CASH TREASURY SMART-ORDER-ROUTING (SOR) FOR AGGRESSIVE CROSSES

SHANKAR NARAYANAN
JIUYUE ZHOU
PURU SARATHY
MARCH 12, 2025

EXECUTIVE SUMMARY

- We implement a new logic for the cash treasury smart order router (SOR) for aggressive crosses, which allocates quantities across venues based on a ranking algorithm that uses the venues' market impact and fill ratio.
- Simulation results using market data are shown for illustration, demonstrating a value-add of around $1/4^{th}$ of the BTEC minimum price increment ($39.0625) compared to the naive or random venue selection approach, specifically for aggressive crosses only.

INTRODUCTION

An execution algorithm typically has two parts: passive fills to capture the spread and aggressive opportunistic crosses. However, there are vital differences between treasury and future execution, such as: 1) treasury has multiple venues, and 2) those venues have different characteristics, such as tick sizes and heterogeneous dealer streams, in addition to CLOB venues. Therefore, SOR becomes a critical problem for treasury. The upstream execution methodology is similar to our broad algorithm suite, and determines the rest of the logic for sizing, crossing, and placement. Previous QB articles[1] discuss liquidity of different venues, but they are not explicitly used for order routing. This article describes our SOR logic for selecting the venues for aggressive crosses. We also show the simulation results for illustration to show the value added.

PROBLEM

The SOR venue selection for crosses comes into play only when we have a tie between different price targets. Table 1 illustrates the problem by showing hypothetical offer sizes at different levels of different venues.

In the example, the target is to buy 85 million at 105-05 or a better price, so the downstream algorithm will sweep through the levels 100-04 to get 10 million as the target size is greater than the total displayed quantity across all the venues at 100-04. However, there are a few choices to cross the remaining 75 million at level 100-05. While the total size across venues exceeds the target of 75 million, no particular venue will suffice for the need on a standalone basis. So, our current method would send multiple

Illustrative Example

	Venue 1	Venue 2	Venue 3	Venue 4
	CLOB	CLOB	CLOB	NONCLOB
100-05+	40	60	10	100
100-05	60	50		20
100-04+			5	5

TABLE 1. The table illustrates the realistic problem at hand. It shows the offer side displaying the sizes of different venues. For the example mentioned, the venue selection is crucial at levels 100-05, where the total size to execute the remaining order size of 75 million, but none of the venues on their own can complete the order.

child orders simultaneously to different venues. Currently, it would chose venues based on either the client's preferred venue or randomize the venue choices.

However, the analysis of our child orders revealed that specific venues tend to have a higher market impact after the crosses than the others. And while BrokerTec (BTEC) has higher displayed quote sizes, they are not necessarily better in market impact than other small-tick or non-clob venues such as FENICS or LiquidityEdge (LE). The liquidity of a venue such as FENICS can be much higher than the displayed volume or quote size. It is more so for non-clobs like LE, where we don't see traded volume in real time.

Conversely, the displayed sizes on small-tick venues like FENICS and LE with are low quote, which means we go for multiple clips to execute our order sizes. The repeated actions signal our intentions to the market. This would manifest in either the quote fading or the price moving away from us. So, this brings us to modeling the two critical factors of impact against completing the order.

MODEL

We use a loss function as a combination of the two critical factors to rank the venues. The two factors are:

- Market Impact: The price change caused by trading at a particular venue by crossing. It is relevant as it will impact our future child orders.
- Variance: The risk of selecting a venue that provides insufficient quantity or price moving away due to latency and other factors, leading to delayed execution at a price worse than the arrival price. We model variance and not volatility to have a quadratic form.

