Quantitative Algorithmic Trading

Sponsored by

September 2007

Management Summary

Percentage of overall trades executed using algorithmic trading – On the equities side the percentage of trades executed via algorithmic methods stands at around thirty to forty percent. In the debt area, this figure is between zero and nine percent depending on the institution. In FX there is a greater differential, with uptake ranging from singe figure percentages to over sixty six percent of trades.

Trends driving growth – The major trends driving the growth of algorithmic trading are an increased uptake of electronic trading methods, decreasing margins, speed of markets combined with a multiplicity of execution venues, a focus on cross-asset class opportunities and the money made by the major hedge funds.

Two fronts of algorithmic development: Tactical alpha generation versus effective execution – The leading sell side houses in particular are all working on developing their algorithmic trading solutions. A commoditisation cycle will remain in place as cutting-edge algorithms will first be utilised by the banks' proprietary trading desk before offering them to clients.

Market leaders – Interviewees cited Credit Suisse, Goldman Sachs, Morgan Stanley and Deutsche Bank as the leading sell side players. They were also very much impressed with what Citadel had achieved.

The hedge fund edge – Certain hedge funds have garnered a lead in this area through the adoption of swift, purpose built IT infrastructures, the hiring of expensive trading talent and a willingness and ability to manage a greater degree of risk. However, the short termism of certain statistical arbitrage funds, combined with assumptions made on incomplete data, has led to heavy losses during the August 2007 period of volatility.

Increased data flows – The vast increase in market data, from both existing exchanges and from new data sources has been matched by demand to capture and study data that had previously been largely ignored, such as full order book information. This has caused an explosion in the demand for market data storage, which has been paralleled with an increase in the demand for speedy analytical processing. These factors place an immense strain on many bank's existing systems as they attempt to stay in touch in the algorithmic arms race.

Back-testing – Back testing and regression-testing are essential for the development of algorithmic trading, not least due to regulatory pressures. In order to hone the most sophisticated algorithmic tools, banks are looking for richly detailed, high quality data lasting for duration in excess of three hundred trading days.

Existing data infrastructure and the development of algo trading – Internal latency and the challenge of ramping up volumes without losing speed were identified as the main problems with existing infrastructure.

Storage and Analytics – All participants within the research cited the need for as much data as possible, which could accessed and analysed at as low a latency as possible as key to the successful development of an algorithmic trading.

Benchmarking data requirements – At the present time there can be no set benchmarks as to how much data is required and the requirements as regards speed of access. The pace of change and the fiercely competitive nature of the algorithmic market mean that no minimum standard is being set in stone by the major houses. Furthermore, it is unlikely that such standards will develop in the short term. There is much secrecy surrounding who is using which techniques to drive their algorithmic development so the only standards that are likely to be stated will already be obsolete.

Leveraging compliance for competitive advantage – Banks are looking to leverage the vast amount of data that they will be forced to store under MiFID and Reg NMS to generate information on client behaviour and on market movements, as well as client responses to specific market situations.

Integration of risk management platforms – This is considered an important factor to facilitate cross asset trading. However, specific traders' risk can still be managed at localised level via spreadsheet. With increasing velocity and volume of data flows, and the risks and limitations inherent to spreadsheet will become increasingly apparent. Those who can adopt their risk management systems away from traditional closing prices, and make use of the flow of rich data in real-time will have significant competitive advantage.

Real-time risk modelling – The demand for real-time, on the fly modelling is very much dependent upon the market. In a fast moving exchange trading space, the ability to rapidly apply risk models is vital. Over the next twelve to eighteen months this area will grow in importance.

The death of the trader – The human trader is not headed for extinction just yet. However, coming years could see a drastic reduction in headcounts and shift from traditional market making to creation of trading models. Traders will focus on devising more complex algorithmic strategies, whilst the algorithms will act as a vehicle to execute these strategies.

The application of complex event processing to algorithmic trading – Whilst there is a great deal of chatter around complex event processing, uptake has been limited. However, as many of the dislocations that current algorithms are designed to take advantage of are arbed out then the application of more complex techniques will become necessary.

Strategic priorities – The adoption of algorithmic trading is probably the number one strategic priority for the major broker-dealers at the present moment.

Introduction

All across the investment banking industry there is a buzz with talk of algorithmic trading. These techniques can be defined as the placing of a buy or sell order of a set quantity into a quantitative model that automatically generates the timing of orders and the size of orders based on goals specified by the parameters and constraints of the algorithm. As a number of hedge funds rake in extraordinary revenue via their algorithmic techniques, investment banks are racing to catch up.

The application of algorithmic trading threatens to revolutionise the entire trading environment. With the rise of the machines, trades will be conducted ever faster and at ever higher volumes. The most radical proponents have argued that algorithmic techniques threaten obsolescence for the human trader, replaced by the split-second accuracy of a delicately coded, multi-purpose suite of automised trading techniques.

The leading investment banks find themselves in a sprint towards ever faster and more accurate executions, with latencies that seemed inconceivably rapid one quarter rendered sluggish the next. Techniques once used only in the equities sphere are spreading to other asset classes and, indeed, to the burgeoning number of cross-asset class trades.

This Executive Summary examines the actions of players across the industry, the current trends and drivers in the market, data and risk issues and what lies in the future. Full acceptance of algorithmic techniques will transform the securities industry and this report will investigate how banks hope to develop their algorithmic trading systems to take advantage of present and impending opportunities.

Current Trends

The rising tide

That the trading environment is being transformed by algorithmic techniques is beyond doubt. However, it is important to gain a measure of precision in what we mean by algorithmic techniques. Algorithmic trading is more than simple electronic execution, these methods can be applied to a diverse array of areas, from looking to minimise execution costs to pattern-seeking for minute arbitrage opportunities.

For tier-1 institutions there is a perception within the market that, on the equities side currently approximately thirty to forty percent of trades are executed algorithmically. There are a number of reasons for this. Firstly, equities are concentrated volume products – the ratio of the number of equities to the number of market participants is much higher than it is for options or corporate bonds – thus there tends to be a higher turnover ratio. Equities have a relatively high bid-offer spread to volatility, something that algos can profitably exploit. In addition, equity products are typically exchange traded and there are significant volumes traded electronically.

The uptake across other asset classes has been nowhere near as advanced. Each asset class poses a distinct set of challenges to the algorithmic trader, affecting the way they can apply their techniques. On the debt side, for example, the reported figure traded is much lower than for equities, merely a single figure percentage. The fundamental issues in the bond market are a lack of transparency and an advanced electronic trading framework by which algorithms can be applied to the market. One source qualified this, by arguing that there was a definite realisation of the need for greater adoption of algorithmic trading methods across a broader range of products. The feeling within the leading broker-dealers of a need to make a move has been growing in the last 18 months, and now, they are finally laying down detailed plans of action. There are also fundamental differences in the way aspects of the debt sphere are traded that makes some more suited than to algorithmic techniques than others. There is much less opportunity for trading corporate bonds because it lacks a critical mass of active market participants required to stealthily apply algorithmic techniques. Currently, the bund, bobl, shatz and rate futures market's huge liquidity is way past this critical mass thus participants are not exposed to the liquidity risk that can arise from high volume algorithmic trading.

Across FX, the uptake of algo trading is mixed. Although superficially more suited to algo trading due to the electronic nature of the market, uptake has yet to reach the level of equities. Some houses boast over two thirds of their FX trades executed algorithmically with a larger number of price indications being generated by the same methodologies. However, Lepus research has indicated that some large scale global institutions, with significant emerging market interests, execute as little as three percent of their trades via algorithm. As the "old money" increasingly looks towards currencies as a method of portfolio diversification and alpha generation, demand for effective algorithmic execution looks likely only to increase. Such execution is facilitated due to the deep liquidity in the FX space, meaning large scale pattern trades can be undertaken without detection, and because its an increasingly electronically traded market.

