

Researched and written by

A-TEAMGROUP

BACK TO THE FUTURE

Historical Data in High-Frequency Trading

June 2009

An industry briefing prepared
by **A-Team Group** for



London
Stock Exchange



Executive Summary

- Financial markets volatility, exacerbated by low trading volumes, is challenging algorithmic and quantitative trading models, making it imperative for developers to make use of the highest-quality historical data as they create, test and release new trading strategies.
- The wide scale acceptance of algorithmic trading strategies is driving order and tick volumes, as client orders are sliced into so-called 'child orders', creating more and more incremental data points for models to analyze and execute on.
- Meanwhile, more stringent regulations in the wake of the Credit Crunch is forcing increased scrutiny of past trading activities, with specific regulations requiring firms to be able to recreate the precise market conditions at the time of any particular trade, a major data challenge.
- This combination of market conditions and regulatory climate is making it imperative for trading firms to apply the highest quality historical data available, both to their models and to their compliance processes.
- Quantitative models, in particular, require every tick in order to take into account true market conditions, and are less effective when driven by 'conflated' tick data, whereby consecutive orders at the same price are combined into a single update, or tick.
- Historical order book data can be used to rebuild order books at the time of any given trade, allowing compliance departments to address regulators' queries, particularly around best execution.
- Market practitioners are increasingly turning to the original source of historical data as they seek to meet the more onerous requirements posed by current market and regulatory conditions.

The London Stock Exchange offers comprehensive, officially sourced data services for both tick and order book historical data. To learn more, contact:

Ian Wood

+44 (0)20 7797 3827

iwood@londonstockexchange.com

www.londonstockexchange.com/tradedata

Introduction

The ongoing financial markets volatility – sparked by the highly public failure of several major financial institutions and exacerbated by recent low trading volumes – is testing brokers' algorithmic and quantitative trading models to the max.

As prices have spiked and troughed, the application of historical tick and order book data to models have become fraught, making it imperative for developers to use the highest-quality historical data as they build innovative trading models. At the same time, new regulations are forcing institutions to recreate history, by storing huge quantities of trade and quote data in order to prove best execution according to current market conditions at the time of the trade.

Both developments have converged to underscore the importance of data quality in the development and testing of trading strategies, and creation of audit trails. And achieving a core level of data quality is, in itself, being challenged by the ever-growing volumes, particularly in orders, as algorithmic trading models slice customer orders into much smaller and manifold 'child orders', each generating order data.

Historical tick data, in the form of latest-trade (print or stamp) and order data, is emerging as a key enabler for trading firms as they compete to differentiate the services they offer to clients. While real-time market data is widely held, and remains, essential to the ability to compete into today's fast markets, firms have become reliant on the trading strategies and models they use in order to identify market opportunities or execute client orders to the letter.

The shift to the mainstream of algorithmic trading techniques, in particular, has upped the ante with respect to the historical data requirement. By significantly boosting the volume of order data, algorithmic trading has driven demand for more granular data. The trend toward algorithmic trading has also increased the number of trades, as child orders are executed (See graph, below).

