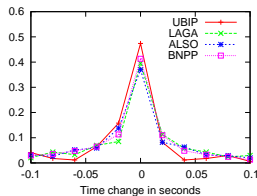- ► Remark : multiple run

Performance of trade matching

Analysis of four Euronext Paris stocks (Ubisoft, Lagardère, Alstom, BNP Paribas) on March, 9th 2010, 10am - 4pm

- ▶ One run
- ▶ Liquidity issue (but obviously not only)

Changes of timestamps

After the reordering of events, time stamps are changed to
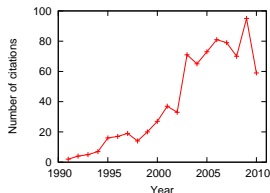retrieve a chronological order

		UBIP	LAGA	ALSO	BNPP
All Events	nEvents	19591	49359	86473	146926
	Changes	0.87%	1.36%	1.47%	2.98%
	Average	-0.010	0.003	-0.002	-0.005
Trades	nEvents	467	1297	2507	9056
	Changes	13.70%	12.61%	16.27%	10.50%
	Average	0.018	0.028	0.028	0.020

▶ Resynchronization with minimum change
▶ Delay to update quotes ?

Consequences on trade signature

- Lee and Ready algorithm (1991) :
    1. comparison to the mid-quote (which one ?)
    2. if non-conclusive, comparison to the previous trade

- Standard in financial literature

- Lee-Ready errors compared to perfect matches :

UBIP	LAGA	ALSO	BNPP
11.94%	7.04%	11.24%	5.49%

Issues of order book reconstruction

- If timestamps were perfect and no quote omitted:
  - Perfect identification of market orders
  - Perfect identification of executed limit orders (FIFO rule)
- Even if timestamps were perfect:
  - Non-unique identification of cancellations

Date	Time stamp	Side	Level	Price	Quantity
2010-03-09	36000.335	B	3	56.12	100
2010-03-09	36000.335	B	3	56.12	200
2010-03-09	36000.335	A	2	56.19	700
2010-03-09	36000.335	B	3	56.12	100
2010-03-09	36000.335	A	2	56.19	600
2010-03-09	36000.335	A	2	56.19	400
2010-03-09	36000.335	A	2	56.19	500
2010-03-09	36000.349	A	2	56.19	400
2010-03-09	36000.352	B	3	56.12	200

  - Clustering of orders' volumes
- When modeling :
  - Point processes modeling (no simultaneous events)
  - Problem usually not treated (Random shuffling, others ?)