bConnected
powered by Google

Kevin Ramlal <kevinramlal@berkeley.edu>

Quoting Behavior and Information Leakage Project

Benjamin Connault <benjamin.connault@iextrading.com> Tue, Jun 11, 2019 at 12:38 PM
To: Thomas Moore <thomas_moore@mfe.berkeley.edu>
Cc: Kevin Ramlal <kevinramlal@berkeley.edu>, Weipeng Shao <weipeng_shao@berkeley.edu>, Junjie Rao <raojunjie2014@berkeley.edu>

Hi all,

Thank you for this detailed update. Great to hear you've already made good progress. If you'd like, I am happy to jump on a call and chat a bit about the project, maybe answer potential questions you may have. No pressure though, up to you.

TAQ is definitely the right data for this project. I know the data is not easy to work with and I read through your jupyter notebook – great job unpacking the data already. Here are a couple of additional pointers:

- The SIP distributes (and TAQ redistributes) 3 tables: "trades", "quotes", "nbbo".
- The "nbbo" table name is misleading. It cannot be used by itself and it's really a complement of the "quotes" table for the purpose of building the nbbo.
- There are two roads to building the nbbo:
  - Use "quotes" + "nnbo" and the NATBBO_IND column.
  - Just using "quotes" and the quote condition logic (QU_COND).
    - This is the road that we take at IEX internally (we don't look at the SIP/TAQ "nnbo" table at all).
    - It allows you to create a richer version of a custom nbbo table, including for example the set of exchanges on the nbb / nbo (probably useful for this project). See below for a screenshot of our internal nbbo table.

date	sym	time	bexs	bsizes	bsize	bid	ask	asize	asizes	aexs
2018.01.04	AAPL	09:30:00.114854924	enlist "T"	enlist 51i	51	1726000	1727100	1	enlist 1i	enlist "T"
2018.01.04	AAPL	09:30:00.179964944	enlist "P"	enlist 1i	1	1726300	1727100	1	enlist 1i	enlist "T"
2018.01.04	AAPL	09:30:00.192815687	enlist "P"	enlist 1i	1	1726300	1727100	2	enlist 2i	enlist "T"
2018.01.04	AAPL	09:30:00.216050853	enlist "X"	enlist 1i	1	1726400	1727100	2	enlist 2i	enlist "T"
2018.01.04	AAPL	09:30:00.242303475	enlist "P"	enlist 1i	1	1726300	1727100	2	enlist 2i	enlist "T"
2018.01.04	AAPL	09:30:00.258104228	enlist "P"	enlist 1i	1	1726300	1727000	1	enlist 1i	enlist "Z"
2018.01.04	AAPL	09:30:00.265703274	enlist "P"	enlist 1i	1	1726300	1727100	2	enlist 2i	enlist "T"
2018.01.04	AAPL	09:30:00.419196064	enlist "P"	enlist 1i	1	1726300	1726700	1	enlist 1i	enlist "X"
2018.01.04	AAPL	09:30:00.419537911	enlist "P"	enlist 1i	1	1726300	1727100	2	enlist 2i	enlist "T"
2018.01.04	AAPL	09:30:00.426891904	"TP"	5 1i	6	1726300	1727100	2	enlist 2i	enlist "T"
2018.01.04	AAPL	09:30:00.428479371	"TP"	6 1i	7	1726300	1727100	2	enlist 2i	enlist "T"
2018.01.04	AAPL	09:30:00.430316889	"TP"	7 1i	8	1726300	1727100	2	enlist 2i	enlist "T"
2018.01.04	AAPL	09:30:00.492765733	"TP"	7 1i	8	1726300	1727100	3	enlist 3i	enlist "T"
2018.01.04	AAPL	09:30:00.505282524	enlist "P"	enlist 1i	1	1727100	1727100	3	enlist 3i	enlist "T"
2018.01.04	AAPL	09:30:00.505380370	"TP"	7 1i	8	1726300	1727100	3	enlist 3i	enlist "T"
2018.01.04	AAPL	09:30:00.522023025	"TP"	7 1i	8	1726300	1727100	2	enlist 2i	enlist "T"
2018.01.04	AAPL	09:30:00.522087340	"TP"	7 1i	8	1726300	1727100	1	enlist 1i	enlist "T"
2018.01.04	AAPL	09:30:00.522090449	"TP"	7 1i	8	1726300	1728000	17	enlist 17i	enlist "T"
2018.01.04	AAPL	09:30:00.524159767	"TP"	7 1i	8	1726300	1727100	1	enlist 1i	enlist "T"
2018.01.04	AAPL	09:30:00.544454718	"TP"	7 1i	8	1726300	1728000	17	enlist 17i	enlist "T"
2018.01.04	AAPL	09:30:00.545599610	"TP"	7 1i	8	1726300	1728000	18	17 1i	"TZ"
2018.01.04	AAPL	09:30:00.545671400	"TPZ"	7 1 1i	9	1726300	1728000	18	17 1i	"TZ"
2018.01.04	AAPL	09:30:00.545720327	"TPKZ"	7 1 1 1i	10	1726300	1728000	18	17 1i	"TZ"
2018.01.04	AAPL	09:30:00.546314235	"TPKZ"	7 2 1 1i	11	1726300	1728000	18	17 1i	"TZ"
2018.01.04	AAPL	09:30:00.551261172	enlist "T"	enlist 12i	12	1727500	1728000	18	17 1i	"TZ"
2018.01.04	AAPL	09:30:00.571368371	enlist "T"	enlist 11i	11	1727500	1728000	18	17 1i	"TZ"
2018.01.04	AAPL	09:30:00.571402510	enlist "T"	enlist 10i	10	1727500	1728000	18	17 1i	"TZ"
2018.01.04	AAPL	09:30:00.571465939	enlist "T"	enlist 5i	5	1727500	1728000	18	17 1i	"TZ"
2018.01.04	AAPL	09:30:00.571466174	enlist "T"	enlist 4i	4	1727500	1728000	18	17 1i	"TZ"

In terms of quote conditions, the quote that can contribute to the NBBO are the following: O, R for all symbols + Y for tape C symbols.

Once you have nbbo-building logic, you can send me an example and I can validate it against our internal version.

About time aggregation:

- You have access to (non-rounded) nanosecond timestamp so there is 1 approach to the project where you don't do any aggregation.
  - This may allow you down-the-line to distinguish and compare two cases such as:
    - Nasdaq creates a bid and is "quickly" followed by BATS.
    - BATS creates a bid and is "quickly" followed by Nasdaq.
    - If you aggregate you will see that BATS and NASDAQ created bids.
- Aggregation is still a reasonable way to approach the project. I just wanted to point out the above for you to consider – please proceed as you prefer.
- If you do aggregate, I would suggest:
  - Write code so that you easily play with the aggregation window (1 sec, 100 ms, etc.) for potential robustness checks
  - Take some time to look at the distribution of # of quote creations by time window.

Please feel free to reach out with any question!

Ben

[Quoted text hidden]