



Who is my peer?

Darren Toulson, Head of Research, LiquidMetrix.

Following any presentation to a buy-side of the top line performance metrics in a TCA / Execution Quality Report, the natural question is “So, is this performance good or bad?”

An absolute implementation shortfall (IS) number of 8.6 BPS may sound good – most people talk about IS numbers in double digits – but is it really good? Or is it simply that the types of orders you are doing are relatively ‘easy’. Similarly if broker A has a shortfall of 4.6 BPS and broker B has a shortfall of 12.6 BPS, is this simply because you’ve trusted broker B with your tougher orders?

One approach to the difficulty of interpreting performance metrics is to compare each of your order’s outcomes to some kind of pre-trade estimate of IS and risk, based on characteristics of your order such as the instrument traded,

percentage ADV etc. This may work fairly well when looking at the relative performance of brokers for your own flow, where you can take relative order difficulty into account. However, in terms of your overall, market-wide TCA performance, how can you be sure that the pre-trade estimates you’re using are at the right absolute levels? Many pre-trade estimates themselves come from broker models. How realistic are these for all brokers’ trading styles and how can you be certain that they’re not over or under estimating costs relative to the market as a whole and thus giving you a false picture of your real performance?

To get an idea of how well your orders are performing versus the market you need to compare your own performance to orders done by other buy-sides.

“The principal difficulty with any kind of peer comparison lies in the fact that you will be comparing your orders with orders from many different buysides; each with different investment styles in different types of instruments, given to different brokers with different algorithms.”

Darren Toulson, LiquidMetrix

The principal difficulty with any kind of peer comparison lies in the fact that you will be comparing your orders with orders from many different buysides; each with different investment styles in different types of instruments, given to different brokers with different algorithms. So if you compare your orders to some kind of ‘market average’ how meaningful is it? Who exactly are your peers?

For the results to be meaningful and believable it’s necessary to be open on how exactly peer comparisons are constructed so as to be sure that we’re comparing like with like.

Order similarity metrics

The starting point in any type of peer analysis is that your orders should be measured against other ‘similar’ orders. But how do we measure similarity?

Consider an order to buy 6% ADV of a European mid-cap stock, starting at around 11am and finishing no later than end of day. Apart from the basics of the order, pre- or post-trade we can also determine many other details such as: the average on-book spread of the stock being traded, the price movement from open of the trading day to start of the order, the price movement in this stock the previous day, the annual daily volatility of the stock, the average amount of resting lit liquidity on top of the order book for this stock and the number of venues it trades on, etc. It’s easy to come up with a list with many different potential ‘features’ that can be extracted and used to characterise an order.

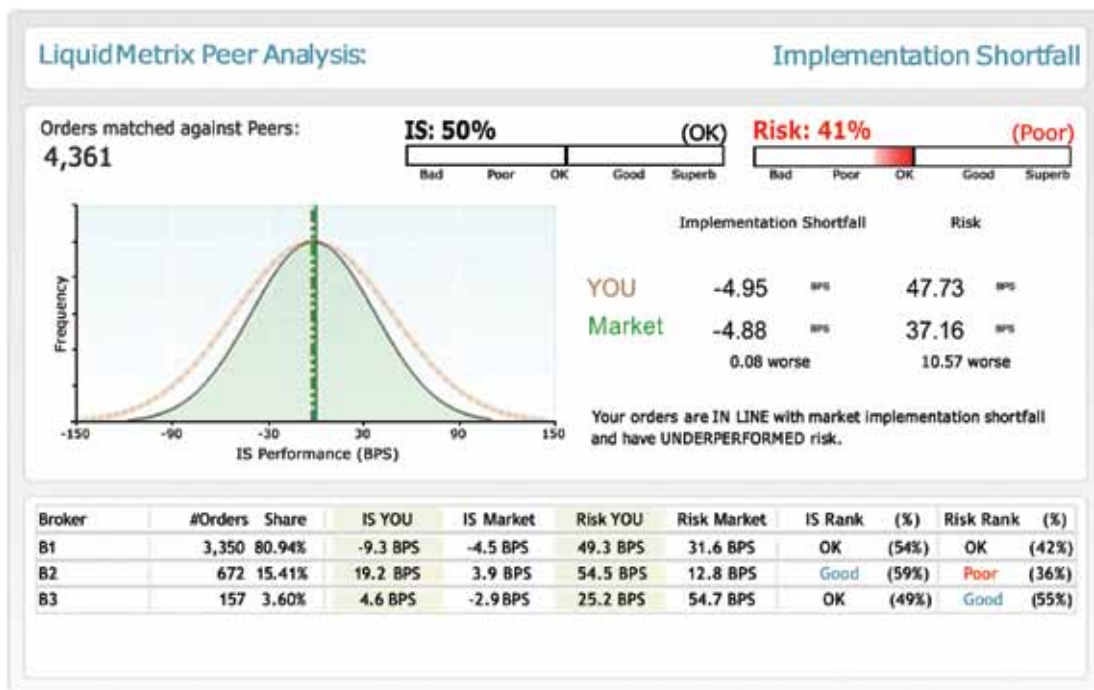
Each of these features may or may not be