For each venue i, the loss function to rank is:

$$\text{Rank}_i = \text{Market Impact}_i + \lambda * \text{Variance of left over quantity}_i$$

The market impact on the left side of the equation is measured as the price change ΔP in the subsequent **T** seconds after the fill using historical production crosses; more on the choice of **T** in the next paragraph. The market impact coefficient is modeled as follows:

$$\frac{\Delta P_i}{\sigma} = \beta_i * \frac{x_i}{Q_i}$$

Here, σ is the instrument's volatility per minute and is venue invariant as it is fundamental to the instrument. Q_i is the displayed quote size of the venue i at the target price level, and X_i is the size to be traded at the specific venue. We used quote size as a deflator in the market impact function because we also have access to dealer venues, which tend not to display traded volume in real-time.

The second term is the variance of left over quantity, and is measured as:

$$\text{Variance of leftover quantity}_i = (\frac{X - x_i}{Q_i})^2 * \sigma^2 * \text{T}$$

where,

$$x_i = min(\textit{fill ratio}_i * X_i, Q_i)$$

In the above equations, *fill ratio* is the characteristic of the venue and depends on the stickiness of the displayed size. X_i the remaining order quantity to be allocated at the price tie level. Q_i is the same as described above for the market impact function. So, going back to the example mentioned in Table 1, we had 85 million to cross, and so X_i will be 75 million, which was left after crossing 10 million at 100-04, and Q_i will be the quote sizes of the venues at level 100-05.

Here are a few other comments on the choice of the form and parameters. **T** is based on our analysis that showed an average time to decay to be around 30 seconds. Therefore, in our current setup, **T**=30 seconds. The λ is a subjective choice that gives weight to one over the other and is the same across all the venues and instruments. We set λ to 0.1 per price unit based on our willingness to tolerate risk for the sizes we trade and the preliminary results we observed. In some sense, we put much more weight on minimizing the market impact over the variance of the leftover quantity.

Figures 1 and 2 show the market impact and the fill ratio of different venues based on production fills. Notably, some small-tick venues like FENICS and LE tend to have a lower β, although they have a lower quote size. Consequently, when there is a tie, the above rank will favor the small-tick venues for smaller sizes. In contrast, for large sizes, it will favor the venues with larger displayed quote sizes despite the higher impact, as the variance component will be lower for a venue with larger displayed quote sizes.

We also identified additional factors besides those mentioned above, but our new logic builds a strong foundation. It also produced intuitive results, as mentioned in the following sub-section. The subsequent iterations will be an improvement to the same. For example, previous QB articles addressed the machine learning approach for feature engineering[2] and can aid in our forthcoming versions for measuring the market impact and fill ratios. Additionally, there could be other forms of the loss function across all venues that would be more accurate, but we chose to use the above loss function to rank all the venues with a price tie only once, as it is fast and straightforward in real time.

RESULTS AND CONCLUSIONS

We use market data for the chosen venues to show the value-add and compare the above loss function for $\lambda = 0.1$. For a particular day, we sample the quotes of five venues

FIGURE 1
Market impact per lot across different venues measured using production crosses for CT10. It is noteworthy that BTEC has a higher impact than the other small-tick venues despite larger quote sizes.