Tier-2 banks are lagging behind their bulge bracket counterparts. Many do not have algorithmic trading strategies in place in the rates arena, but these institutions are generally hoping to bring algorithms online over the course of the next year. One source estimated that within the next two months they would be looking to execute eighty percent of their proprietary trades via the medium of "smart trading" systems, whilst steadily building their "pure" algo offerings over the course of the next two years. There are however, some problems with adopting a smart auto-hedge system. If a simple algorithm is applied, operating on a trade-by-trade basis, then the bank will always be paying the bid-offer and frequently hedging unnecessarily. Instead, certain limits for trades need to be set, say, on a day-to-day basis and when trades exceed these limits then auto-hedging can come into play.

On the buy side, some hedge funds are looking to generate profit purely through the development and application of algorithms. Other players, whose sources of alpha are not derived directly from algorithmic techniques, are happy to purchase algorithms from sell side developers in order to reduce execution costs.

Any estimates of a particular banks' volume of algorithmic trading should only be interpreted with reservations. The sensitive nature of the information means that some banks dissimulate over the exact figures for algorithmic executions, exaggerating or understating volumes in order to disguise their position and to represent superior performance.

Drivers for growth

The primary factor fuelling the growth of algorithmic trading is a rise in electronic trading across all asset classes. The ability to execute trades electronically provides the mechanism by which algorithmic trades can function. This is particularly true of exchange traded products. However, increasingly instruments that were once purely OTC are moving towards a more formalised trading model, resembling that of an exchange. Government bonds for example, are heading towards an equity trading paradigm. Moves in the Foreign Exchange market, such as the creation of FXMarketSpace, are also indicative of this trend. The creation of electronic platforms alone, however, is insufficient for the large scale application of algorithmic trading; the listed options market is a good example of an electronic market that does not have enough concentrated product liquidity for algo trading to be particularly viable.

This rise in electronic trading has also led to a squeezing of margins. With an increase in market participants, spreads have tightened and traditional market making sources of revenue have declined for the sell-side. Thus, there is pressure on the investment banking industry to move towards a high volume of trades, facilitated by algorithmic execution, for an ever-growing client base in order to replace the revenue lost from providing a specialist service.

Declining margins also put pressure on banks to make cost-savings. Traders are expensive to employ, and if simple vanilla trades can easily be executed via algorithms, man hours and the related costs can be cut. Another area proving troubling for the human trader is the sheer breadth, depth and speed of the market. There are practical difficulties for the trader in even perceiving the rapid movements in prices that are commonplace in the highly liquid global markets. This has been further compounded by the multiplicity of directly competing execution venues and platforms with, increasingly, the same or fungible offerings that define the new global market environment. The numerous data feeds coming from these venues provide simply too much information to be processed by even the brightest human trader. The issue has been brought to the forefront, not just by declining margins from conventional trading, but also by regulatory pressure on banks to ensure best execution for their buy side clients. Compliance requires the digestion and interpretation of immense quantities of data, something that, if current trends continue, will only be possible via machine. However, the buy side uptake of algorithmic strategies remains, fundamentally, business driven. They desire a lower cost of execution and the ability to carry out large scale orders with minimal market impact.

A further problem facing the broker-dealers is the rise of the competitive threat from the hedge fund industry. While many of the leading sell-side players have seen the margins shrink to the thinness of a razor, a number of hedge funds have managed to extract money from these markets through the application of sophisticated algorithmic techniques. One source referred to a certain element of "greed and envy" felt by their bank over the profits that Citadel et al were squeezing from the markets. This has led the banks to focus much of their proprietary trading strategy on developing similar algorithms in order to capture and exploit pricing anomalies. A particular area in which two sources pointed to a definite lead for the hedge funds is in the sphere of cross asset trading. The ability for example, to spot arbitrage opportunities across bonds and swaps, or across different currencies was highlighted by two respondents.

However, the expansion of the mega-funds does not tell the whole story of banks' relationship with the hedge fund industry. While they may be direct competitors in the markets, hedge funds also provide considerable revenue streams via prime brokerage services. The supply of algorithmic trading solutions via prime brokerage was highlighted by one source as a major influence in the development of their algorithmic offering.

In addition to this, not all funds will succeed in the spectacular fashion of Citadel. A considerable portion of trading talent left its original home at the major investment banks to work in the new sector. However, there has been a steady flow back, as traders realise that the grass is not always greener on the other side of the hedge. This was cited by one source as a catalyst for further algorithmic development. These returning traders bring with them a considerable repertoire of knowledge from their experiences working for funds, which is then shared with, and absorbed by, their new employers.

What to look for in an algorithm

A complex algorithmic trading solution is, inherently, a multifaceted beast. There are a number of dimensions that come into play when one is being developed.

The headline grabbing advantage offered by algorithmic trading is the raw speed of execution. The ability to pin point and rapidly execute trades has seen a speed arms race between various banks as to who can develop the quickest algorithm. One reason for this may be that it is easier to use a quoted "speed" as a marketing tool, as opposed to the more intangible and ephemeral notion of what is a "quality algorithm". By far the biggest concern in the algorithmic trading space for the traditional sell-side is latency reduction. This is desired in order to guarantee Best Execution and to avoid possible reputational and regulatory risks that would, in future, come into play with the implementation of Reg NMS and MiFID.

However, the myopic focus upon speed has been questioned by some industry figures. Heavy investments have been made in both the equities and FX area in order to minimise external latency. Large investment banks have been installing their own hardware on exchange in order to cut out the middleman from their interaction. It could be argued that the industry has become excessively preoccupied by speed between the bank and the exchange and too little focus on the speed at which a trade can get "out of the house". Whilst reducing any external delays retains its importance, there is little purpose in achieving a three millisecond external latency if a risk calculation on a trade can be measured in seconds. Furthermore, the excessive reliance on two-tick and mean reversion arbitrage trades, as opposed to a variety of strategies, has come painfully unstuck when faced with volatile markets, as exemplified by the problems suffered by quant funds this summer.

Speed of execution will remain important, however many of the leading firms are now running into barriers in external latency reduction, in that they are now functioning at the same latency as the exchange itself. In this situation, and as more players scythe down their external latencies, the returns from reduced external latency diminish. Instead, their will be an increased focus upon internal speed reduction, and as these hurdles are removed, in the amount and calibre of data that can be drawn upon to formulate more sophisticated algorithms. Whilst the algorithmic arms race has been fought on the battlefield of latency, the terrain will soon shift to quantity and quality of data.

As well as ensuring reduced latency, another criterion for an algorithmic tool is fitting the right model to the right market. The flexibility and ease of modification that a particular model allows is important as it means they can be easily adapted and applied to new market conditions. This is important both in terms of market by asset class and market by region. In certain situations, for example, some pipes will need to be quicker in Europe than the US. In the current market place, it has been argued that the US markets are more open, with more points of dislocation and thus potential profit opportunities. This contrasts with Europe where, in the rates area, one source saw the market as tightly locked down. However, the threatened post-MiFID fragmentation of liquidity could provide an array of opportunities for the enterprising algorithmic trader, exploiting the late-adopters who have failed to develop the infrastructure required to deploy algorithmic strategies.

When applying algorithmic techniques, there are a number of "soft" issues that must come under consideration. One of the most important of these is "trust". Traders still need to have overall control of the process to avoid algo machines running off track. Any number of dystopian science fiction films have portrayed a situation where the advanced computational power of machines leads them to run awry from, and in some cases specifically opposite to, their original intentions. For this reason, it is always necessary for the traders to have some general understanding of the inner workings of a specific algorithm, even whilst the solution may remain a "black box". A useful analogy here is that of an autopilot on an aeroplane. The pilot is responsible for setting the overall flight path, which the autopilot then follows. Should anything untoward happen the pilot can use his expertise to step in. In the same way, the trader should set the overall strategy, which can then be executed via algorithm, with occasional interventions from the trader to guide the process towards profitability. In essence, this source saw the algorithm as a vehicle for the trader, rather than an autonomous alpha generator.