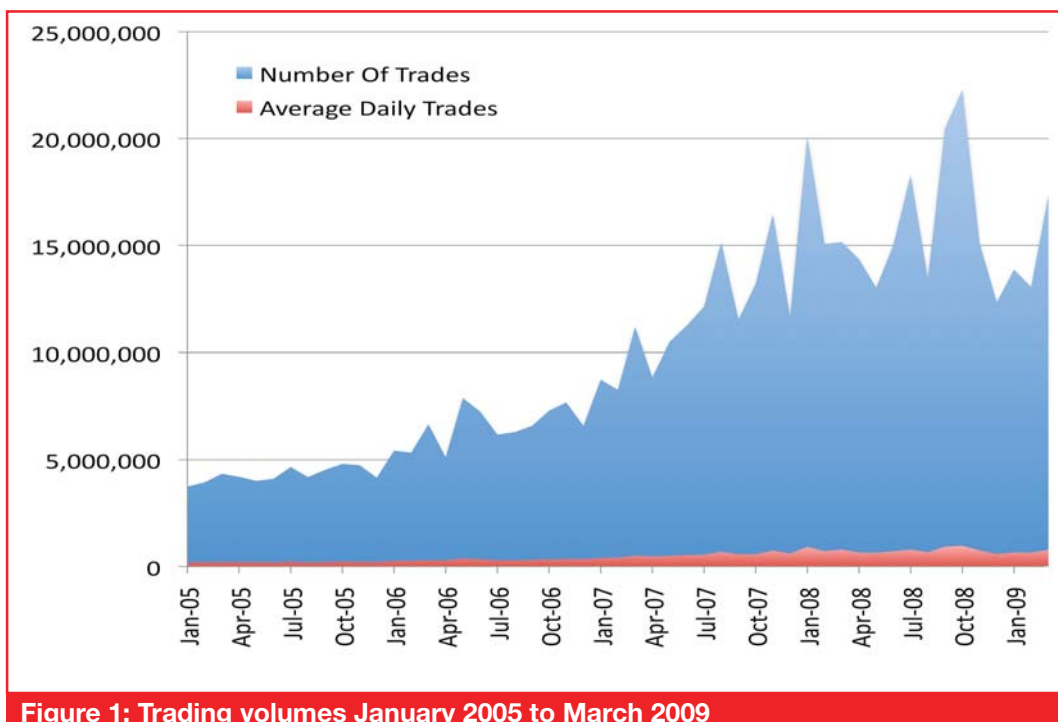


Figure 1: Trading volumes January 2005 to March 2009

In the past, rising volumes have led to conflation of tick data – the commingling of multiple instances of a trade or a quote at a single price into a single update. But today, firms seeking to test their trading strategies, or recreate market conditions at the time of a past trade, increasingly need every tick. That means every last trade, but also every bid and offer recorded by the marketplace in question.

Historical data of this kind – generated by exchanges and trading venues themselves – is now used by a range of market practitioners. These range from hedge funds, mutual funds, brokerages, investment banks and market-makers, through to data vendor redistributors and exchanges themselves.

The types of use for this data are similarly wide-ranging. Perhaps the most common use is in the development and pre-production testing of quantitative and algorithmic trading models. Indeed, the disciplines involved in these activities can be applied more widely into the realm of programme trading, market-making, statistical arbitrage and building transaction cost models, all of which involve running historical simulations and performing back-testing.

Another important area of interest, though, is in compliance. In particular, with recent U.S. and European regulatory scrutiny of practices around providing so-called best execution, firms are required to build and maintain detailed audit trails of their trading activities. This commonly involves the storage and analysis on demand of vast quantities of historical tick and order data.

And the quantities of data involved are, unsurprisingly, vast. The London Stock Exchange, for example, supports two historical data services, a tick database of last-trade data and a so-called Rebuild Order Book service, offering historical order data that allows market practitioners to recreate order books from recorded bids and offers.

The Exchange's data services extend back more than a decade, to September 1996. The tick database file currently consists of approximately 1 million trade details on peak-volume days, while the Rebuild Order Book service hits 20 million or so records on peak days. While volumes clearly have grown, particularly over the past few years, the entire history equates to a perhaps incomprehensible amount of data, which needs to be stored, accessed and manipulated by those firms subscribing to it.

With both business need and physical volumes of data escalating rapidly, historical trade and order data represent a growing challenge, for a widening range of professionals within financial institutions. Today, the requirement for collecting every tick, and storing, maintaining and manipulating it in a standardized way, is a core function for trading strategists, market information specialists, compliance officers and the IT and data staff that support them.

“With both business need and physical volumes of data escalating rapidly, historical trade and order data represent a growing challenge, for a widening range of professionals within financial institutions.”

