

The Waiting Game

Capacity is a slippery topic in the CTA industry. Smaller capacity allows for a more meaningfully diversified portfolio but today I want to talk a little about how capacity affects execution.

If you run a large CTA, your team probably produces beautiful trading cost analysis (TCA) reports for each of your markets that show that, despite running \$10bn, \$20bn or \$40bn of AUM, realised slippage tracks model slippage very well. Your TCA reports benchmark your own trades against a TWAP/VWAP algorithm, and your execution algos perform very well. Your execution team is operating like a slick machine, on a tight schedule, with each trade monitored and executed within specified parameters and specified time.

For the volumes you trade, efficiency is crucial.

So indulge me as I make the case for making execution as inefficient as possible.

Allocating risk...

For a quant CTA, allocating risk is everything. If I invest a unit of risk in a single asset for a day, I may not be making money, but I am taking on a unit of variance.

If I invest for a year and each business day was an independent bet, I would be accumulating over 250 units of variance and approx. 16 units of volatility.

So, for example, if I try to run a CTA with a target annual volatility of 16% that basically means my daily target is to hold “a unit of risk”: a position with ex-ante P&L volatility equivalent to 1% of AUM. That unit of risk is allocated among assets and strategies so that the aggregate ex-ante volatility matches our target unit of risk.

Giving the execution desk some risk

Let us imagine you run a mid-frequency CTA running a *single* asset. You trade a *monthly* momentum signal and have (on average) one unit of risk per day in the asset. You turn over your position *monthly* so on average, trade just about 0.05 units of risk a day (assuming at least 20 business days). You throw that 5% of risk as buy/sell-requests over the fence to the execution desk and ask them to fill the order.

Your CTA is running two books now: The original CTA book, running 100% unit of risk per day and the execution desk book, running approximately 5% of risk per day.

How much does execution cost?

For a reasonable sized CTA focused on the super-liquid futures, costs will be, ballpark, 10 basis points of Sharpe. As we move to less liquid assets, we may move to 30 – 40 bps of Sharpe. Market impact is, unfortunately, very real.

There are two ways for the execution desk to save money: by being passive and by being clever.

- Being passive is about reducing market impact by providing liquidity and waiting for the market to come to you.

- Being clever is about using clever algorithms to predict short-term price movements.

We are going to ignore the passive approach for the moment and ask the simple question: How much of the costs can be saved by being clever?

Unfortunately, next to nothing.

And the reason is very simple: the trading book has very little risk to play with.

Estimating the risk of the execution desk book

Our initial estimate was that the execution book risk is 5% of the big book but that is an over-estimate: the trading desk can only decide on the *timing*. By end of the day, it must fill the order: the desk is a “day-trader”: they never accumulate risk on weekends or overnight. Their risk starts when the order arrives after start of day. Their position must decrease throughout the day until some time before close of trading. This reduces the execution book risk to, at the very most, 3% of the main book.



Figure 1: The risk profile of the execution desk through the day

You may think the execution book will struggle to make a difference with only 3% of risk but it's just about to get much worse...

Correlation hath given, and correlation hath taken away...

A CTA trades around 100 markets. Since these are quite correlated at mid-frequency, we scale the positions (and hence trades) down quite a bit to make sure we still have “one unit of risk” daily.

Except these trades do not look very correlated to the execution desk. Position correlation is much higher than trade correlation and as we move into faster time horizons, correlation between markets break down. The P&L of the execution desk per each order is essentially uncorrelated, making its overall risk (when compared to the CTA book) at most 1% in size.