Two fronts of algorithmic development: Tactical alpha generation versus effective execution

Existing algorithmic development is being seen as a supplement to the skills of the trader, both in terms of handling flow trades and also, increasingly, as a way of generating excess proprietary returns. The spectacular success of the dedicated statistical arbitrage funds, exemplified by total earnings of over a billion dollars for quantitative hedge fund managers James Simmons and Kenneth C. Griffiths, has drawn attention to potential profit opportunities that can only be accessed via algorithmic methods.

Evidently, the weighting of algorithm development is dependent upon a banks prior business strategy and client base. If a bank sought to generate large returns through their proprietary book then, naturally, their focus would be on developing algorithms that could help both reinforce and grow their returns in this area. However, if the bank's revenue stream was dependent on the provision of execution services to the established buy-side, then this would provide the sounder foundation for algorithmic developments.

The traditional buy side is not necessarily the catalyst for this revolution in trading techniques. However, there is a definite awareness of, and increasing interest in, algorithms from the buy side. The perils of MiFID, and also the success of beta generating index funds, have forced a large portion of the major "old money" players to closely examine how their equity trades are executed. An increasingly sophisticated buy side means a need for banks to offer services in this area.

In the debt sphere, some bulge bracket institutions are dedicating all their resources to proprietary trading and alpha generation. The illiquidity and lack of electronic exchange platforms in Fixed Income further compounds difficulties in applying algorithmic techniques. Long only fixed income mutual funds may, at some point in the future look to algorithmic techniques. However, due to the present lack of a clear and liquid market for most types of bonds, there simply is not the level of demand to merit the dedication of significant resources to the development of client focused tools. It is worth mentioning that such a view is not unanimous within the industry, however. Some banks are in the process of developing their algo trading offering, evenly splitting their resources, in anticipation of an awakening of interest in algo techniques from the traditional buy side and a move towards a more exchange based trading environment.

Turning away from specific asset classes, development appears likely to follow a typical commoditisation cycle. Cutting-edge algorithmic developments will be focused, initially, on the proprietary area. However, as the market matures and, more importantly, new algorithms are developed, these algorithms could then be given away or sold on to clients. The ramping up of the hedge fund industry has meant that many sell side firms are looking to leverage their infrastructure in order to offer a variety of algorithmic techniques to smaller scale hedge funds. Many banks, in the debt sphere, are primarily looking to the alternative investment funds as their buy-side revenue source, and are hoping to be able to offer their algorithmic trading services through their ever expanding prime brokerage arms.

It is commonly acknowledged that many investment banks are playing catch up in generating proprietary trading revenue via algorithms, as the leading hedge funds have seized profit opportunities that most players were not even aware of. For many, it was not necessarily a case of leading the market, or that there was definite profit identified, but rather that they simply could not afford to be left behind. Much in the way that equity trading is a vital component in a "full service" investment bank even though it may not necessarily be particularly profitable, and may indeed be loss leading, algorithmic trading is rising in prominence because those who cannot prove their presence in the area risk losing clients across the board. Algorithmic trading provides another "hook" for clients, who can then be cross-sold the banks' other offerings as part of the banks' overall business strategy.

Market leaders

Speak to any figure within the industry, and the same name will crop up as the leader in developing advanced algorithms: Kenneth C. Griffin's powerhouse Citadel hedge fund. The ability for the fund to squeeze juicy margins from markets that had seemed as dry as the Sahara has captured the attention of the world's investment banks. There is now a race on between the leading broker-dealers to capture some of these returns before they are arbed out of the market.

When speaking of their peers, Goldman Sachs' formidable equity and futures algorithmic tools were rated highly by Lepus interviewees. Other mentions went to Morgan Stanley, Deutsche Bank and Credit Suisse for their equity offerings. It was the opinion of one buy side practitioner interviewed by Lepus that the major investment banking houses had the tools to draw up algorithms to a roughly equivalent standard. However, the recent period of increased volatility has called into question this interpretation, or at least the wisdom of how these tools were applied.

On the debt side, there seemed to be some confusion; no one was entirely sure of the actions of other players in the market, apart from the aforementioned Citadel. One interviewee commented that he suspected that Deutsche Bank would be an early mover amongst the bulge bracket sell side banks.

One thing to bear in mind during the course of this debate is the multitude of strategies, across multiple asset classes, employed by an array of market participants, with their own specific objectives, which can be grouped under the banner of "algo trading". Therefore, whilst Citadel et al may storm ahead in generating returns, squeezing dollars from millisecond arbitrage opportunities via algorithms, they are unlikely to ever develop the infrastructure required to run a full service investment bank. Furthermore, the scale of an investment banks' infrastructure and their access to flow data provides investment banks with another source of information by which they can further hone their algorithmic trading techniques.

Indeed, any such move risks diluting the core areas where hedge funds can retain a competitive edge, by retaining a lightweight, agile collection of systems and hardware that facilitates rapid switches in approach to exploit ever changing market conditions. Whilst investment banks are, naturally, looking to develop system architectures that are ever more efficient and well integrated, this does not mean they should look to jettison the potential advantages that such mature infrastructure resources allow.

By exploiting large reserves of data, for example, banks are capable of offering detailed research and analytic services to their buy side clients who may be unwilling, or unable, to develop such large reservoirs of knowledge. Furthermore, the banks can develop and parcel out their algorithms developed on the basis of this research to their buy side client base.

One should also be wary about generalising across the whole of the hedge fund universe. The term hedge fund denotes a myriad of funds that use an array of strategies. Whilst some funds have distinguished themselves through their use of algorithms, others have not been so successful. Apart from the aforementioned Citadel, the other top quantitative players are Renaissance Technologies and DE Shaw.

The hedge fund edge?

The leaders within the notoriously secretive hedge fund industry are in no hurry to share the specific details of their profitable strategies. However, discussion with sources within the industry reveals a number of areas where the leading players have an advantage:

- Specifically honed infrastructure
- Top class traders
- Minimal regulatory interference
- Higher risk tolerance

As one Lepus source within the hedge fund sector commented, many funds have an edge in the area because the whole weight of expenditure on technology is leveraged in order to squeeze the maximum from their algos. This can be illustrated by comparing the different market participants to different vehicles. The leading hedge funds can be seen as a sports car, with an all out dedication to speed. By contrast, the investment bank takes on the role of a family saloon. Whilst it may not be able to match a hedge fund, it is capable of carrying a number of passengers, their clients.

This dedication to speed is a result of the firms' strategic priorities. The leading hedge funds in the algorithmic space devote a spectacular proportion of their expenditure to technology issues. For example, around half of Citadel's two thousand staff are employed in the technology area. The founder of DE Shaw has a background in the development of supercomputing and moved from academia to head up JP Morgan's electronic trading wing only in the mid-80s.

Indeed, the background of many of the employees of the most successful hedge funds is far away from traditional Wall Street. Ren Tech, for example, recruits almost exclusively quantitatively based academic PhDs, with no background in conventional finance. By abstracting trading techniques from the world of finance, and applying techniques analogous with the sciences of physics, chemistry, engineering, computer science, fluid dynamics etc. the staff of Ren Tech have been able to break the trading paradigms of the investment bank, and discover market dislocations that were previously invisible.

The returns on these dislocations are amplified by the higher risk appetite, and lower degree of regulatory interference, of the leading quantitatively based hedge funds compared to their sell side counterparts. By placing faith in the ability of their quants and their automated algos to embrace risk effectively, the foremost hedge funds have been able to exploit ever more adventurous strategies. Furthermore, the hedge funds suffer a lower degree of regulatory intrusion than those on the traditional sell side, however, these players have also been following this trend. Morgan Stanley recently saw net revenue jump by 27% - a result that CEO and Chairman John Mack ascribed to an increased appetite for risk on the banks' proprietary trading desks.

