Modeling executions in a dark pool

Vladimir Markov

Bloomberg

vmarkov2@bloomberg.net

Stevens Institute of Technology

10.29.2015
Talk outline

- Overview of the exchange/dark pool landscape.
- Trade size distribution.
- Zero-inflated Poisson process.
- Hawkes process: clustering of liquidity.
- Effect of the limit prices.
- Dependence between indication and trade intensities.
- Liquidity seeking algorithms.

This talk is partially based on:
VM and Tito Ingargiola, ”Block-Crossing Networks and the Value of Natural Liquidity. The Journal of Trading, Summer 2013
- Fragmented market: 11 exchanges and 30+ regulated dark pools (some of them are not active).
- Exchanges are self-regulated profit centers.
- Most of the profit comes from services such as data feeds, collocation and etc. Trading commission is only a small fraction of the total profit.
- Driving evolution force is the competition between market innovations and lagged regulation.
- Characterized by asymmetric access to information. HFTs invest into infrastructure, data feeds, market micro-structure and quant research in order to get the advantage.
- Matching mechanisms are price-time priority and continuous auction. Viable alternatives are price-size priority; discrete auctions or matching with fixed or random delay (\(\sim 100\) ms) to minimize latency arbitrage.
Dark pools refer to a heterogeneous set of venues where information about an order is not openly available to the public.

Dark liquidity constitutes 14% of the total trading volume in the US.

Institutional investors using dark pools can trade large blocks of stocks without showing their hand to others. It allows to avoid market impact as neither the size of the trade nor the identity are revealed until the trade is filled.

The most informative characteristics of a dark pools is the average trade size. Average trades size varies between 200 and 40K shares.

Independent dark pools (BPoll,Liquidnet,ITG Posit), broker-dealer-owned dark pools (Credite Suisse’s CrossFinder), exchange-owned dark pools, consortium-backed operators (Chi-X).

Dark pools are of various types and can execute trades in multiple ways, such as through negotiation or automatically (NBBO mid-point crosses, interval VWAP, and etc.), through the day or at scheduled times.
Dark pools II

- **Benefits:** price improvement (mid-point executions), no market impact.

- **Risk:** Adverse selection via gaming or information leakage. For example, information leakage via small executions (pinging), winner’s curse (complete fill of an order implies that contra had more liquidity behind his order), midpoint price manipulation.

- **Risk:** Trades are executed at SIP mid-point NBBO which is behind direct feeds used by HFTs. Latency arbitrage.

- **Dark pools/orders vary by the degree of “darkness.”** Some do not display any information about an order, some advertise liquidity via indication of interests (IOI) (info on direction of the trade, sometimes size)

- **IOIs increase crossing rates but leak information. All depends on urgency of the order set by a trader.**
Trade and Order Size Distributions

- The average trade size $T$ is given by the expected value of the minimum of two random order sizes $O_1$ and $O_2$ 
$$T = E[min(O_1, O_2)].$$
- The exponential distribution provide loose fit for the main part of the order size distribution in the US.
- For exponential random variables $O_1, O_2, O \in P_{\text{exp}}(\lambda)$ the ratio
$$\frac{E[min(O_1, O_2)]}{E[O]} = \frac{1}{2}$$

- The average trade size is two times smaller than average order size and doesn’t depend on the intensity $\lambda$. The same ratio for median is 0.5 and does not depend on intensity $\lambda$ as well.
- It explains downward pressure on sizes, as a party with larger size gets a partial execution and is exposed to information leakage and adverse selection.
Mechanics of the crosses

- In order to have a cross the candidate orders have to be active on opposite sides of the same symbol at intersecting times and the price should be marketable for both parties.

- Assuming factorization of the cross probability $P_m$, we have the following factors:

$$P_m = P_{side} \times P_{symbol}(T) \times P_{limit}$$

(2)

where $P_{side}$ controls the sign distribution of the orders, $P_{symbol}(T)$ is the probability to get the cross if an order is exposed for a time $T$, and $P_{limit}$ takes into account limit prices.

