Corvil

Re-Thinking Speed In Financial Markets

BY DONAL BYRNE CEO, CORVIL

CONTENTS

INTRODUCTION

The recent decision over the IEX has once again re-ignited debate about speed and its role in today's electronic traded financial markets. Unfortunately this debate often polarises people into one of two views – speed is good for markets or speed is bad for markets. This is a little simplistic but does fairly describe the argument and actions of people in their respective camps. I believe we are at a point where we should be re-thinking the role of speed in financial markets. The current mental model is limited and needs to be updated. What follows is a four part blog series where we explore the new role of speed in modern market structure.

PART 1

Taming The Speed Monkey

TAMING THE SPEED MONKEY

In 1829 the British built the Royal Calcutta Golf Club (RCGC) in Kolkata. It is today, the oldest golf club in India and was the first outside Great Britain. When the golf course first opened for play, the British encountered a new type of problem. Monkeys would drop out of the trees, scurry across the course, and take the golf balls. The monkeys would play with the balls, tossing them here and there. Some of the time the player's position was disadvantaged by where the monkey dropped the ball and some of the time the player's position was advantaged. The outcome was largely random. This was a major disruption to the game that is governed by a strict set of situational rules and personal etiquette. At first, the golf club officials tried to control the monkeys. They erected high fences around the fairways and greens. This held some promise initially, but pretty soon the officials discovered that these fences were no challenge for ambitious and creative monkeys. Next, the officials tried capturing and relocating the monkeys, but there was no shortage of other monkeys that would take their place. They tried loud noises to scare the monkey's away. Nothing worked. In the end, they arrived at a solution. They added a new rule to the game – "Play the ball where the monkey drops it". This was referred to as "the monkey rule".

The recent debate and argument about IEX, can't help but remind us of the situation faced by the Royal Calcutta Golf Club officials. Unfortunately, it seems likely our industry will repeat the pattern followed by the British golf officials, with similar results. While analogies are never perfect, the reader will recognise some obvious parallels in the actions already taken by regulators, venues, and participants to tame the speed monkey. The introduction of "speedbumps" is the latest action. But this follows a long list of actions already taken in an attempt to counteract effects of speed. Remember the rules for equal length cables in co-location facilities and

Reg. 603a rules for price distribution on direct versus consolidated feeds. What about the attempts to round up those pesky prop-traders and ship them off only to have them re-appear in the form of quant hedge funds. Can anyone see a pattern emerging here? My prediction is that all these actions will ultimately prove ineffective as we look to build a robust, transparent and fair market structure. There needs to be a collective understanding and acceptance of the true nature of speed and the role it plays in market structure. Only then, will it be possible to see what sensible rules might be put in place which cuts through the “Gordian knot” that has emerged from the adoption of high-speed trading technology.

A NEW MENTAL MODEL FOR SPEED

Speed can be thought of as a proxy for price certainty. In a faster market, you are more likely to secure the price as advertised, i.e. the price is less likely to have moved when you attempt to execute a trade. In a slower market, the price advertised is more uncertain. It may have moved by the time you can respond to it, i.e. the price you receive in a slow market might be stale.

An analogy is to think of speed and its impact on market price to be like the weather. The weather is a random natural phenomenon. It changes constantly. It is hard to control. It operates as a complex, dynamic and a highly interdependent system. There are seasons where the weather is warmer e.g. summer. There are geographies where we expect the weather to be colder or wetter. But on any day, of every season, everywhere on our planet there is a level of uncertainty as to the what the weather will be. This is something we accept without much questioning.

What if we decided that it is unfair for all humans on the planet to experience different weather? How would we solve this problem? There are two approaches that one could take. Approach one would be to move all people on the planet to the same location. We would probably have to stack people on top of each other to fit them into a space where it would be precisely the same weather. Approach two would be to build an atmospheric shell around planet and then precisely control the system such that we can deliver the same weather experience to everyone. Most would accept these ideas as ridiculous and futile. We have learned to live and adapt to the weather we are given day for day. Even if we live in a country or state with poor weather relative to another location, we often choose not to re-locate because there are other aspects of where you live that are more important to your quality of life e.g. family.