Historical Data for Trading Strategies and Models

Using tick and order book data for back-testing

Recent regulations – in the U.S. and Europe especially, among them the EU’s Markets in The past several years has seen the rise – and move into the trading mainstream – of algorithmic trading methodologies. Pioneers like Credit Suisse’s Advanced Execution Services (AES) group introduced the concept of trading strategies through the adoption of such concepts as volume-weighted average price (VWAP) and time-weighted average price (TWAP), trading strategies that have since become benchmarks to which buy-side firms pin the performance of their executions.

These models – and others like Time Slice and Point of Value (POV) – started life on the proprietary trading desks of sell-side firms, then later ‘productized’ for consumption by qualified customers; usually, sophisticated hedge fund managers, buy-side traders and smaller boutique brokerages. Their increasing success, and broader uptake, spawned a raft of more proprietary models. Again, Credit Suisse AES set the pace with its Guerrilla

model, followed by the likes of Goldman Sachs’ Sonar and J.P. Morgan’s Aqua and Arid models.

“The proliferation of both standard benchmark and highly proprietary trading algorithms – and the impact they’ve had on market volumes – is well documented. Less lauded, though, is the key role of historical data in their development”

Indeed, a recent poll among readers of *Electronic Trading*, published by

A-Team Group, found that fully 73% of respondents expected the use of algorithmic trading to increase in 2009, despite (or perhaps because of) the market conditions. Of these, 40% expected to see expanded use of adaptive and custom algorithms, further underscoring the potential to gain competitive advantage through the development of proprietary models.

The proliferation of both standard benchmark and highly proprietary trading algorithms – and the impact they’ve had on market volumes – is well documented. Less lauded, though, is the key role of historical data in their development. Simply put, without reams of historical trade and order data to test them against, many of these models would not have seen the light of day.

By happy coincidence, growing demand for more adaptable algorithms, able to react swiftly to changes in market structure and sector performance, has emerged as new technologies able to process huge quantities of data have become more affordable. These technologies include complex event processing (CEP) platforms and other high-performance database systems, supplied by software houses like Progress Apama, Aleri/ Coral8, Kx Systems, Vhayu Technologies and others.

Their high-performance processing platforms act like very fast databases, allowing vast amounts of data to be processed in order to replicate market conditions from historical data. These are used in production trading systems and market surveillance systems, but importantly often form part of the apparatus for back-testing trading strategies. Tick-by-tick historical data allows quantitative analysts – teams of whom are working

round the clock to keep firms' algorithmic models up to speed with market developments – to look for trading patterns and test hypotheses around market behaviours. By running their trading models against historical tick data, sometimes along with live real-time data, quants are able to conduct what-if scenario testing, allowing them to assess how the models would behave under certain market conditions.

Taking a block of historical data – going back five or 10 years, say – an analyst could compare how a particular instrument, sector or stock reacted to historical events, looking for patterns that could be incorporated into a trading model. He or she could apply a benchmark like VWAP or TWAP to the same data, and compare the new model's performance against that benchmark. The data could be captured in charts and graphs in order to illustrate to users how the model's performance is affected by different market conditions and scenarios.

Key data quality considerations

In order to build a realistic view of market conditions, however, it's important to reconstruct the moment using the most accurate data available. Indeed, there are several data quality and standardization factors that need to be taken into consideration if the quants' 'what-if' scenarios are to give a useful view of how a particular model might behave.

First, it's important to ensure that the historical data being analyzed is recent enough to be relevant to the task in hand. While some analysts may seek to assess the impact of a model over an extended period – some may routinely look at data extending back 10 years or more – the results may be questionable if the data is, in essence, old. The period it covers must be able to yield market conditions that have a bearing on today's situation. Indeed, some analysts may opt to receive daily trade and order histories, and re-test their algorithms' hypotheses on a daily basis, particularly during periods of extreme uncertainty or volatility.