Strategically, the advantages possessed by the top quantitative hedge funds are interrelated and self-reinforcing. The development of an agile infrastructure facilitates profit by allowing algorithms to act faster upon the market than competitors. The profit generated allows the hedge funds to attract top class employees, needed to keep their bleeding-edge technological platforms up to date. The combination of a swift and extremely accurate technological base with the raw brainpower of the quant enables the development of more sophisticated algorithms and the successful management of a higher degree of risk. This leads to further profit and the cycle continues.

However, in the recent period of market volatility the reliance upon pure quantitative modelling combined with a heavy risk appetite has backfired to a spectacular degree. Those funds that were previously touted as the luminaries of the algorithmic trading sphere have seen their dazzling gains turn into equally extravagant losses. The list of quant funds suffering in August rings like a roll call of those previously cited as “best-in-class”. For example, Goldman Sachs had nigh-on 30% wiped out from its Global Alpha, Global Equity Opportunities and North American Equities Funds. The losses stemmed from a heavy reliance on short term, high volume statistical arbitrage strategies, executed algorithmically. These losses were explained by Goldman Sachs as arising from 25-standard deviation events, several days in a row. It does not take a quantitative PHD to realise that a 25-standard deviation event works out as one that should occur roughly every 100,000 years and nor does it take one to realise to realise that the models underpinning such calculations are fundamentally unsound. Instead of making such assumptions over probabilities from incomplete, and possibly inaccurate, data, there is a need to concentrate on the capture, storage and analysis of far higher quality and far richer data to avoid the potentially devastating consequences of such suppositions.

In addition to this, even when the market returns to relative calm, as investment banks plough money into their algorithmic systems to chase this success, statistical arbitrage opportunities will decrease. In some markets, there are currently but one or two players deploying crawling algorithms, sniffing for dislocations. However, no technique in finance can remain exclusive for long. Once all the major broker-dealers, together with the leading hedge funds, have seen a new source of profit, they will leap into the market with similar techniques. Profit opportunities today will be ironed out in an increasingly competitive market. Market participants will need to leverage their experience, both in terms of personnel and the data their past algorithms have generated, looking for ever more esoteric opportunities for gain.

Data, Data Everywhere…

It has been oft-commented that the combination of new regulation and the interconnected rise in algorithmic trading has led to a profusion of market data. This explosion of data needs to be processed and ideally utilised as a potential profit avenue. Many banks will soon find themselves swimming in a sea of data, however without the platforms to exploit this information, firms may find themselves in a situation that can be likened to that of the Ancient Mariner: “Water, water everywhere/nor any drop to drink.”

Sourcing Data

The changing nature of the European exchange-based market provides a concrete illustration of the data challenges associated with the new trading environment. These currently produce approximately 600,000 messages per second, a massive increase on previous years. The 27th of February 2007 was a record for most markets and this number jumped to periods of 800,000 messages per second. Taking this as a base and, factoring in a continued rise in the number of trades executed algorithmically along with the expected impact of Reg NMS and Mifid, and it is easy to envisage a situation where the markets produce upwards of a million messages, even two million messages a second by 2008. The ability to store the four billion data ticks a year that this works out at will strain the current data architecture of even the most advanced bank. And these are merely the ticks and quotes for the underlying instruments. Add in the assorted collections of indices; news items; weather; and numerous other types of ever more abstruse time stamped data that people may wish to analyse and pattern match to this, we can quickly get to a point where we are talking about truly vast time series. This amount of data is not only beyond the boundaries of comprehension for the human trader, it is straining the limits of technological possibility.

Things are not about to get any easier, with market data volumes growing by between 100 and 120% per year currently. The implementation of MiFID will see this data flow explode by an estimated 300%. This growth is not just limited to the equities market. Options quotes have more than doubled in the last five years and, in the US the advent of penny quotes will see a growth in messages from an average of around 177,000 messages per second to upwards of 450,000. This eruption of data forms a self-reinforcing feedback loop, the more data that spews into a banks’ system the harder it is to be processed purely by a human trader, the greater the reliance upon algorithms, the greater the number of orders and so on. With this in mind, banks do not only need to build data infrastructures that can cope with today’s surfeit of market data, they need to future proof their systems to handle data that will be rising exponentially in volume year in, year out.

Strategic planning in order to cope with these data flows needs to look beyond simply average message per second measures and take into account spikes in the data flow, where the number of orders may be many multiples higher than the mean. Thus there needs to be enough spare capacity within the system to take into account such spikes. This is crucial because such spikes will typically be the result of an important market event fuelling turbulence. By capturing all the messages, banks will be well placed to develop their algos to cope with exceptional market situations.

These difficulties are further compounded by the need for rapid import of data from direct feeds. Traditionally, banks had sourced the market data information from a number of vendors, who aggregated information for an array of feeds. By buying data directly from the exchange banks can reduce the number of hops the feed has to take before it can be plugged directly into the banks technological applications. Direct feeds, however, bring with them a whole new set of difficulties. Directly acquiring data means that banks must iron out any imperfections in the information themselves. The vast number of sources, - there are forty-four European equity exchanges alone - systematic internaliser's dark liquidity pools, data vendors, multi-dealer platforms all need to be scoured for the best possible prices. Although it must be said that many firms are capturing and storing 'dirty data' as they need to model this along with 'clean or cleansed data' as they do not wish their algorithms to be inappropriately triggered by, for example, a "fat finger" error at the exchange.

If this were not complicated enough, each exchange will use its own data format and naming conventions. Feed handlers then are needed to normalise the data and consolidate multiple feeds into a single stream. From the feed handler comes the stream processor, which searches for relevant data, separating the useful from the dross. The desired data then is sent to an event processor that analyses the data even closer for specific data conditions that trigger "events," such as the creation of a quote, order or order cancellation. In addition, banks also need a high-speed messaging bus so that once the data is delivered it gets transported and processed quickly. This data not only needs to be fed into the banks system, in order to exploit it rapidly through algorithms in nigh-on real time fashion. In addition, Money Markets, Bonds, Repos, Options and Futures all have different quote conventions and it makes no sense to try to aggregate or massage them into a common format. Some appear to a practical solution is to rely on data tables with over a hundred columns for these differing datatypes, despite the obvious flaws with such a plan. Industry estimates sit around seven different convention standards, multiplying difficulties in data usage.

Back testing

Reference has made to the development of algorithmic trading through back testing algorithms on historical data. The FSA has insisted that all firms thoroughly test and model their algorithmic trading tools before applying them to the market so as not to expose their clients and shareholders, to undue risk. They have also pointed to the potential systematic risk issues of large scale algorithmic trading without sufficient attention paid to their back testing – although they have yet to clearly clarify what levels of back-testing need ideally to be performed. Thus, the ability of an algorithmic trading platform to back-test profile and allow the refinement of new strategies in advance of deployment is fundamental to a successful algorithmic trading strategy.

A source, from one of the world's largest hedge funds, commented that when looking to purchase algorithms from the leading houses he expected them to draw upon around five-years of pristine tick data to ensure their reliability. Those who have the capability to guarantee this will have a distinct advantage both in terms of developing cutting-edge proprietary algorithms and also in the primer brokerage sphere going forward.