However, there are two really important things we do as a society in dealing with weather. Firstly, we attempt to gain better insight into weather patterns and make better predictions with weather forecasts. This becomes critically important for people whose daily lives depend on knowing the weather e.g. fishermen. Equally, being able to predict major storms and taking appropriate action is a consequence of us dealing with something that is inherently random and difficult to control. Second, we now finally realize the importance of protecting the planet from global warming. We understand as a global society our actions are impacting the integrity of how our weather system works. This is where we need our governments to step in and help. Therefore, there are three really important things that regulators need to do relative to the speed question in electronic traded markets:

1. It must provide the right level of speed transparency and visibility to all players
2. It must build a market structure that can accommodate speed variation and diversity
3. It must not damage the operating integrity of the system by the action it takes

PART 2

Policing Speed

POLICING SPEED

If we assume that we are unlikely to make trading via computers illegal any time soon, I have come to the belief that all attempts to regulate and police speed to artificial limits will ultimately prove ineffective.

Faster markets enable narrower spreads and greater price certainty. Slower markets tend to have wider spreads and less price certainty. No market is uniformly slow or fast. Proximity to market is not the only determining factor. Varying traffic patterns on trading networks can cause corresponding variation in delay due to congestion. These delays are sometimes larger and more difficult to manage compared to deterministic transit delays. Unless there is zero traffic, this phenomenon will always be in play. This is why it is possible for someone in a co-location facility to lose a fill opportunity to a participant not in the co-location facility. We have seen this happen and it is straightforward to understand why it happened when you have the right framework and right data to analyze.

Adding speedbumps is not a good general answer but does provide differentiation local to the venue e.g. IEX. Speedbumps introduce speed asymmetry between interacting participants and venues. This adds more complexity to an already complex system because it is harder to figure out the likely execution outcome when there is a mix of fast venues and slow venues. It is speed asymmetry that creates the opportunity for speed arbitrage. Even if you block off a speed play at one venue, you typically end up creating (not intentionally) multiple new speed plays across the population of interacting venues, resulting in proliferation of new order types, rules and infrastructure enhancements that either encourage or discourage exploitation of the effect.

It is impractical to build a marketplace where speed

is guaranteed to have no impact. To do this, we need an environment where everything happens so fast relative to the time required to execute a transaction that we can consider all actions to effectively happen simultaneously. The SEC recently declared that events in US Equities Markets within a millisecond to be de minimis i.e. they can be treated as if they happened simultaneously. In today's markets, high speed algorithms respond to price updates and make trade decisions in less than 10 microseconds. Participants can execute orders within 25 microseconds. Hundreds of market orders can be executed across multiple venues within one millisecond of time. Unfortunately, the SEC declaration is out by about a factor of 1000. It would be more accurate to say events within a microsecond are de minimis. To make this a reality, we would need all venues and all participants to be no more than a microsecond apart in time, i.e. everyone co-located within the same building. Even then, we would have to deal with the random delay variations that occur over short-timescales due to varying traffic patterns and dynamic capacity limits causing random congestion and uncertainty in outcome. In fact, as transaction times get faster we would need to co-locate all participants and venues on the same computer chip. This is theoretically possible in the future but not really a practical solution. Therefore, we have to live with the realities of speed and be able to deal with the impact of co-location, speedbumps, direct data feeds, consolidated data feeds, random delay gateways, random traffic congestion, and venues at distant locations e.g. west coast.

A good market structure design must be able to handle diversity and unforeseen actions. No one rule or one size will work for all people all of the time. Just as democracy allows for people with opposing views, beliefs and actions to coexist within its system, our market structure should allow for any expression of speed (fast or slow) or speed variability that may arise naturally or may be employed purposefully.

PART 3

The Speed Rule

THE SPEED RULE

We are entering a new chapter in speed where information about speed will become more valuable than speed itself.