“From a logistical standpoint, the timestamp, data format and delivery mechanism must all be consistent for the period of data being analyzed”

In terms of standardization, data being analyzed must be consistent over the full period being analyzed. From a content standpoint, this means the same segment or sector constituents, the same data points and the same rules around order types, for example, are required if an 'apples to apples' view of the data is to be achieved over a period of time.

From a logistical standpoint, the timestamp, data format and delivery mechanism must all be consistent for the period of data being analyzed. This stress on 'clean' data is especially important for the increasingly automated process of back-testing, where human input is minimized so as to allow for the handling of huge quantities of data.

Finally, from a performance and quality standpoint, the more granular the available data, the more accurate the outcome of the analysis will be. For example, some originators publish conflated order details, rather than full tick-by-tick data, with the effect of making the data more opaque. Conflation involves the commingling of orders posted at the same bid or offer price within a given time-band, perhaps a second. Clearly, a true view of market movements – particularly during periods of wide price fluctuations and volatility – it's crucial to consider each individual order, where possible, in order to build a truly accurate picture of what was occurring in the marketplace at that particular moment in

time. This can have repercussions, too, for regulatory reporting. Indeed, some practitioners argue that sourcing the data from the original issuer can ensure consistent quality, in part because not all redistributors get access to all data available from the originators. Moreover, different redistributors may have different approaches to data cleansing, which can cause problems if data from different third parties is used.

Frequency of update is important. The London Stock Exchange now provides millisecond time-stamping, offering a high level of granularity, which can be crucial in times of high volatility.

Challenges posed by intense market volatility

There is something of an ongoing industry debate on how best to deal with periods of high volatility. The jury is still out, for example, on whether analysis of the market gyrations that followed the collapse of Lehman Brothers last autumn can provide useful insight into future market movements. Clearly, spikes and troughs in market numbers – for either orders or trades – can impact the outcome of ‘what-if’ scenario testing.

Historical Data and Regulatory Compliance

Aside from its value in helping analysts develop trading models, historical tick data is seeing demand driven by new regulations, particularly those around best execution. The EU’s Markets in Financial Instruments Directive (MiFID) is a case in point.

MiFID requires sell-side firms to demonstrate how they achieved best execution for their clients. Increasingly, this involves storing large quantities of historical data, and being able to reconstruct from that data the precise market circumstances at the time of any particular trade.

This can be used to offer an assessment of execution quality. In order to rebuild the order book for a specific moment in time, it’s necessary to capture and store tick-by-tick order data. Indeed, MiFID requires storage of data relating to trades for a period of five years, allowing regulators to query firms on their best execution performance.

Some data suppliers – such as the London Stock Exchange’s Proquote, for example – are offering execution quality applications to their clients, which draw on historical exchange data in order to rebuild order books from past trading environments.

Help is at Hand: Officially Sourced Data from the London Stock Exchange

Standard and Poor's has long been the authoritative, objective, and credible information source. For U.K. market participants, the London Stock Exchange has two main historical tick data offerings that can answer practitioners' concerns about data coverage, comprehensiveness and quality. Many players use the London Stock Exchange's historical tick, or trade, data stretching back to 1996. It also offers a service known as Rebuild Order Book, which carries quote data from the same period. Customers are able to select the period for which they require coverage.

Data is available for all U.K. securities traded on the London Stock Exchange, including SETS, SEAQ and SEATS Plus stocks. International Retail Service and International Order Book securities are also included. And an optional best price file can be provided with the tick data, showing the best prices available at the times each of the trades occurred.

Data is sourced from the exchange's TradElect trading platform, and its Infolect distribution mechanism. As such, it includes all trades and orders, and is not manipulated by the exchange in any way. This guarantees that the data is generated directly from its source, giving users a pure view of each tick on the exchange, both in terms of trades and orders.

