NO. 8 | JUNE 2015

Liquidity Clustering in Dark Venues

Introduction

The US equities market is fragmented, with dozens of dark pools and exchanges offering hidden order types. Each operator finds something to love in its own pool, but skeptical observers have questioned how much unique liquidity, or unique value of any kind, the proverbial 31st dark pool really provides.

To begin to address this question, we analyzed correlations in midpoint liquidity across two dozen trading venues. If everyone effectively has access to all the dark pools—either directly or indirectly through trading intermediaries—there may be little benefit to accessing them all individually.

Contrary to the most skeptical view, our analysis shows that at the moderate time scale of a minute, venues matter. The strongest pattern shows that liquidity tends to cluster roughly into two groups, exchanges and dark pools, but there are additional idiosyncratic correlations between venues that suggest that having access to a broad range of dark pools provides practical value. Brokers who have not established a broad set of interconnections with other dark pools, or who favor their own dark pool to the exclusion of others' may be missing liquidity.

In a forthcoming research note, we will explore other

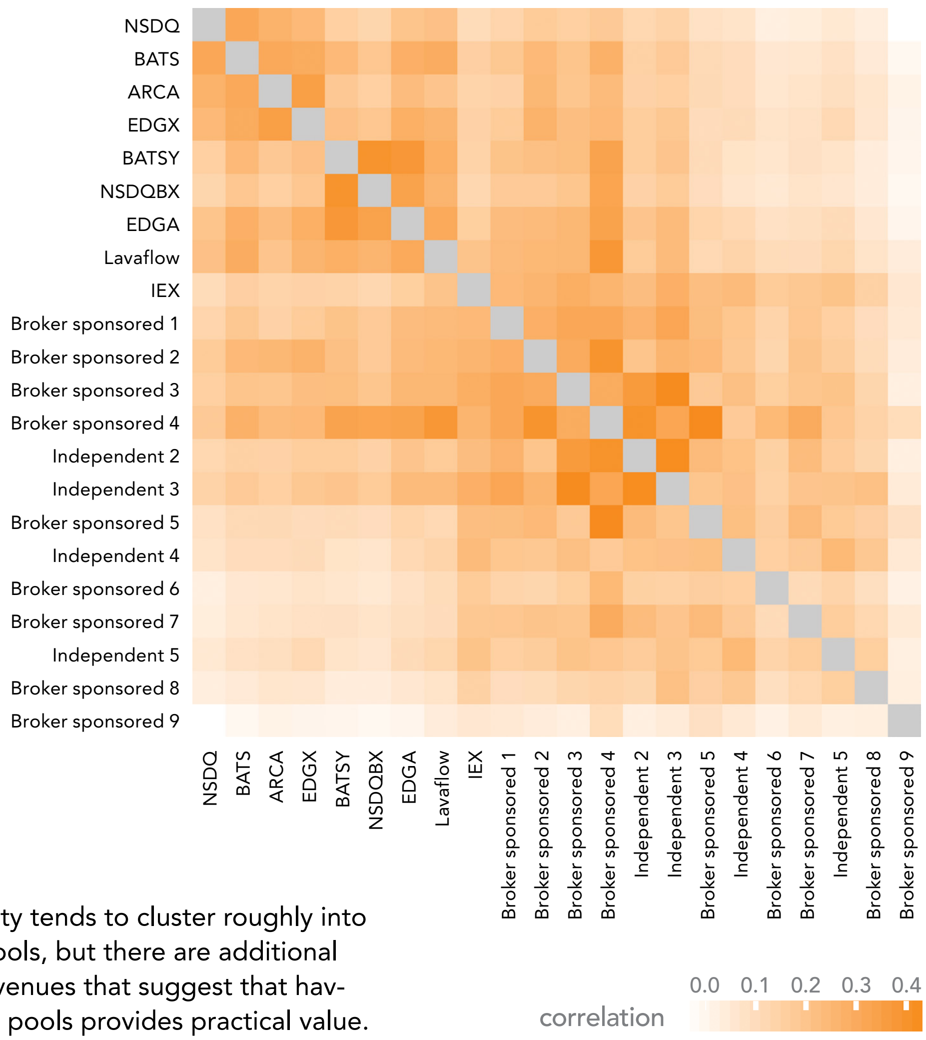

FIGURE 1

Liquidity Correlation

characteristics of dark pools around liquidity and execution quality, and see the extent to which venues differentiate themselves in these dimensions.

Data

We use a proprietary dataset of Pragma's dark algo orders from Q4 2014, which were routed to about 30 venues, a mixture of dark pools and exchanges. The data set includes approximately 71,000 parent orders with 1.3 million individual fills.