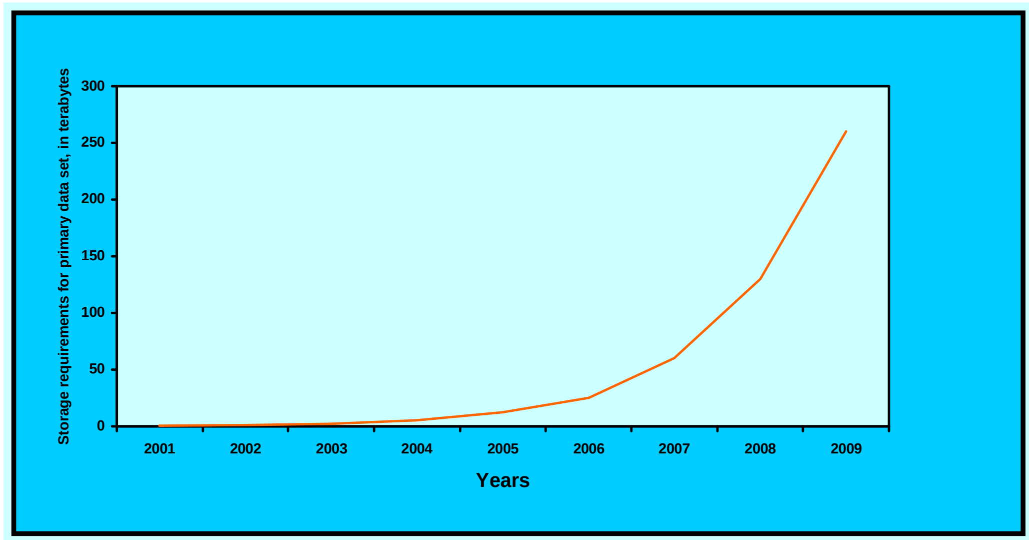
This view was confirmed by one Lepus interviewee. The source claimed that accurate back testing was an extremely high priority for their organisation. However, currently they lack sufficient quantity of accurate tick by tick data, let alone quote data, to formulate their algorithms. This is a legacy of the profligacy of exchanges, market data providers and end users in their former treatment of market data: Much historic data is limited to approximately ten years of trade data, with longer histories consisting of only closing prices, or at best openings, highs and lows. The ability to carry out regression analysis on rich historical data is key to further progress in the area. The source further opined that the lack of reliable reference data was a problem that was too often ignored in the rush towards real time data feeds. Without sufficient attention to reference data issues a banks' algorithmic infrastructure will grind together like unoiled gears.

However, the conventional view of back data has been challenged by some. One source, speaking from the debt side, remarked that his tier-1 bank was adopting a forward looking, pragmatic approach in the algorithmic sphere. Their primary focus was very much observing day to day flows and looking to iron out any imperfections as their algorithm development progressed. For this, they carried out real-time on the spot testing based on short-term historical back data. However, this was qualified by the relative immaturity of the algo trading market. There would come a time, in this sources' opinion, when more back testing would become necessary. As algorithms advance, banks will need to search harder and harder for the scraps of profit they can take, and long term historical data will be rendered more interpretively useful as it will correspond with a more mature algorithmically traded market. There was no point, according to this source in extensively, and expensively, back testing algorithms when the market terrain could shift under their feet. Thus, whilst not relying on long term historical data currently, data capture remained extremely important. The source described the data requirements associated with the rise of algorithmic trading as "huge". The bank was seeking to capture "all" data possibly available to it, from as many feeds as possible. These data requirements encompassed a broad range of factors, from transaction data through to specific news events.

Data capture was not the only area of importance, however. It is vital that multiple data feeds can be filtered through to a centralised storage site, enabling the information to be leveraged with ease across the bank and not simply by a specific business silo. As banks move towards a centralised data storage system, however, they need to ensure that they retain flexibility and scalability. Different functions within the bank have very different requirements from their data. In the rush towards centralisation, banks must ensure that data retains maximum utility across the enterprise. This also creates huge, and often unanswered concerns, as regards Business Continuity Planning (BCP) and Disaster Recovery (DR) in a realtime dynamic environment with multiple updates occurring in multiple sites simultaneously.

Discussion of this data issues concerning algorithmic trading frequently focused upon the necessity of capturing and storing trade data. However, the importance of quote data should not be neglected, particularly in a transformed regulatory environment with strenuous best execution requirements and an increasing focus upon liquidity processing. All banks involved in the algorithmic arena are thirsting for a greater breadth and depth of data of all types. The precision required in developing the best algorithms in a fiercely competitive market, necessitates the capture and storage of, simply, as large an amount of data as possible. Industry estimates suggest that, for its primary data set alone, a globally operating investment bank storing all the information available to it, will have to have stored approximately twenty five terabytes of market data since decimalisation. With replications included this leaps to some 200 terabytes. Following the explosions of MiFID and Reg NMS and the rise of algorithmic trading, the data storage issues are multiplied.

Figure 1: Data storage requirements for primary data set of a global investment bank, past and projected.



Source: Lepus

Whilst recognising the need for as much data possible, it is still of the utmost importance that this data remains comprehensible. Another recent development is combining back testing and graphical representation. The concept is to place all of banks complex strategies for trading into the system, and then observe how they would have played out over, say, the past thirty-day trading cycle. The results of these surveys are then represented graphically in the manner of an equaliser, i.e. when an algorithm was performing badly there would be a red light, while profitability could be indicated in green. This ease of interpretability facilitates the rapid modification of strategies, providing a bank's internal infrastructure is capable of handling the processing challenge.

On a more sophisticated level, the leading quants are looking to synthesise biological and mathematical science in order to produce Genetic Algorithms (GA). Genetic algorithms work as a computer simulation analogous to evolution, whereby a number of competing solutions (in this case, algorithmic strategies) are placed in a context where they must compete for scarce resources (profit) in order to reproduce. At the end of the simulation, the evolutionary process should provide a best-fit algorithm to the specific market scenario.

Back data requirements

During the course of the research, Lepus also asked sources to provide figures on their requirements for the length of back data they thought was required to provide a basis for the development of algorithmic trading strategies. The results are summarised below:

Figure 2: Back data requirements for the development of algo trading

Questions	Bank A –Tier-2 European (Rates Area)	Bank B – Tier-2 (Rates Area)	Bank C – Tier-1 European (Rates Area)	Bank D – Tier 1 European (Equities Area)
What is the duration of back data you are currently using to develop your algo trading?	Not currently storing/holding any data.	Not currently storing/holding any data.	90 days	This is very much strategy dependent. If you're looking at secondary harmonics – just half an hour will do, in other areas could be looking at 3 years plus
What duration of back data, in your opinion, is ideally required for algo trading?	365 days+	180 days	365 days +	365 days+ (for some strategies)
How do you anticipate your data requirements (in terms of time of the duration of the back data required) will change in the algo trading space in the next 2 years?	Begin to store market and internal pricing data in order to refine algorithmic strategies	Begin to store more and more data	The more data the better, and the longer periods they hold data for the better	Longer data specifically related to algo trades in order to develop heuristic systems. Need data to develop strategies that can spot discontinuities so algos do not backfire

Source: Lepus

It is important to qualify these responses with an awareness that, on the fixed income side, algorithmic developments are behind other asset classes. When building an algorithmic platform to compete in either the equities side or in FX the requirements are considerably more stringent.

On the equities side, data requirements are very much strategy dependent. For some algorithms, less than a day of data is necessary, for others it can be three years or more is required. The major problem is trying to develop heuristic algorithms so in the event of a discontinuity the algo stops trading and can be recalibrated before it loses a lot of money.

For rates, the market players are building their systems as we speak, so are not currently holding much data. However, as the market matures, leading firms will be looking to pull on 365 days or more of back data for testing.

The prevailing view in the industry is that the more data that can be captured stored and rapidly accessed the better. However, there are practical problems associated with this as conducting detailed analytics on rich data over a vast time span absorbs considerable computational power. The resource consumption required can be off putting to some institutions, who may perhaps feel a more pragmatic solution is available to them. Furthermore, the processing time required for such detailed testing can delay the application of the algorithm to market, meaning profit opportunities escape. Those who will win the algorithmic battle will be the banks that can rapidly carry out testing, whilst at the same time drawing on a richly detailed store of back data.

Existing data infrastructure and the battle with latency

The most important issue for data infrastructure cited by respondents was a ramping up of the speed, with both internal latency and calculation timescales seen as too long to fully capitalise on the opportunities that could be available through algorithmic trading. Here the hedge funds have a definite advantage with their small scale infrastructures, tailored specifically to the needs of algorithmic trading solutions.