helpful in determining how ‘easy’ it might be to execute an order at a price close to the arrival price. Some features make intuitive sense. Trying to execute 100% ADV in a highly volatile stock with a wide spread is likely to be much more expensive and risky than executing 0.1% ADV in a highly liquid stock with tight spreads and little volatility. But how relatively important might yesterday’s trading volume in the stock, or market beta be in predicting trade costs? Which are the best features to use?

Assume we’ve come up with 100 different potential features that might help characterise an order. If we’re designing a similarity metric that uses these features to find similar orders to compare ourselves to, we need to do one or more of the following:

- Identify which amongst the 100 features are best at predicting outcomes such as Implementation Shortfall (Feature Selection).
- Combine some of our selected features to reduce any data redundancy or duplication of features which are telling us basically the same thing (Dimension Reduction).
- Come up with a weighting of how important each remaining feature is when looking for similar orders (Supervised Statistical Learning).

The good news is that there are decades of academic research on how to do most of the above. Methods such as stepwise regression, principal components analysis, discriminant analysis and statistical learning (KNN, Support Vector Machines, Neural Nets) all lend themselves well to this type of analysis.



The upshot of all this is that for each buy-side order we wish to analyse, we produce a similarity measure that can be used to find, from a market-wide order-outcome database, a set of the, say, 100 most-similar orders done by other buy-sides. They do not have to be necessarily orders done on the same stock, just on the same type of stock. Assuming we've done our analysis well, the TCA outcomes of these orders should represent a good target to compare our order with. If we do this for each of our orders, we discover how well we've really done versus 'The Market'.

An example of peer analysis done well

What might this kind of analysis look like in practice? Figure 1 shows one way of presenting peer results. A set of client orders has been matched, using a similarity metric as described above, against a market-wide order database. We're looking in this example at implementation shortfall (IS); one can look at other TCA metrics such as VWAP or post order reversions in exactly the same way. Using the matched orders we're

able to see the distribution of our buy-side client IS outcomes with market-wide IS outcomes (green).

This tells us how well both the IS and risk (standard deviation of outcome) of our client orders compare to the market average for similar orders. Based on the number of orders analysed and a measure close to a statistical 't test' we can then also translate the differences in performance into a significance scale from 0 to 100 to qualify how much better or worse than average a client's performance is.

Conclusion

A fundamental question buy-sides want to know from any kind of top level TCA analysis is how the costs associated with their orders compares to their peers. The danger of any kind of peer analysis is being certain that you really are being compared fairly to other participants. The solution to this is to ensure that any kind of peer comparison only compares orders with like orders, rather than orders from similar companies, preferably from a large database of orders from different buy-sides. ■