The exchange's trade data service includes every trade executed for the trading day, which runs from 8.30 a.m. to 4 p.m. The trade data file includes ISIN instrument code, data and time, to the second, price, size and a unique trade code. Rebuild Order Book is also delivered on a daily basis, usually distributed around 8 p.m. to 9 p.m. The service includes an end-of-day summary, featuring a best price file with all instrument details, bid and offer price and mid-price. Since March 2008, order book data has been time-stamped to the millisecond.

The London Stock Exchange has invested significantly to ensure data quality in the face of huge rises in volumes in recent years. The trade data file typically contains about 300,000 to 400,000 trades per day, and peaks at around one million trades. Due to the rising incidence of algorithmic trading, the number of orders is significantly higher, sometimes peaking at 20 million per day.

The exchange rationalized the number of data fields it delivers some 18 months ago, dropping redundant columns and adding a reference file number for each entry. It also added millisecond time-stamping in March 2008. Data files are zipped up for delivery, either via CD or FTP, depending on the amount of data required by the client. Data is delivered in flat file format, separated by commas.

Clients, the exchange says, may take varying amounts of this historical data for their back-testing needs. Most have their own internal software and systems to run their testing or rebuilding order books.



London
Stock Exchange

www.londonstockexchange.com/tradedata

Datalect – the natural source of financial data

The London Stock Exchange's Datalect suite incorporates a wide range of reference and historical data products:

- SEDOL Masterfile
- Corporate actions
- Datasync
- Company data
- Trade data
- Valuation data
- Market report service.

Our Datalect products are powered with robust data from the heart of the financial markets, ensuring you know what is happening in your own organisation and across the financial world. These products are widely trusted and are often used as the benchmark to drive decisions.

For more information contact:

Ian Wood

+44 (0)20 7797 3827

iwood@londonstockexchange.com

www.londonstockexchange.com/tradedata

A-TEAMGROUP

www.a-teamgroup.com

A-Team Group, founded in 2001, is a publishing, research and events group that provides the global community of IT and data professionals in financial markets with the business intelligence they need to excel in their roles.

We're widely known for our focused series of publications, research and events across mission-critical functions including high-performance trading infrastructure, low-latency market data and connectivity, enterprise data management, reference data, market structure, risk and regulation and more.

A-Team Group's publishing division publishes a range of online and in print news services including Risk & Regulation IT, Reference Data Review, Market Data Insight, AsiaMarketsIT.com, Low-Latency.com, FinTech-Infrastructure.com, and A-Team IQ magazine. Find out more at www.a-teamgroup.com/publishing

A-Team Group's research division collates data, performs analysis and delivers results for internal product development and marketing, as well as publishing in-depth reports for external use, examples of which include:

- OTC Valuations - Pricing Assets in the Post Credit Crunch World
- Buy-side Firms Take a Hard Look at Data Practices: Derivatives Push the Data Management Envelope
- Buyer Persona - The Influence Behind Data Management Decisions, and more...

Find out more at www.a-teamgroup.com/research

A-Team Group's events division partners with companies to produce highly-targeted seminars on specific business challenges, attracting senior-level industry practitioners, such as its recent series of Algorithmic Trading seminars with Dow Jones across Europe and North America, and upcoming events with GoldenSource and Sybase.