- The probability $P_{symbol}(T)$ is proportional to the idiosyncratic historical intensity $\lambda$ of trades and Poisson or Hawkes models can be used for the modeling.

- $P_{symbol}(T)$ grows with time, i.e. the longer the exposure time the higher probability to get the cross.
Zero Inflated Poisson Model: Sparsity of Liquidity

- Dark (block) liquidity is often sparse (most of the time there is no volume). Zero inflated Poisson (ZIP) model allows to statically model the expected number of dark trades per unit time $x_i$:

\[
Pr(x_i = 0) = \pi + (1 - \pi) \exp(-\lambda)
\]  

\[
Pr(x_i = n) = (1 - \pi) \frac{\lambda^n \exp(-\lambda)}{n!}
\]

where $\lambda$ is the expected Poisson intensity and $\pi$ is the probability of zeros.

- $\lambda$ and $\pi$ can be estimated via maximum likelihood method by solving two equations:

\[
\bar{x}(1 - e^{-\lambda_{ML}}) = \lambda_{ML} \left(1 - \frac{n_0}{n}\right)
\]  

\[
\pi_{ML} = 1 - \frac{\bar{x}}{\lambda_{ML}}
\]

here, $\frac{n_0}{n}$ is the observed proportion of zeros and $\bar{x}$ is the sample mean.
A conventional trader’s wisdom says that liquidity begets liquidity.

It can be formalized via the Hawkes process (Hawkes [1974])) which is a generalization of the Poisson process that has a memory kernel $h(t)$ which takes into account the influence of past events at time $t_i$ on the current conditional intensity:

$$\mu_H(t) = \mu_0^H(t) + \sum_{t_i=0}^{t} h(t - t_i), \quad (7)$$

here $\mu_0^H(t)$ is a base intensity that determines the rate of arrival of idiosyncratic, history independent, exogenous first order events (immigrants) which trigger clusters of endogenous descendants (offsprings) with the rate controlled by the memory kernel $h(t) = \alpha \exp(-\beta t)$.

Parameter $\alpha$ is called immigration intensity and inverse $\beta$ defines the timescale of the clustering.
The positive feedback level can be measured by the branching ratio $n$ which is equal to the average number of descendants per one immigrant

$$n = \int_0^\infty dt \, h(t) = \frac{\alpha}{\beta}. \quad (8)$$

The close to zero branching ratio shows the absence of clustering and the branching ratio close to one provides an example with strong positive liquidity feedback with a significant number of trades arriving in clusters.

For $< n > = 0.5$, 50% percent of all trades happen via clustering mechanism.
A buy order at price $P_0$ becomes unmarketable if the price is higher than the limit price $P_0 + x_0$.

Assuming that the price follows Brownian motion with diffusion constant $D$ (volatility at time scale $T$ is $\sigma^2_T = 2DT$), the first passage probability density $F(x_0, t)$ is given by:

$$F(x_0, t) = \frac{x_0}{\sqrt{4\pi Dt}} e^{-x_0^2/(4Dt)}$$  \hspace{0.5cm} (9)

We apply the reflection principle that says that the price will be half of the remaining time above the limit price. The average time the order is not marketable $< T_{nm} >$ is given by:

$$< T_{nm} > = \int_0^T dt \ F(x_0, t) \times (T - t)/2 =$$

$$= \frac{T}{2} \left( (1 + y^2) \text{erfc}(y\sqrt{2}) - \sqrt{\frac{2}{\pi}} y e^{-y^2/2} \right), \quad y^2 = \frac{x_0^2}{2DT} = \frac{x_0^2}{\sigma^2_T}. \quad (11)$$
Effect of the limit prices II

- If an order is exposed for a day ($T = 1$), then the volatility is daily volatility $\sigma_D$ and $y$ is the limit price in daily volatility units.
- The ratio $< T_{nm} > / T$ is plotted below:

For example, with the limit $x_0 = 0.2\sigma_D$ (i.e $y = 0.2$) the order is not marketable 36 percent of the time during $T = 1$ Day.
There is no match if at least one order is not marketable and the match is not possible (we use the average values) for 
\[ < T_{nm} > = T_{nm} + T_{nm}. \]