The particular behavior of an algorithm can introduce strong biases or artifacts into after-the-fact venue analysis.[1] By restricting our attention to mid-point orders sent from a single dark algo, and by looking at coincidence of fills only during periods when orders were present at pairs of venues, we minimize such artifacts. However, understanding something about the behavior of the algorithm is still valuable when interpreting the data we present here.

Pragma's dark algo takes a novel approach to deciding where and how much to send to each venue.[2] It maintains an empirical distribution of likelihood of fill for each order size for each venue using both historical and real-time data. At any given venue, the higher the minimum quantity specified for an order, the less likely the order will get filled and the lower the expected trading rate will be. The algo continually adjusts to use the highest minimum quantity such that the desired overall trade rate is achieved and thereby minimizes the footprint of the order in the market. It does this by minimizing the total number of distinct executions and by avoiding "pings" that reveal the existence of the order without providing adequate compensating liquidity. Then, the leaves of the parent order are continually redistributed across the available venues to achieve a desired trading rate with the highest possible minimum quantity using an approach similar to one described as "optimal" in Ganchev et al.[3]

This allocation algorithm naturally results in blanketing the venues in order to find liquidity when possible, and actively searches for liquidity when necessary.

1 Pragma Securities LLC, "Venue Analysis: What is it Good for?," 2014.

2 Pragma, "OnePipe 3.0: The Next Generation of Dark Liquidity Aggregator," 2011.

3 K. Ganchev et al., "Censored exploration and the dark pool problem," in *Proceedings of Uncertainty in Artificial Intelligence*, 2009.

Importantly for the data presented here, it does not treat venues differently based on factors outside of the fill probability distribution. Thus, no exogenous or arbitrary correlation structure is built into the order placement logic based on venue identity. We further exclude venues with low fill rates and therefore very few data points.

Methods

We take several approaches to analyzing the relationship among the venues in this note, first and simplest of which is computing the covariance or correlation of our fills at the venues.

After visual inspection, our first approach to analyzing this correlation matrix is principal component analysis (PCA). Our second approach is to look at clustering. The conventional k-means approach to cluster the venues does not work well in the present application because of missing data—there are many times when we do not have orders at one or more of the venues. Instead, we borrow an idea from graph partitioning called *community structure* as an alternative approach to grouping the venues.

We use a window size of 60 seconds to define simultaneity because our focus in this study is how liquidity flows among venues, not the high-frequency structure of order placement or execution. In addition, this time window allows us to better accommodate low-volume venues. With a short window, the correlation among venues is lower; with a longer window, the correlation is higher. However, we find that the correlation structure remains similar within different window lengths.

Results and Analysis

First, we compute the correlation matrix of liquidity for fills—that is, each cell shows the correlation of finding simultaneous liquidity at a particular pair of venues conditioned on having active orders on both venues. Note that correlations are not on fill quantities, but on a fills table. The fills table has entries of true for having a fill, false for not having a fill, and null for having no orders outstanding. We show the correlations matrix in Figure 1, with deeper orange indicating higher correlation.

Although some structure can be seen directly in the correlation matrix, the structure of the correlations is more clearly revealed by PCA. We show the first two principal components in Figure 2. The first indicates a