For more information visit

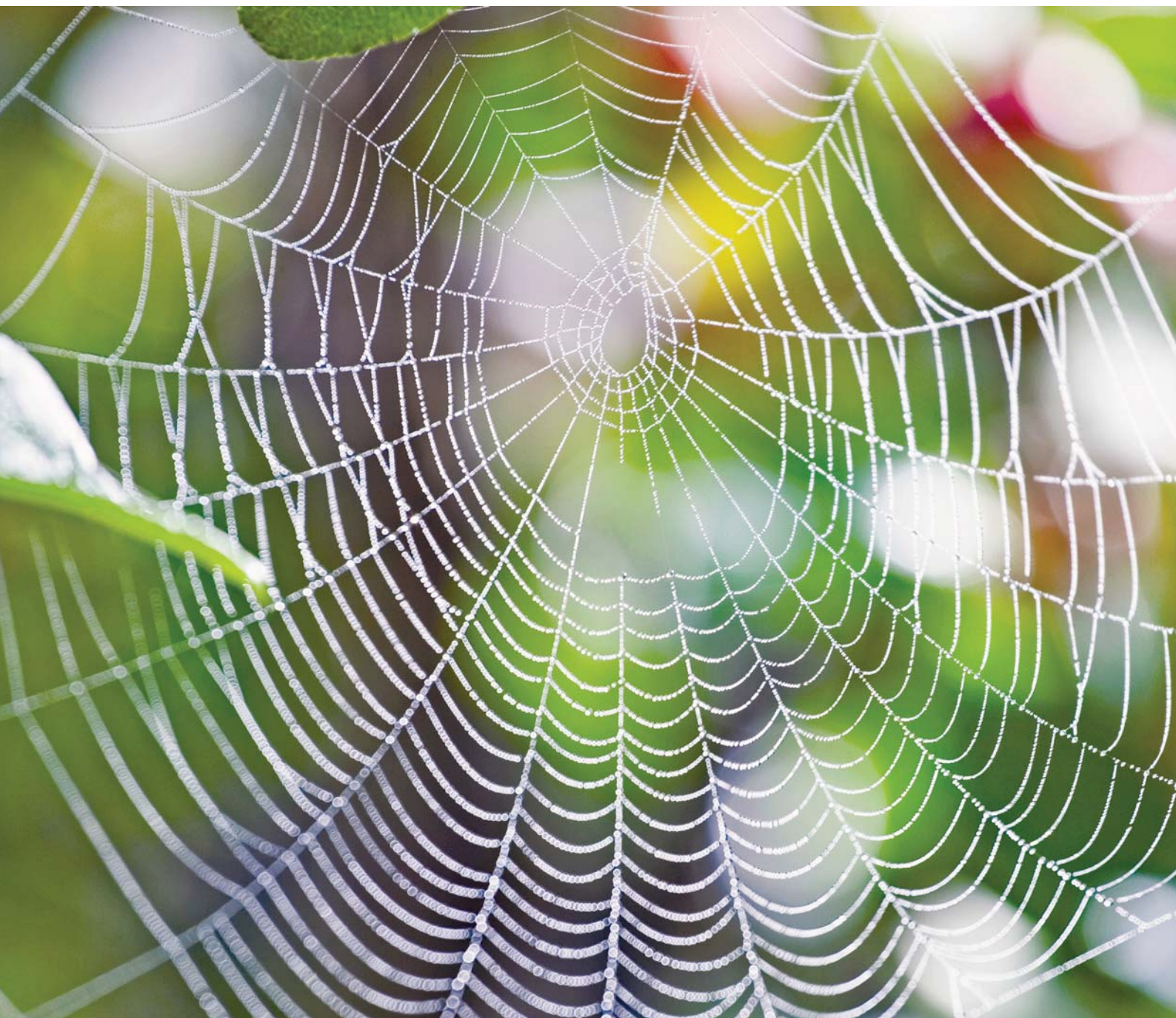
www.a-teamgroup.com



A-TEAMGROUP



London
Stock Exchange



London Stock Exchange

10 Paternoster Square
London EC4M 7LS
T +44 (0)20 7797 1000
www.londonstockexchange.com

Copyright © June 2009 London Stock Exchange plc.
Registered in England and Wales No. 2075721.
London Stock Exchange plc has used all reasonable efforts to ensure that the information contained in this publication is correct at the time of going to press, but shall not be liable for decisions made in reliance on it.
London Stock Exchange, the coat of arms device and SEDOL Masterfile are registered trade marks of London Stock Exchange plc.
Datsync is a trade mark of the London Stock Exchange plc