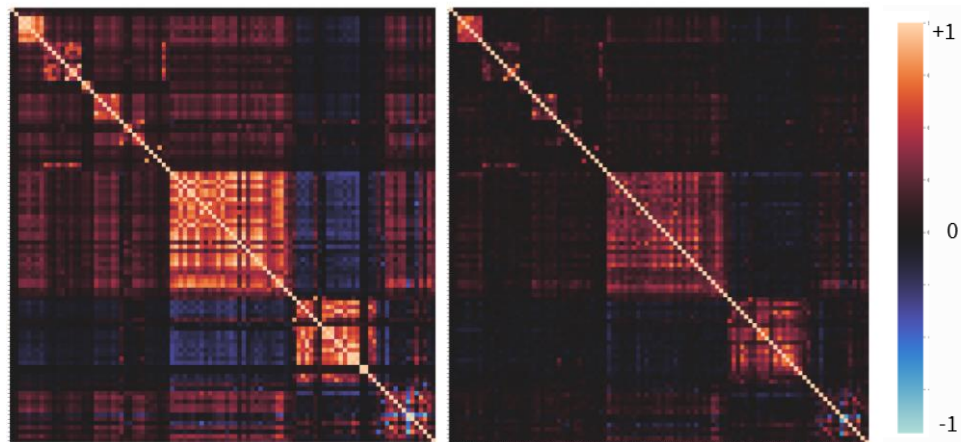


Figure 2: Position correlation (LHS) vs trade correlation (RHS) for 103 future markets running mid-frequency trend: Black regions represent areas of low correlation. (source: GQ analysis)

Trapped by increased capacity?

And now imagine you are a CTA running \$10bn. As you got bigger, your trade size became bigger, and you observed real slippage. The reason is that you are too big and, guess what, your costs *have* increased. Even if trading liquid markets, you are now paying north of 20 bps in Sharpe for market impact.

Your clients got a little antsy about capacity and your TCA reports looked messy. You have a Eureka moment: you split the trading during the day into waves: you trade, say, every 2 hours, 10%-20% of the portfolio. On the face of it, slippage has gone back down because each trade's executed fills now track the recent wave's price print that generated the trade very well. Unfortunately, the waves are not independent and the slippage from the morning wave also affects the afternoon wave. Worse, you now have higher autocorrelation between successive orders and other participants in the market can see it. It is a sad mathematical fact of life that the trading costs of the optimal *constrained* execution algorithm cannot improve on the trading costs of the optimal *unconstrained* algorithm.

Most relevant to my argument is that you constrained your execution desk's risk *even further*. Instead of trading 5% of the asset's risk at any time, the execution trades at most a fifth of that. So how much risk is the desk left with?

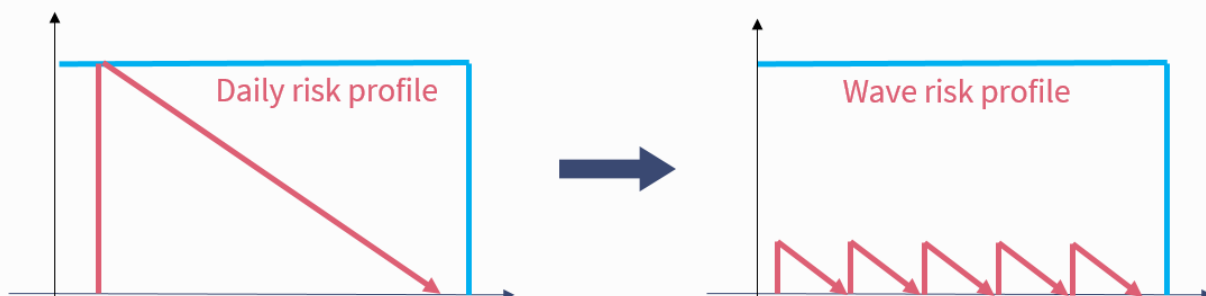


Figure 3: A 5-fold decrease in risk profile for the execution desk due to the introduction of 5 waves.

For a modern, large AUM, mid-frequency CTA, the execution desk runs ~0.2% of the overall book risk.

Can you make any money by being clever?

No. No price prediction algorithm would be able to make the slightest bit of difference to your execution costs. You can buy level-2 order book data, use AI, you can spend money on GPUs, direct market access or a large execution research team. But all for naught: even a Sharpe-5 model (at which point you should resign, return the AUM to your investors and run your own prop shop) on 0.2% of the book risk, would only reduce your costs from 20 bps to 19 bps Sharpe.