For the bulge bracket investment banks, their traditional infrastructure can be considered relatively fast; calculating, sending and receiving data for vanilla products in around ten milliseconds. However, any speed improvements they want to make will be closely examined. Large scale infrastructure overhauls often have unintended consequences, with the suboptimal implementation of one component rebounding negatively upon the overall architecture of a bank's IT process. Existing data infrastructure may well be product dependent. One source commented that their pricing engines were up to scratch, however they were running into definite problems on the market data repository and derivative processing front. The challenge, for this bank, was to keep the rapid speed of the infrastructure whilst also ramping up volumes to meet the challenges of shrinking margins.

It can be argued that the existing data infrastructure of a large scale investment bank is both a burden and a boon to banks' algorithmic efforts. Whilst hedge funds can build their systems devoted to a particular strategy, or focused set of strategies, the generalist needs of the investment bank will always require a more robust architecture. Despite this, the generalist infrastructure provides numerous business opportunities that hedge funds will never be able to avail themselves. For example, banks can leverage their infrastructure to facilitate outsourced risk management and analytics offerings to their clients.

How to juggle the competing priorities of keeping an agile, low latency infrastructure, whilst also being able to leverage large quantities of data and offer a diverse suite of offerings to clients, is a central question for the banks' senior management. Lepus polled a number of institutions on specific latency requirements as regards the development of algorithmic trading techniques. There results are displayed below:

Figure 3: Internal Data Latency Requirements

Questions	Bank A –Tier-2 European (Rates Area)	Bank B – Tier-2 (Rates Area)	Bank C – Tier-1 European (Rates Area)	Bank D – Tier 1 European (Equities Area)
What is your current data latency in the algo trading space?	N/A	N/A	Between 10 and 90 ms. Would not be any more specific due to the sensitive nature of the information.	10 ms at the moment, leading players are orders of magnitude faster.
What is your ideal data latency requirement in the algo trading space?	10 ms for now	90 milliseconds	10ms	Sub 10 ms – although this is very much strategy dependent
How do you anticipate your data latency requirement will change in the next 2 years?	Latency will have to decrease significantly	Latency will have to decrease	Will definitely go down further in coming years – witness FX space, now talking in timeframes of fractions of milliseconds for latencies	Obviously, the lower the better
Do you have a centralised or multiple databases for algo trading?	N/A	N/A	Multiple, one for each asset class. Looking to integrate to facilitate cross-asset algos	Multiple – Equities, FX

Source: Lepus

When looking across the financial world as a whole, the internal data latency requirements for algo trading are fairly easily stratified:

1) Leading Hedge Funds and some time sensitive strategies employed by leading investment banks can be measured in fractions of milliseconds.

2) Standard strategies employed by top broker/dealers banks demand latencies of lower than 10 milliseconds.

3) Lagging slightly behind are those full service banks that do not derive their revenue primarily from trading who are looking at a 10 milliseconds benchmark figure.

There can, however, be no constants as regards the specific requirements for a technological infrastructure. The competition is to develop ever faster and leaner technological infrastructures, in order to give the quants talent within a bank or hedge fund the best technological background on which to develop money-making algorithms. As the number of market participants who are able to access the market at low latency increases, the previously developed algos will cease to be effective, as dislocations are ironed out. Thus, the algos advance at the same time as the technological backbone, and the development of this technology by competitors fuels the necessity to decrease latencies to capture ever-more minute arbitrage opportunities. It cannot simply be stated whether technology is keeping apace with the desires of the quants, because without the technology for low latency calculation they may not even be aware of the possibility for the application of certain algorithmic strategies.

As time goes on, and latency ceases to be the be-all and end-all of the algorithmic debate, resources will be increasingly devoted to more long term trend finding algorithms, looking for correlations and covariance across asset-classes and time spans, looking for new avenues for profit as millisecond dislocations are arbed out.

The costs of entering this race are prohibitive for certain institutions. The need to maintain a nigh-on pristine tick history, to employ teams of advanced quants to work with this data, whilst at the same time retaining an agile technological infrastructure, has limited the development of cutting edge algorithms to either the top flight investment banks or to specifically dedicated funds. However, there are a number of vendors in this space who are leveraging their specific development expertise in order to offer platforms that provide a development infrastructure at a considerably reduced cost, whilst retaining performance.

Leveraging compliance for competitive advantage

The combination of the implementation of MiFID and Reg NMS with the rise of algorithmic trading means that more data than ever is available to investment banks. Whilst client offerings are one way to leverage this data, there is another, namely to garner a clearer picture of the market place as a whole. The leading bulge bracket banks have plans to make use of the data required under MiFID and Reg NMS come implementation day. The terabytes of data that this will produce can be divided into three broad categories:

- Customer behaviour style – here banks are looking at what customers made or lost on trades and the hedging strategies that they applied.
- Market behaviour – observing how the market responded to thick or thin liquidity situations, changes in volatility and monitoring movement at certain times.
- Customer market data – observing how market conditions may change customer behaviour in order to be ahead of the market.

The data necessitated by MiFID is seen as something of a boon in this area as it will ease the production of analytics. Others have argued that much of the information was already readily available, for example through TRACE in the USA, and this could be useful to monitor flows. This underestimates the sheer volume of data required by MiFID, and its analytic utility. Gathering this market data, however, poses a whole new set of challenges. The potential fragmentation of liquidity resultant from MiFID means that data will need to be aggregated from multiple execution venues. This, when combined with the large volumes traded via algorithm, poses a significant burden on banks' existing processes. Once banks build or buy their infrastructure up to a sufficient standard to cope with the ever-rising data tide, they will have formidable resources which they can leverage to produce ever more accurate and advanced algorithms.

Indeed, by observing the internal data flows produced by client activity banks can anticipate big trend movements that can then be exploited algorithmically, as opposed to simply concentrating on latency driven stat arb strategies.

By combining instrument and customer data, banks can hone their algorithms to understand the interplay between asset class, buyers, sellers within particular contexts, and develop algorithms to work within these specific conditions.

The Data Challenge Illustrated: Risk and Algorithmic Trading

Integration

The idea of integration and consolidation of platforms has been frequently seen as a panacea to internal latency issues, whilst also facilitating a more efficient utilisation of resources across the enterprise. The importance of such efficiencies was reinforced by the opinion of a leading figure interviewed by Lepus. This source commented that the bank was devoting considerable resources to the development of a cross-product risk platform to facilitate the exploitation of cross-product trading opportunities. The second priority, on the debt side, was to implement risk updates every thirty seconds. The current trading situation on the debt side does not necessitate the investment of resources to ensure real time data feeds. One source argued that the most effective way to cope with the risk issues resulting from algorithmic trading was to run risk calculations from individual traders' spreadsheets, which would be connected to a centralised model library. The robustness and flexibility of spreadsheets render them easily modifiable, thus facilitating the rapid overhauling of a particular model or algorithm come a market shift. This ease of amendment, however, brings with it a concomitant set of risks. Changes to a particular spreadsheet can have ripple effects through out an enterprise, particularly one relying on a centralised data storage system. The consequences of these changes are amplified in an area as data sensitive as algorithmic trading. Thus, any bank relying so heavily on spreadsheets, and allowing modification "on the ground", will need to ensure they have a thorough set of controls in place to avoid negative impacts on their operation as a whole.

Real-time risk

Spreadsheets alone, however, will be insufficient to cope with the analytic work load required to run numerous real-time risk calculation. The pressure of speed as relates to the development of algorithmic trading has already been mentioned in this report. There is a constant demand upon banks to produce solutions at exponentially quicker rate, something only exacerbated by the development of algorithmic trading. With markets in a state of constant flux, and arbitrage opportunities vanishing no sooner than they have arrived, there is an increasing focus on attempting to model risk "on the fly".

As was highlighted by a source from a leading hedge fund, the combination of real-time streaming analytics with low latency electronic trading is the "holy grail" for trading. The faster risk can be analysed the faster algorithms can be applied to the market, and the more chance of gaining an edge over competitors and thus generating profit. However, with the avalanche of market data descending upon the investment banking community, the traditional in-memory databases used for on-the-spot modelling will soon be insufficient to cope with the demands of rapid fire, yet highly precise, analytics.