In trading, two important attributes that market participants seek are liquidity and best execution. The ability to provide both of these is critically dependent on price certainty. We define price certainty as the probability of being able to access the price advertised by the market. Fast venues with fast participants generally operate with greater price certainty. However, slow participants on fast venues will have less price certainty compared to fast participants. Price certainty on slow venues will be less than that on fast venues. Fast participants on slow venues will have less price certainty advantage over slow participants as all participants have less price certainty. If only it was that simple.

In US Equities Markets we have a mix of fast participants and slow participants, fast venues and now slow venues (IEX), all interacting electronically based on a complex set of rules (e.g. order protection rule) and sophisticated order types - immediate (IOC) orders and now "extended life" orders (Nasdaq). We have co-location, speedbumps, random order processing, and now clock-synchronized multi-venue order submission (see US patent application14/451,356). All techniques used to leverage or de-leverage the impact of speed, or more accurately differences in speed, on execution outcome. Add to this, random traffic effects creating random delays that complicate prediction of execution outcome and quality. This is why I believe it will prove futile to regulate speed. It is impossible to eliminate the effects of speed from the equation unless you ban computers completely. What to do?

At the risk of simplifying, the faster you can receive market price and the faster you can respond to that

price, the more likely you will be able to hit that price. The more a price is delayed and the slower you are in responding to a price, then the more likely the price is no longer available. In general, a fresh price is better than a stale price. Fresh prices are more certain. The "age" of a price (i.e. the difference in time between when a price was first created and when you received it) becomes a critically important metric to account for the potential effect of speed on execution quality.

Think of "age" as a risk coefficient to be used with "price". It quantifies the probability of the price existing when you look to execute on it. We are all familiar with the "best before" date that typically accompanies perishable products. When we go to our local grocer we intuitively ask "is this today's bread?", "is this meat freshly cut?". We always check the "best before date" on packaged perishables. In exactly the same manner, advertised market prices are perishable and need to be consumed before they go stale.

This then leads to the maxim I call the speed rule:

The Speed Rule - Decide to trade, once you know when the price was made.

Once we know the age of a price, we have transparency. We can then reasonably judge the validity of the price on offer and make an informed decision to trade or not trade based on our prior experiences of trading with prices of a certain age on certain venues. This can be applied uniformly across all venues irrespective of their speed. Of course, you may choose to weight your order routing decision to a fast venue or a slow venue depending on your execution experience and overall strategy. The point here is that once you observe the speed rule, then we no longer have to prescribe a specific speed or speed limit before we decide to trade. We simply deal with the speed environment that exists, and make informed decisions to trade that make the best of the situation. Just like how we deal with the weather.

So what is the catch? Unfortunately, there is one. We cannot accurately determine the age of a price on today's markets because we don't know time with sufficient granularity and accuracy. To determine the age of a price we need two things:

1. A timestamp of when the price was generated
2. A timestamp of when the price was received

The age is the difference between both timestamps. The problem is that timestamps used today by venues and participants lack the necessary granularity and are not synchronized to a common time reference across all venues and participants. Therefore, the age calculation is inaccurate. Sometimes the calculations using today's venue timestamps are grossly inaccurate e.g. negative age - implying that the future happened before the past. This means we have to guess or deduce the real price of the market. This lack of precision synchronized time is the urgent problem to be solved by regulators. If we can't tell time in the machine traded world, and can't see exactly what is happening, we will never have a market structure that everyone accepts to be transparent, fair and robust.

PART **4**

The Need for Machine-Time Data

THE NEED FOR MACHINE-TIME DATA

Machine-time data is essential to operate electronic traded markets transparently and fairly.

While speed improves price certainty, it also makes it harder to see what is going on. The faster an object travels, the more difficult it is for an observer to make out the details of the object. I believe a big part of the human concern over high-speed trading is born from the fact that we can't easily or precisely see what the machines are doing. It's all a blur.

I believe the root problem is not speed. It is time.

Albert Einstein gives us a hint at the solution to our problem. He explains "time is what prevents everything from happening at once". Time is the essential mechanism that allows us to live our life in an orderly fashion. Time makes sense of the progression of existence and events that occur in irreversible succession from the past through the present to the future. Without time, we could not make sense of events that happen. We could not tell the past from the present from the future.