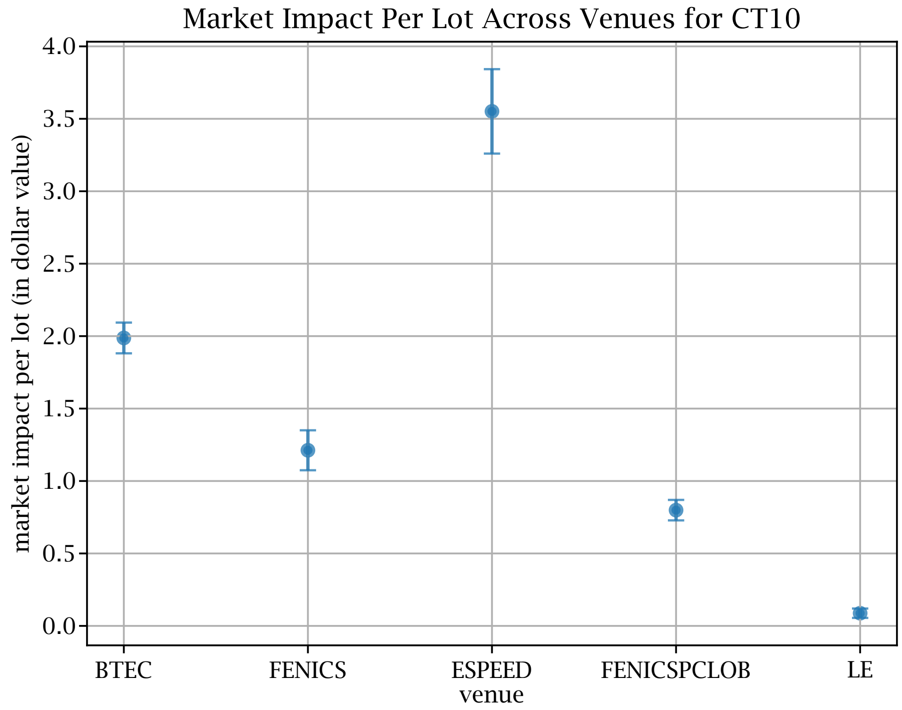
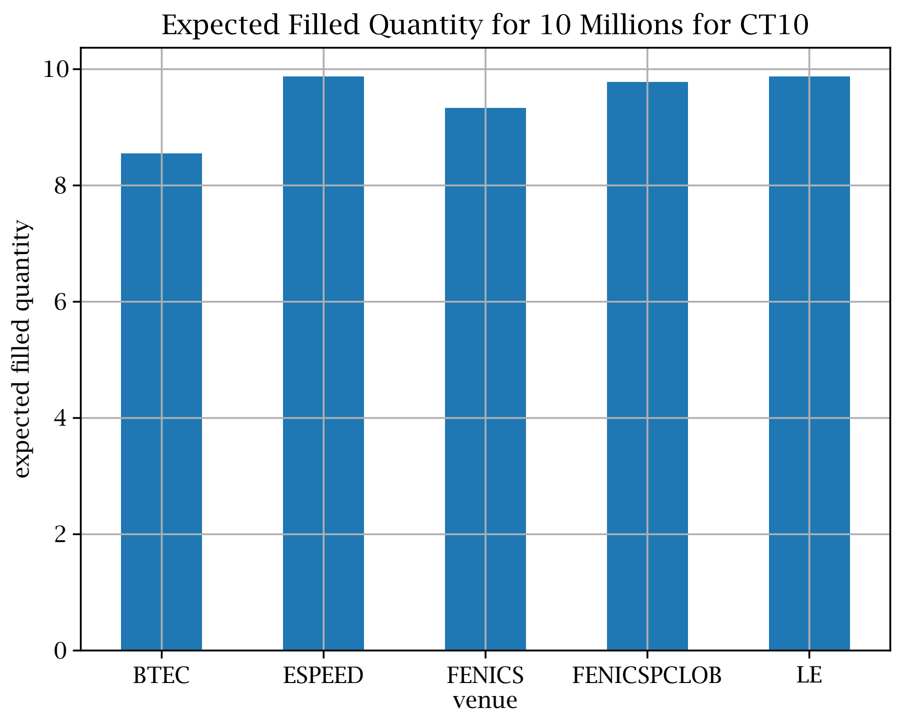