Further to this, algorithmic trading places a whole new assortment of challenges on risk management. Firms are moving towards an increasingly advanced suite of strategies, specifically tailored for certain circumstances, a process that is only likely to intensify in the future. There will be times when some algorithms are more active than others dependent upon market conditions that can alter by the minute. The traditional method of setting limit exposures shared between trading desks will mean that the full capability of a firm's algorithmic suite cannot be brought to bear on the market at one point. The ability to manage risk in real-time allows a more flexible approach to algorithmic deployment and development.

However, the importance of speedy ad-hoc modelling is very much dependent on the product area and model involved. The key issue in many markets is rapid fire decision making. Thus many firms are looking at discretely running risk initiatives, breaking down the risk so it is not assessed on a portfolio level. The overall risk assessments for the banks' entire portfolio can still be run via End of Day (EOD) overnight batch processed risk management; however in twelve to eighteen months time, as the market matures, the ability to carry out this kind of risk modelling on the fly will become increasingly relevant.

In order to facilitate this real time risk management in the future, it will be vital to corral and utilise risk data correctly. Data challenges are the "key" area for implementing real time risk management. The most important issue here is the streamlining of existing systems in order for them to function at sufficient speed to take advantage of cross product anomalies, which may disappear in a matter of milliseconds. All the banks questioned were looking to restructure and centralise their market data storage to facilitate ease of access by both risk management and the Front Office.

Another data challenge is the complexity of models required for algo trading and the struggle to keep up with this constantly moving area of business. However, when the risk management of algo trading is applied to more vanilla products, apart from the need to increase volumes, the data challenge recedes into the distance.

Further to this, many banks still operate with an artificially imposed conception of "closing prices" – failing to fully appreciate, and exploit, the globalised and continuous nature of many markets. There needs to be an awareness gained of the subtlety and flow of intraday movements across the board, and an acknowledgement that "money never sleeps". The advantages of a move to real time risk are not simply found in increased speed but also in a richer and more detailed understanding of the subtleties of everyday market movements.

The problem facing risk managers is that the demands of both clients through prime brokerage and the banks' own quants will increasingly be looking for on-the-spot risk calculations as a basis for applying their algorithms to market. The waves of data that will wash through the banks in the coming years *can* be exploited, and those who look to move away from risk management based upon close prices and towards real time will reap significant rewards, whilst others will find themselves marooned in the rapidly changing algorithmic trading environment.

Infrastructure overhaul

As exemplified by the issues surrounding risk, an environment dominated by algorithmic trading places a whole new set of infrastructure challenges on the investment banking industry. There are ever expanding requirements for more data, richer data, more accurate data all of which needs to be rapidly accessed. If a bank seeks to compete in the algorithmic age they need to align their overall infrastructure to meet these new demands across all functions.

Into the Future

The death of the trader

The rise of algo trading will allow banks to pare away the dead wood amongst the trading staff. However, the brightest traders will always find their way to the money. There remains a need for experts to devise new and more sophisticated algorithms to exploit market opportunities. The advantage of algorithms is that they can take traders away from the more menial, simple tasks, rendering the overall operation more efficient. This will also free up time for the "stars" of the trading floor to devote themselves to seeking out ever more esoteric routes to profit.

Going forward, banks will look increasingly to distinguish between their automatically executed algorithms and those that provide alerts to traders, alerting them to patterns and correlations within market data, and then requiring human intervention to allow the algorithms to interact with the market. Currently, the prevalent focus upon speed and latency reduction has meant that many algos are based around statistical arbitrage and executed in a timeframe measuring in the milliseconds. However, as the major houses adopt these techniques, such dislocations will be ironed out of the market. Instead of just speed, banks will also need to focus upon developing ever more complex algos based upon a high quality, in depth data set. Such algorithms are likely to have multiple stages of execution, not the one or two tick timeframes of most stat arb algos, and consequently will carry substantially higher operational risk burden, not only allowing but necessitating the supervision of a skilled human trader.

The human trader is a necessity; an experienced trader will have a feeling for market sentiment, he will be able to appreciate the "colour" of the market, foreseeing its volatile and unpredictable movements in times of stress. We can also envisage the role of the human trader as managing by exception, seeing where the machines may be going wrong and guiding them back on path. The trader can be left to devise the overall strategy in place, which could subsequently be executed via algorithms.

Another role for which the human trader is uniquely suited is the structuring of complex, at times loss making, deals for particular clients. Whilst the bank may not make money from such deals, they may need to offer their clients' solutions in order to encourage further business from the client that will actually generate revenue. A computer will not be particularly suited to the demands of calculating such a trade. And, unlike a human, a computer cannot articulate or interpret a bank's overall business strategy. Whilst algorithms may advance, they will remain tools to be utilised rather than completely replacing the human trader.

Complex Event Processing

Complex event processing (CEP) refers to the ability for computers to handle and react to data from incidents that have a multiplicity of effects. First developed in The Academy by a mixture of computer programmers and advanced quant PhDs, the idea is finding increasing traction when applied practically to the financial markets.

Theoretically, this should allow the development of sophisticated algorithms that can exploit news feeds, detect fraud and manage risk in a nigh-on real time fashion. However, the current clamour is something of a case of “all mouth and no trousers”. Whilst CEP may be oft discussed, there is little actual application on the ground. Across the industry, views over the application of CEP remain decidedly mixed. Whilst it was dismissed in some areas, Lepus discovered that several tier-1 firms employed it for high volume, real time pattern matching.

One area that is being looked at, however, was the reaction to complex events such as market shocks. The unpredictable volatility that could result from such events means that the “standard” algorithms the banks had in place would be of limited, or even negative, value. A source opined that it was possible that, in this situation, CEP could prove extremely useful. Indeed, this does illustrate one flaw within the algorithmic sphere. Their development corresponds with a period of low volatility almost unmatched in the history of the financial markets. Should a market shock transform this situation, as has occurred in the wake of the subprime crisis in the US, all the algorithms developed on the basis of this data would be rendered useless. Furthermore, there is a great deal of overlap employed in the algorithmic strategies of the major players, meaning that their losses were magnified. Thus, it was thought important to look towards developing algorithms that could anticipate price movements in extraordinary market conditions, not just in terms of volatility, but also in the case of a liquidity squeeze - be it isolated to a specific market or occurring across asset class. However, there are definite difficulties in programming algorithms to cope with these challenges as historical market data can never be a fully accurate predictor of how the market will respond to new shocks. By storing a lengthy quantity of historical data, including past market shocks, and analysing this in tandem with quasi real time data feeds banks will be able to ride out, and even profit from, times of stress.

One way of coping with periods of stress is through the development of heuristic algorithms. These algorithms could be developed by making reference to historical back data, then deploying a recalibrated best-fit model to a market in a stressed situation in order to rapidly take advantaged of changed market conditions.

For most players, as it stands, current algorithmic development, aimed at squeezing dollars from market dislocations and arbitrage opportunities, is sufficient. Launching into CEP without a comprehensive algorithmic toolkit, could be seen as analogous to attempting to run before you can walk. Until there are fifteen major players deploying crawling algorithms to squeeze out the margins, resources assigned to CEP would be better deployed elsewhere. To draw a comparison directly from the investment banking world, the profit opportunities would be similar to the process of a product from exotic through complex to vanilla. Whilst a product is new, large margins can be extracted, but as it becomes increasingly accepted and competition becomes more intense and margins disappear; the sell side then moves on to the next, ever evolving, offering.

One area that could prove important in the near future is the automation of trading off news data via algorithm. With the ubiquity traditional pricing data sets, operators are striving to quantify news data in order to find another source of alpha. Although the data challenges associated with algorithmic trading are formidable, the number of market participants engaged in building systems means that traditional arbitrage opportunities may soon be exhausted. Theoretically, the rapidity with which an algorithm can process news data should allow it to respond more quickly than any human trader, gaining an edge of only a few milliseconds over the rest of the market, which would be enough to guarantee substantial gains. The opportunities for such trading are amplified in world of live global news and information feeds and where a political or economic development in one country can ripple out across a variety of markets.

