Order flow imbalance

Enea Shaqiri

Let p_n^A, p_n^B be the $n-$th ask and bid prices, and q_n^A, q_n^B be the $n-$th ask and bid volumes, for $n = 1, 2 \ldots$. We will keep track of two times: the physical time (measured by clocks), and the event time which increases when there is a new bbo event (change in price or in volume). We denote physical time by t and the event time by n. We introduce the random variable τ_n, which denotes the physical time of the $n-$th event. We define

$$e_n := \chi_{\{p_n^B \geq p_{n-1}^B\}} q_n^B - \chi_{\{p_n^B \leq p_{n-1}^B\}} q_{n-1}^B - \chi_{\{p_n^A \leq p_{n-1}^A\}} q_n^A + \chi_{\{p_n^A \geq p_{n-1}^A\}} q_{n-1}^A,$$

where χ is the indicator function.

The variable can be positive or negative, with positive hinting at a positive price movement and negative hinting at the opposite. For instance, if the $n-$th event is that the ask price moves up, then $e_n = q_{n-1}^A$, which is the volume at the best ask before the price moved (that is the available supply that left the market either because it was taken or cancelled).

We assume that the data has been recorded in the interval $[0, t_{end}]$, and we fix two equally spaced grids on this interval, $\{T_i\}_{i \in \mathbb{N}}$, $\{t_k\}_{k \in \mathbb{N}}$, with the second grid finer than the first one. We define ofi_k on the intervals $[t_{k-1}, t_k]$. To do so, we first introduce the function $N(t) := \max\{n \in \mathbb{N} | \tau_n \leq t\}$, which maps physical time to event time. With this notation we set

$$\text{ofi}_k := \sum_{n=N(t_{k-1})+1}^{N(t_k)} e_n.$$

ofi_k is a measure of imbalance between supply and demand in the interval $[t_{k-1}, t_k]$. A negative (positive) imbalance can manifest both as an increase in supply (demand) and as an increase (decrease) in price.

We also consider the mid price changes

$$\Delta p_k = \frac{p_k - p_{k-1}}{\delta},$$

with p_k price at time t_k and δ tick size. We now claim that the relationship between the price changes and ofi to be linear. To formulate that precisely we recall that every interval of the form $[t_{k-1}, t_k]$ is contained only in one interval of the form $[T_{i-1}, T_i]$. To make this dependence explicit, in the next formula we use $k(i)$ to denote that the index k is related to the interval i.

The relationship we want to study statistically is

$$\Delta p_{k(i)} = \alpha_i + \beta_i \text{ofi}_{k(i)} + \varepsilon_{k(i)},$$

where the regression coefficients are assumed to be constant on the interval

$[T_{i-1}, T_i]$, and ε_k is the error in the linear approximation coming from ignoring the effect of deeper orderbook levels. The β coefficient, called price impact coefficients, determine how sensitive the price change is to the order flow imbalance. Intuitively, the more volume there is in the book, the less sensitive the price is to reductions in volume.

Data

todo

Results

todo