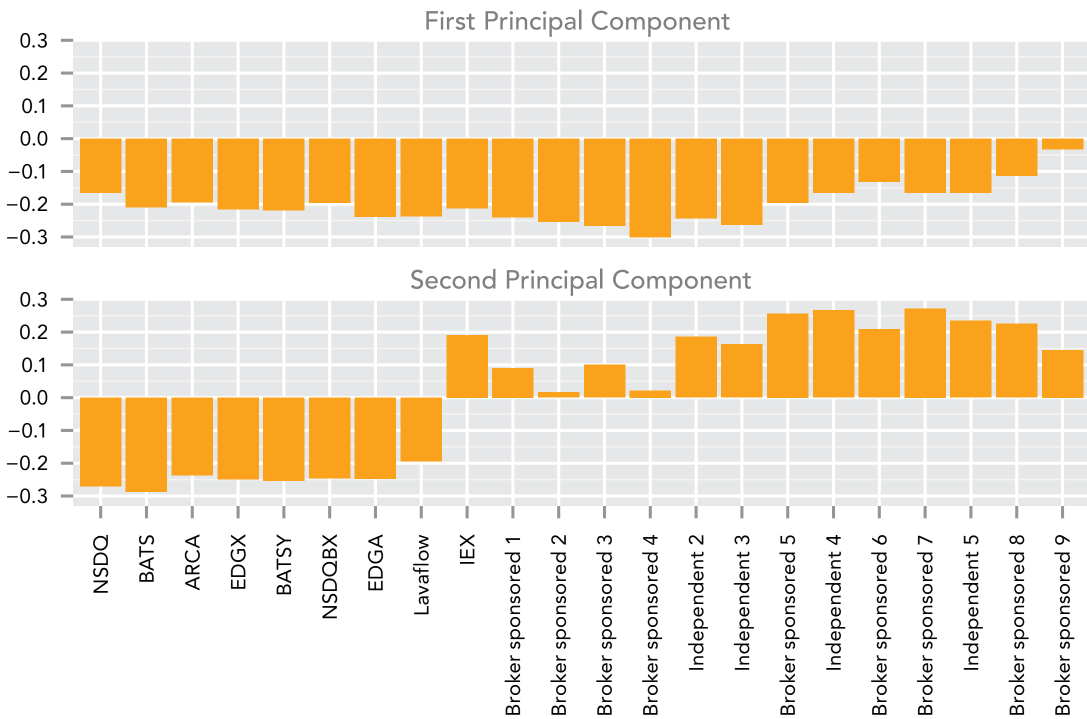

FIGURE 2

First 2 principal components of liquidity correlation matrix

market-wide component of being more likely to get fills everywhere given that we've been filled somewhere. We speculate that this indicates the effect of an individual aggressive trader on the other side, or an overall trade imbalance when on balance the market is aggressive on the other side.

The second principal component indicates a category specific component—liquidity tends to cluster within dark pools or within lit exchanges. That is, given a midpoint fill on an exchange, it's relatively likely that you will also see a fill on an exchange and unlikely you will see one at a dark pool, and vice-versa. We speculate that this pattern reflects the nature of the trading tool being used by an active counterparty in the marketplace with whom we are interacting at that moment. Although our dark aggregator places its hidden midpoint orders based only on fill probability, an aggressive counterparty may be more likely to route to multiple lit venues if they are using an SOR or a schedule based algo like a VWAP, and more likely to route to dark pools if they are also using another dark pool aggregator or liquidity sourcing algorithm. Therefore, our inference is that other algorithms may make a stronger distinction between lit and dark venues than our dark algo does.

To look explicitly at clustering, we borrow an idea from graph partitioning. The community structure approach partitions a network into exactly two groups by minimizing the cut size, i.e., the number of edges connecting the groups.[4] In graph theory we can express the strength of the connections between nodes, i.e. the venues in our application, via a weighted adjacency matrix. The correlation matrix provides a convenient proxy for the interconnection among venues. The community structure technique then labels each venue as belonging to one of two groups, in a way that minimizes the total connectivity among venues that are assigned to different groups. Of the 23 venues under consideration, 8 are exchanges and the rest dark pools, so we apply the community structure to split the venues 8 vs. 15. In the first group, we have the exchanges plus one large broker-sponsored dark pool, and in the second group we have Lavaflow, IEX, and the other dark pools. The similarity of the grouping reinforces the conclusions from the more traditional PCA approach described above.

Conclusion

The question behind this research note is whether, in today's market, there is such a thing as unique liquidity. A skeptic might suppose that everyone effectively

4 M. Newman, "Finding Community Structure in Networks Using the Eigenvectors of Matrices," *Physical Review*, 2006.

has access to all the dark pools—either directly or through trading intermediaries—and that therefore there is little benefit to accessing them all.

This note shows that the answer is no—at moderate times scale of a minute, venues matter. Having access to a broad range of dark pools is valuable. If liquidity flowed freely among all the venues, we would expect to see a uniform grey—conditioned on a fill in one venue, we're equally likely to see a fill in any other venue—or perhaps a simple gradient from highest volume to lowest. That is not what we see. We see that liquidity tends to cluster roughly in two communities: lit exchanges and dark pools. We tend to get fills together within each group of venues, but not as much between groups. There is further structure indicating strong connections between certain pairs and groups of dark pools that could be caused by different routing arrangements and preferences among broker algos. The data suggests that brokers who have not established a broad set of interconnections with other dark pools, or who favor their own dark pool to the exclusion of others' may be missing liquidity.

The results of this study also point to a potentially more efficient way of allocating liquidity among venues by taking advantage of the correlation structure. For example, especially when working with smaller orders or probing for orders with larger minimums, we might route away from smaller venues that are highly correlated with others. When we get a fill, we might also route dynamically and with a preference, to venues within the same group to increase our probability of a fill.

In a forthcoming research note, we will explore other characteristics of dark pools around liquidity and execution quality, and see the extent to which venues differentiate themselves in these dimensions.

TO SEE ADDITIONAL THOUGHTS FROM PRAGMA, VISIT OUR BLOG AT HTTP://BLOG.PRAGMATRADING.COM
FOR QUESTIONS OR COMMENTS, PLEASE EMAIL DR. ERAN FISHLER, CHIEF OPERATING OFFICER (RESEARCHNOTES@PRAGMATRADING.COM)