Indeed, the move towards automated trading off news may come sooner than expected. One source, speaking from a hedge fund perspective, commented that a number of his peers were already developing algorithms to take advantage of news stories. During the course of producing this report Reuters launched a system designed to quantify the sentiment within news stories in order for the data to be utilised by algorithmic trading. The system works by assigning numerical "sentiment scores" to words or phrases which are then processed to give an overall positive, neutral or negative score to the company featuring in the news article. These scores are added together to calculate the current sentiment for a company, a sector, an index or, indeed, to quantify the sentiment of the global market. This builds on the two products launched during December 2006 aimed at directly importing economic statistics and company earnings announcements into quantitative algorithms. Reuters are not the only company making an offering in this space. As of March 2007, Dow Jones launched their “Elementized News Feed“. This product aims to deliver news stories on corporate or economic issues via XML so as to be easily interpreted by a computer. Although the vendors are keeping details of their client base close to their chests, it has been strongly implied that a number of leading banks’ proprietary trading desks and several hedge funds have been amongst the early adopters.

Building a suite of complementary, cooperative algorithms

Another area for the future is the simultaneous application of a variety of algorithms that operate cooperatively. By way of a simple example, a bank may look to employ a strategy that uses historical analysis for volume-based slicing, but simultaneously take an FX position to hedge against currency risk.

Plans for cooperative algorithms remain in their infancy. Some progress has been made on the equities side. However, for debt, the leading players are looking to combine two to three algorithms as a base, on which they will then build. One source argued that any further work in this area was very much market dependent. If more competitors began utilising algorithm suites for a specific market, then banks would be forced to throw resources into the development of an ever more complex toolset of algorithms designed to function in harmony. For those in the second strata of algorithmic trading, the development of complementary suites remains only in the realm of conjecture. It is first necessary to develop a comprehensive algorithmic toolkit to serve the bank, and clients’ needs, before attempting to find synergies between tools. One can, question, however, whether the terminology of a “suite” of algorithms is, in fact, particularly meaningful. Instead a collection of algorithms serving one trade could merely be seen as one, albeit complex, algorithm.

Algorithmic development as a strategic priority

The clamour in the financial press as regards algorithmic trading can often lead to a misplaced sense of the actual situation on the ground. Whilst various hedge funds may have been shovelling in the profits via algorithmic trading it is only of late that the major sell side players have pushed algorithmic trading up their agenda, particularly outside the equities sphere. In the past six to twelve months, applying algo trading in the rates area has only been a topic of idle discussion. It was mentioned in passing meetings, but actual action was limited. However, now there is a definite awareness of the importance of developing algorithmic techniques. Work is now progressing apace, with the leading banks senior management realising they need to react to a transformed trading environment. Amongst the second strata of banks, long overdue work on algorithmic trading is now taking place as well, although a realisation of the full strategic importance of algorithmic techniques remains lacking.

There appears to be general dissatisfaction from those charged with implementing algorithmic techniques with the failure of business planners to fully appreciate the shifting paradigm for the trading world, and whilst, on the equities side, there had been considerable work, the application of algorithmic tools across the other asset classes had not advanced as much as it could.

However, awareness of the new trading paradigm is dawning amongst senior figures. Budget allocations are expanding faster than any other area of investment banking, as leading players realise that they are faced with a transformed business environment. Industry estimates for spending on algorithmic trading vary from some $700mm to just shy of two billion dollars for 2007. Whilst sources within banks keep their actual spending figures close to their chests, most stated that this expenditure would have to grow by 20 to 50% year-on-year in order to remain competitive. If anyone were in any doubt of the strategic importance of algo trading, Citigroup's $680mm acquisition of Automated Trading Desk, a company devoted to high speed automated executions, in July 2007 should dispel them.

One definite strategic priority is the development of techniques to take advantage of the opportunities offered by cross-asset trading. However, before this can be accomplished there needs to be substantial work carried out to bring the standard of algorithms in other asset classes up to the same standard as those on the equity side. This problem is aggravated by the inability to port algorithms directly across from the equities sphere over to other instruments. Each market has its own specific demands. For example, many of the equity algorithms have been devised to crawl for liquidity, something that is not as much of an issue in either the FX or government bond markets.

Looking further into the future, banks will be seeking to utilise their e portal technology as an access point for offering algorithms to clients. One bank Lepus surveyed stated that they were pushing their e portal as a cross product centralised access point for all of their clients. The more ambitious players in this space have been looking to ape Amazon, devising sophisticated customer analytics that can automatically hone the offering of algorithms to clients based on their past trading habits.

Conclusion

Quant driven algorithmic trading is now seen as a “must have” toolset in any bank that wishes to be seen as a major player in the broker/dealer market. Although realisation may have been slow to dawn, many firms are seeking to apply algorithmic techniques across the full spectrum of conventional asset classes.

However, whilst work may be afoot across all asset classes, it is still in the equities sphere that algorithmic trading has seen the greatest level of acceptance. Lepus research has revealed that across the market, between thirty and forty percent of equities traded by the leading broker/dealers are traded via algorithmic techniques. On the debt side, the level of trades carried out via algorithmic trades lags significantly. However, the vast majority of those within the industry envisage a significantly greater uptake over the next twelve to eighteen months across all asset classes.

This Executive Summary has attempted to highlight some of the key issues both in the current market and future market. The major demands are for speedy risk modelling, effective back data capture and utilisation, and reduction of internal latencies in order to capitalise on ephemeral market dislocations.

All houses are looking to have an algorithmic trading platform in place that can deal with these major pain points. Speed-out-of-shop is a fundamental requirement that has until now recently received insufficient attention. Indeed, many of the leading banks could be accused of external latency monomania, as the need to achieve best execution under MiFID and Reg NMS has consumed all attention. It is difficult to estimate exact figures for the speed requirements out-of-shop, as the goalposts are moving constantly. In effect, banks find themselves in an arms race, similar to the one being fought over external latency, as the pressure of maintaining competitive advantage drives ever faster execution. As external latencies and internal latencies decline, increasing battles will be fought over the quality and quantity of data that banks can store.

Banks are looking to capture and store more of the back data from their trades in order to hone more precise and efficient algorithms. The data explosion as result of the interplay between regulatory moves, business pressures and the application of more and more algorithmic techniques is creating a feedback look, feeding the demand for algorithmic trading and straining the infrastructures of all banks. The specific data requirements remain strategy dependent. However, the more data that can be stored, and the quicker this data can be accessed, the easier the development of more efficient algorithms. Thus, banks will look to the most efficient data warehousing solutions in order to ensure an in-depth store of high quality data that can be rapidly leveraged to develop new techniques required by market developments.

Synthesising the issues of back data and reducing speed out-of-shop, banks are looking to develop algorithms that can cope with market shocks as rapidly as possible. The development of algorithmic trading has taken place in an extremely low volatility environment. These algorithms have not been tested against a high volatility, or low liquidity, shock scenario, and when confronted with this situation have lost money heavily. There remains a risk management requirement to quickly alert traders of a malfunctioning algorithm so a machine can be recalibrated as rapidly as possible minimising down time and profit loss. One way of doing this is to utilise market data on past algorithmic performance in order to develop heuristic algorithms.

In this new environment the role of the trader will change. Instead of dealing with a stream of simple flow orders, traders will be left to develop new money making techniques whilst supervising the function of an array of algorithms. Regardless of progress, or lack there of made to date in regards to adopting algorithmic trading, it is clear that all banks will need to develop a robust and flexible platform for their algorithmic strategies that can both store and exploit large quantities of data, whilst retaining the speed to allow rapid readjustment and nigh-on real time risk management.