But does slippage look like it has gone down?

Yes. Slippage is now “well understood”. The model slippage and the theoretical slippage now track very well: we have managed to reduce potential deviation (and potential improvement!) from the benchmark model to near zero.

Does it matter?

In some sense, no... the execution desk is running a (order updating, wave based) VWAP which is a sensible algorithm. Your TCA reports look beautiful, all your benchmarks look good. Your performance suffers but you are unaware of the reason: you feel like you can even increase capacity. The reality is that performance degradation and price impacts due to your AUM increase, are now “hidden” in the trend model itself and that the very sensible actions you took to reduce slippage, made your trading more aggressive and increased your trading footprint and market impact.

Making a difference

As far as I can tell, there are only two approaches which would allow execution desks to make a difference: either trade faster models (for HF firms, their whole book is the execution book) or execute slower.

Executing slower has several effects: Firstly, your execution book risk share goes back to the 5% range, if you allow (and can back-test) multi-day execution. If you do have a clever, short term, price-predicting algorithms, (or just good traders) they may have some risk to work with.

Secondly, you can speed up your book: CTAs artificially slow their book to ensure a given speed and lose alpha in the process. But you can just let the trading desk do that for you.

The last and the most important effect, is that you can be more passive. Being passive is very distinct to predicting price action: it is about reducing your footprint, providing liquidity rather than taking it. And waiting. You still need to work hard to source liquidity and get a good execution, so improvement is not guaranteed, but for a CTA execution desk, the waiting game is the only game in town.

Yoav Git
Quant Research



This material is provided for informational or educational purposes only and does not constitute a solicitation of any securities in any jurisdiction in which such solicitation is unlawful or to any person to whom it is unlawful. Moreover, it neither constitutes an offer to enter into an investment agreement with the recipient of this document nor an invitation to respond to it by making an offer to enter into an investment agreement.

This material may contain “forward-looking” information that is not purely historical in nature. Such information may include projections, forecasts, estimates of yields or returns, and proposed or expected portfolio composition. Moreover, certain historical performance information of other investment vehicles or composite accounts managed by Nuveen may be included in this material and such performance information is presented by way of example only. No representation is made that the performance presented will be achieved, or that every assumption made in achieving, calculating or presenting either the forward-looking information or the historical performance information herein has been considered or stated in preparing this material. Any changes to assumptions that may have been made in preparing this material could have a material impact on the investment returns that are presented herein by way of example.

This material is not intended to be relied upon as a forecast, research or investment advice, and is not a recommendation, offer or solicitation to buy or sell any securities or to adopt any investment strategy. The information and opinions contained in this material are derived from proprietary and non-proprietary sources deemed by Nuveen to be reliable, and not necessarily all-inclusive and are not guaranteed as to accuracy. There is no guarantee that any forecasts made will come to pass. Company name is only for explanatory purposes and does not constitute as investment advice and is subject to change. Any investments named within this material may not necessarily be held in any funds/accounts managed by Nuveen. Reliance upon information in this material is at the sole discretion of the reader. Views of the author may not necessarily reflect the view s of Nuveen as a whole or any part thereof.

All information has been obtained from sources believed to be reliable, but its accuracy is not guaranteed. There is no representation or warranty as to the current accuracy, reliability or completeness of, nor liability for, decisions based on such information and it should not be relied on as such. For term definitions and index descriptions, please access the glossary on nuveen.com. Please note, it is not possible to invest directly in an index.

Past performance is not a guide to future performance. Investment involves risk, including loss of principal. The value of investments and the income from them can fall as well as rise and is not guaranteed. Changes in the rates of exchange between currencies may cause the value of investments to fluctuate.

This information does not constitute investment research as defined under MiFID.

Nuveen, LLC provides investment solutions through its investment specialists.

3478823

GCM-3442071PE-00224X