To make sense of time we must observe it with appropriate granularity and accuracy. Let's say you are at your local grocer and you are buying milk. You look at the "best before" date and see that it only has the year specified i.e. no day and no month. You look at another milk producer's product and discover the same thing except it is quoting a different year. Then you think "what date is it anyway"? You look at your watch, and all it has is a single number – the year. You know however that milk will go sour within a week. The problem therefore is that if you buy the milk, you have no valid time data to inform your decision to purchase. You have no idea if you are buying fresh milk or sour milk.

The exact same situation arises in electronic trading

when our ability to observe events is limited to a time granularity and accuracy that is much larger than the time it takes a machine to trade. The SEC recently claimed a millisecond to be de minimis for events in US Equities Markets. In the machine world, a decision to trade can be made in 10 microseconds or less. We are now seeing sub one microsecond algo decision times in FPGA implementations. Therefore, a millisecond time granularity in the machine world would be equivalent to us living our daily lives where time could only be measured with a granularity and accuracy of a single day. Everything we do in a day would be considered to happen at the same time. Needless to say, this would make life very confusing and problematic.

The problem is that we cannot tell time in today's markets.

We must be able to tell time in a machine world so we can observe and control accurately the actions of machines that we entrust to trade on our behalf. People often use the term "real-time" to infer that we can see things as they actually happen. In a human world we typically equate real-time to be approximately a second. If you get a response within a second, we generally consider this real-time. I call this "human real-time". This explains why most humans wear a time management device on their person with a granularity and accuracy of one second. We call this a watch. A machine world is different. Machines act much faster than humans. Their idea of real-time is much closer to a microsecond. Roughly a million times faster. I refer to this as "machine real-time" or "machine-time" for short. We define machine-time as the time within which a machine can act or make a decision. We therefore need a machine-time watch for a machine-time world.

There are signs that this understanding is happening. MiFID II in Europe has embraced this understanding in part with its recent rules mandating that all business clocks involved in high-speed trading must be synchronized to within 100 microseconds of Coordinated

Universal Time (UTC) with a timestamp granularity of a microsecond or better. The only fly in the ointment is that 100 microseconds accuracy is not sufficient to provide the levels of visibility required to detect market abuse reliably. The details are referenced in ESMA RTS-25.

MiFID II also requires investment firms to maintain a record of all machine data involved in a high-speed trade transaction. This machine data needs to be synchronized to UTC and maintained for a period of five years. The value and quality of this data is highly dependent on the accuracy and granularity of the timestamp associated with each piece of data. For today's trading machines, we need timestamps with a minimum of a microsecond granularity and ideally a microsecond accuracy relative to UTC. Commercial technologies for UTC clock synchronization (e.g. GPS with PPS/PTP signal distribution) can deliver approximately three to five microseconds accuracy. This would be good enough.

We refer to microsecond time-synchronized data as "machine-time data". Machine-time data is the new type of data essential for orderly operation of any electronic trading business and becomes the main data source for assuring trade transaction transparency, execution quality and detection of potential market abuse. If regulators, venues and participants don't address the fundamental need for accurate machine-time data, we will continue to pursue ineffective agendas and ultimately fail to build trust and confidence in the operation of our electronic financial markets.

I will leave you with this final thought as it relates to dealing with speed in financial markets:

> "Most people spend more time and energy going around problems than in trying to solve them."

- Henry Ford - Businessman (1863 to 1947)

Corvil

Corvil safeguards business in a digital world. We see a future where all businesses trust digital machines to algorithmically conduct transactions on their behalf. For some businesses, this future is now. We provide big data analytics products that examine digital machine communications, in machine time, and apply analytical and statistical methods to deliver new levels of trusted, streaming intelligence needed by business, IT and security operations teams to safeguard the transparency, performance, and security of critical business applications and services. Corvil was forged on Wall St where it is trusted by leading financial institutions to safeguard their businesses in a digital world.