TICK-LEVEL ANALYSIS

.

1. The Structure of Tick Data

1.1. **Terminology.** The basic important elements of market state change are trades, bids and offers. We use the term *tick data* to denote high-resolution observations of these numbers, and usually assume it is available as of any given time in which the market is open. The trades are collectively considered to comprise the *trade tape* while the bids and offers collectively form the *order book* as of any given moment.

If the market is an exchange, we can consider one or more orders to be at the *top of book*. For typical exchanges this is the set of highest-price bids and lowest-priced offers that have not yet *matched* or been *cancelled*, and have not *traded out*. The top of book is normally also ranked with highest priority to the oldest orders, in which case the exchange is said to use classic *price-time priority*. Incoming new orders are said to comprise the *order flow*.

For a given asset class, the group of exchanges serving it, the rules they use, and the firms accessing them collectively form the *market microstructure*.

2. Interesting Alternative Market Structures

The U.S. equity markets are a*fractured market*, in the sense that the same equities are traded on multiple exchanges. They are now legally linked by *Reg NMS* (for *national market structure*) which specifies rules under which they must send orders to each other when prices are better-looking for the customer.

Foreign exchange *electronic crossing networks (ECN)* allow brokers *last look rights*. Futures exchanges may use a combination of price-time priority and *proportional allocation* in which all orders at the top-of-book get a nonzero fill probability.

Some exchanges allow brokers to give higher priority to *internal crosses* that join the broker's own customers. The U.S. equity markets now have similar structure, in effect, as orders arising from *retail* order flow are internally crossed by Citadel and other firms who make it their business to purchase the flow, cross where sensible, and then send the *exhaust* as new orders to the usual markets.

Certain futures markets are traded *pro rata*, in which there is no time priority given, or time priority is given only to the very earliest order..

Not all exchanges make the orders visible to participants. These *dark pools* may also have minimum size rules for matching.

Eric Budish suggests discretizing not only price, in order price ticks, but also times, by altering the matching process so as to hold periodic auctions.

3. Tick Level Statistics

3.1. **Basic Book Statistics.** The most important tick-level statistic is the *mid price*, defined as

$$\text{mid} = \frac{1}{2}(\text{bid} + \text{offer})$$

Next most important is the *VWAP* or *volume-weighted average price* defined for trades of sizes V_i at prices P_i to be

$$\text{VWAP} = \frac{1}{N}\sum_{1}^{N} P_i V_i$$

The VWAP is a minimal representation of tradable position changes.

Somewhat less useful is a *TWAP* or *time-weighted average price*

More interesting is an *order-weighted average price*, formed from the observation that if we have many more offers than bids, the *clearing price* at which participants collectively would be willing to trade were tick sizes smaller is closer to the bid than the mid price. This is not an exact science, but given a top of book with N_B bids at price B and N_A offers at price A one reasonable formula weights with square roots

$$w = \sqrt{N_A} + \sqrt{N_B}$$

$$\text{OWA} = \frac{\sqrt{N_A}}{w}B + \frac{\sqrt{N_B}}{w}A$$

Clearly, other weighting schemes, possibly involving further levels of the book, can be incorporated into alternative versions of the OWA.

Arguably, the OWA is a model, since we can use it to predict whether the next trade will occur on the bid side or the offer side of the book.

3.2. **Trade Marking.** A very common derived statistic is *marked* trades, where bids and offers are marked according to whether a buyer initiated the transaction, or the seller initiated it. This process is flattered as the *Lee-Ready Algorithm* and essentially consists of the following for each trade

- Find the best bid and offer as of the trade time
- Compare the trade price to them
- If the trade price equals the best bid, mark as seller-initiated
- If the trade price equals the best offer, mark as buyer-initiated
- (Optional) Otherwise, mark according to whether trade price was below or above mid price

As stated, the algorithm may be too simple. For example block trades in equity markets and spread trades in futures markets may not be appropriate for marking.

Easley *et alia*'s *VPIN* or *volume-weighted probability of informed trading* for a time t compares buyer-initiated versus seller-initiated trades in N windows preceding t. The idea is that if many trades are crossing the spread in the same direction, then the market is likely to have more informed directional buyers than *noise traders* who lack particular information.

We begin by defining *order imbalance* in the nth window by taking V^B to be the volume of buyer-initiated trades and V^S to be the volume of seller-initiated trades, and setting the imbalance as

$$I_n = \frac{|V_n^B - V_n^S|}{V_n^B + V_n^S}$$

We can now set VPIN to be the mean

$$\text{VPIN} = \frac{1}{N} \sum_{n=1}^{N} I_n$$

3.3. **Returns.** We commonly think of the interval T return of an asset as of time t in either absolute terms, as

$$r_t^{(T)} = \frac{P_t}{P_{t-T}} - 1$$

or in logarithmic terms

$$r_t^{(T)} = \log\left(\frac{P_t}{P_{t-T}}\right)$$

but we now see that the idea of knowing a price P_t at some time t is somewhat optimistic.

Let's say we were to define P_t as the mid price. Naturally this stays constant over some time intervals, implying that returns are constant decreasing in absolute value over this time. In addition, the return in any period becomes fairly sensitive to exactly which ticks the period contains.

We cannot entirely overcome the ambiguity, but we can certainly alleviate the tick-sensitivity problem by using an EWMA. We define

$$r_t^{(T)} = \log\left(\frac{P_t}{\text{EWMA}\left(P, \lambda = \frac{1}{T}\right)}\right)$$

which can conveniently be computed without saving past ticks.

4. The Epps Effect

A common issue with tick level data is the long periods of boredom interrupted by short intervals of terror. That is, data updates come at certain times (often bunched) so at a sufficiently small timescale correlations appear to drop to zero.