Assuming \( y_B = y_S = 0.2 \), there is no possibility for a match during \( T_{nm} = 2 \times 0.36 = 0.72T \) and both orders will be marketable only 28 percent of the day and the ratio of the probability of the cross with limits to the probability of the cross without limits \( P_{\text{with limits}}/P_{\text{nolimits}} = 0.28. \)

This clearly shows that tight limits have detrimental effect on the probability of crossing.
Assume buy and sell unit orders arrive according to a Poisson distribution $P(X_B)$ and $P(X_S)$ with intensities $\mu_B$ and $\mu_S$. If an order is exposed to a dark pool for the full trading day, the inverse of the trading rate is the average number of orders per day.

The trading dynamics is controlled by $Z = \min(X_B, X_S)$.

In case of equal intensities of buy and sell orders $\mu = \mu_B = \mu_S$, the pmf of $Z$ looks particularly simple:

$$P(Z = k)_{k>0} = \frac{1}{2} \frac{(2\mu)^k}{k!} e^{-2\mu} - l_k(2\mu)e^{-2\mu}$$  \hspace{1cm} (12)

$$P(Z = 0) = \frac{e^{-2\mu}}{2} (1 - l_0(2\mu))$$  \hspace{1cm} (13)

the expected value of $Z$ is given by:

$$\mu_Z = E[Z] = \sum_{i=1}^{\infty} i \times P(Z = i) = \mu - \mu e^{-2\mu} (l_0(2\mu) + l_1(2\mu)),$$  \hspace{1cm} (14)

The expected value of the poissonian order rates $E[X_B] = E[X_S] = \mu$. 

Vladimir Markov (Bloomberg)
For small order intensities $\mu \ll 1$, the expected trading intensity $\mu_Z \approx \mu^2$.

In a sparse block trading landscape ($\mu \ll 1$) the rate of trading decreases very rapidly to zero. This may help explain why it is so difficult to start a continuous block trading business.

For order intensity rate $\mu = 1$ (one buy or sell indication per day) the trade intensity $\mu_Z = 0.48$ (one trade per 2 days), but for orders intensity rate $\mu = 1/3$ (one indication every 3 days) the trade intensity $\mu_Z = 0.08$ (one trade every 10 days).
Dark liquidity seeking algorithms expose clients’ orders to multiple dark pools.

Dark pools offer a wide range of liquidity. Retail and internalization pools offer frequent but small size (100-1000) crosses. Block venues crosses are less frequent but have larger sizes (10K-100K).

Many pools and institutional orders have high (500-10K) minimum execution size to prevent information leakage.

The practical implementation is comprised of heuristic rules which are necessary to incorporate “stylized” facts of the financial markets, handle edge cases and other departures of the implementation from the theoretical model, and maintain flexibility to satisfy client needs.

Many factors are subjective and not quantifiable. Always be aware of model assumptions and calibration details.
Liquidity seeking algos and multi-armed bandit problem

- Multi-armed bandit problem is a good starting point for building an liquidity seeking algos. Each arm corresponds to a dark pool.
- Bayesian approach (Thompson sampling) updates the probability distribution of each arm whenever additional information becomes available (dynamic heat-map).
- It is important to have both exploration and exploitation phases. Having only exploitation phase (chase past winners only) leads to a linear regret function (optimal solution grows logarithmically) which is not very different from a random choice.
- Exploration phase: use multiple allocations adjusted to a venue historical size pdf to update the heat-map of liquidity across venues, discover liquidity protected by large min quantity.
- Exploitation phase: use clustering property of the Hawkes process.
- Dynamic heat-map of liquidity uses Hawkes intensity plus additional dumping of intensity if posted volume do not get a cross.
- Use min execution quantity to control the rate of execution and max execution quantity to prevent getting blocks at a price spike.
In this talk, we covered a broad range of mathematical methods used in modeling dark pool executions with a focus on block trades. We quantified the basic mechanics of a cross, dependence between indications and trading rates, the effect of limit prices, and clustering of trades using the formalism of Hawkes processes. The problem of optimal allocation between darks pools was also reviewed.