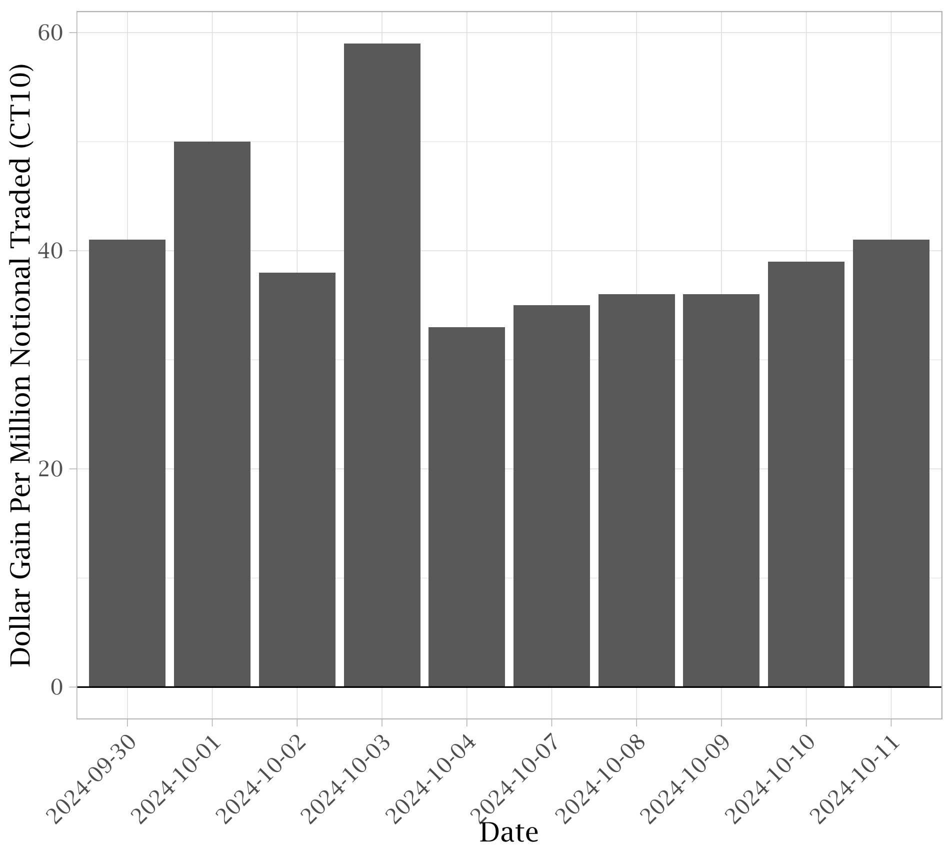
FIGURE 2
Fill ratio per 10 million across different venues of CT10. It is measured using the production data by measuring the quantity filled compared to the displayed size when the child-order was sent.



(BTEC, FENICS, LE, FENICSPCLOB, ESPEED) at the best offer and isolate instances where there are ties between any of the venues. Assuming we are buying, we use our new ranking algorithm and compare against a random selection of venues when we encounter ties at price levels, as mentioned in Table 1. In both cases, we sample the size of the order to be executed from a distribution that mimics QBs orders received from clients on the SMARTDIRECT strategy on all cash treasury products.

The costs are compared between our newly implemented algorithm and the random venue selection approach across the day. Figure 3 shows the daily gain in dollars of the new algorithm over the random one. The gain is per million of the executable sizes when we encounter the price ties. As can be seen, the gain from using the algorithm mentioned above is consistent and significant. The overall gain for a TWAP trader that trades every minute is around \$41 per million notional traded for a product like CT10. The improvement in CT10 was around 27% compared to the random selection baseline. Notably, these savings apply specifically to aggressive SOR orders in price-tie scenarios. As for putting the saving in context, CT10 \$40 per million is $1/4^{th}$ of the BTEC minimum price increment (\$39.0625). Furthermore, these results are based on market data simulations and only indicate the value-add.

FIGURE 3
CT10 dollar gain by using the new SOR algorithm for crosses over a random selection of venues using market data.



In summary, we have the critical components for the SOR algorithm, which combines market impact and the variance of leftover quantities. Our market data simulations are encouraging, although preliminary. We will test the improvements as A/B testing in production and report the improvements. In our forthcoming versions, we will further improve the impact function and the fill ratio measurements.

References

[1] Shankar Narayanan. US Cash Treasuries Liquidity: Comparison of Two Venues. Quantitative Brokers, 2020.

[2] Renyuan Xu and Isaac Carruthers. Machine learning for limit-order routing in cash Treasury markets. Technical report, Quantitative Brokers, Jun 2018.