Automated Analysis of News to Compute Market Sentiment: Its Impact on Liquidity and Trading

The Future of Computer Trading in Financial Markets - Foresight Driver Review – DR 8

Contents

Review Authors:

Gautam Mitra
Director: CARISMA Brunel University & CEO OptiRisk Systems

Dan diBartolomeo
CEO Northfield Information Services, USA

Ashok Banerjee
Professor and Chairperson
Financial Research and Trading Lab, IIM Calcutta, India

Xiang Yu
PhD student at CARISMA Brunel University

20 July 2011

Abstract

Computer trading in financial markets is a rapidly developing field with a growing number of applications. Automated analysis of news and computation of market sentiment is a related applied research topic which impinges on the methods and models deployed in the former. In this review we have first explored the asset classes which are best suited for computer trading. We present in a summary form the essential aspects of market microstructure and the process of price formation as this takes place in trading. We critically analyse the role of different classes of traders and categorise alternative types of automated trading. We introduce alternative measures of liquidity which have been developed in the context of bid-ask of price quotation and explore its connection to market microstructure and trading. We review the technology and the prevalent methods for news sentiment analysis whereby qualitative textual news data is turned into market sentiment. The impact of news on liquidity and automated trading is critically examined. Finally we explore the interaction between manual and automated trading.

Introduction

This report is prepared as a driver review study for the Foresight project: The Future of Computer Trading in Financial Markets. Clearly the focus is on (i) automated trading and (ii) financial markets. Our review, the title as above, brings the following further aspects into perspective: (iii) automated analysis of news to compute market sentiment, (iv) how market sentiment impacts liquidity and trading.

The neoclassical finance theory had embraced the efficient market hypothesis (EMH) which became its central plank. The doubts about EMH and recent surge of interest in behavioural finance (Kahneman and Tversky, 1979, Kahneman, 2002, Shefrin, 2008) not only debated and exposed the limits of EMH but also reinforced the important role of sentiment and investor psychology in market behaviour. The central tenet of EMH was that information arrival, that is, news is rapidly discounted by rational stakeholders; yet the work of Shiller (2000) and Hais (2010) reinforce the irrational contrarian and herd behaviour of the investors. Today the availability of sophisticated computer systems facilitating high frequency trading (Goodhart and O'Hara, 1997) as well as access to automated analysis of news feeds (Tetlock, 2007, Mitra and Mitra, 2011) set the backdrop for computer automated trading which is enhanced by news. The findings of this driver review may be summarized in the following way. Computer mediated automated trading continue to grow in many venues where equities, futures, options and foreign exchange are traded. The research and adoption of automated analysis of newsfeed for trading, fund management and risk control systems are in their early stages but gaining in momentum. The challenge for the trading and the investment community is to bring the hardware technologies, software techniques and modelling methodologies together. The challenge for the regulatory authorities is to understand the combined impact of these technologies and postulate regulations which can control volatility, improve the provision for liquidity and generally stabilize the market behaviour.

1.1 The Market Participants

Over the last forty years there have been considerable developments in the theory which explains the structure, mechanisms and the operation of financial markets. A leader in this field Maureen O'Hara in her book (O'Hara, 1995) summarises in the following way: the classical economic theory of price formation through supply and demand equilibrium is too simplistic and does not quite apply to the evolving financial markets. Thus leading practitioners and specialists in finance theory, Garman (1976) and Madhavan (2000) amongst others, started to develop theoretical structures with which they could explain the market behaviour. Indeed the field of market microstructure came to be established in order to connect the market participants and the mechanisms by which trading takes place in this dynamic and often volatile and tempestuous financial market. Again quoting O'Hara : 'Any trading mechanism can be viewed as a type of trading game in which players meet (perhaps not physically) at some venue and act according to some rules. The players may involve a wide range of market participants, although not all types of players are found in every mechanism. **First**, of course, **are customers** who submit orders to buy or sell. These orders may be contingent on various outcomes or they may be direct orders to transact immediately. The exact nature of these orders may depend upon the rules of the game. **Second**, there **are brokers** who transmit orders for customers. Brokers do not trade for their own account, but act merely as conduits of customer orders. These customers may be retail traders or they may be other market participants such as dealers who simply wish to disguise their trading intentions. **Third** there **are dealers** who do trade for their own account. In some markets dealers also facilitate customer orders and so are often known as broker/dealers. **Fourth**, there **are** specialists, or **market makers**. The market maker quotes price to buy or sell the asset. Since the market maker generally takes a position in the security (if only for a short time waiting for an offsetting order to arrive), the market maker also has a dealer function'. We quote this text as it provides

a very succinct definition of the relevant market **participants** and the **trading mechanisms**. From a commercial perspective there are other **tertiary market participants** such as **market data feed providers** and now **news data feed providers** whose influence can no longer be ignored and indeed they play important roles in automated trading. We observe that the theory is well developed to describe trading by human agents. We are now in a situation whereby trading takes place both as orders placed by human agents and by computer automated (trade) orders placed side by side at the same trading venues. Here we make a distinction between computer mediated communication of orders through Electronic Communications Network (ECN) and its execution and settlement, and orders generated by computer algorithms and its subsequent processing in the above sequence.
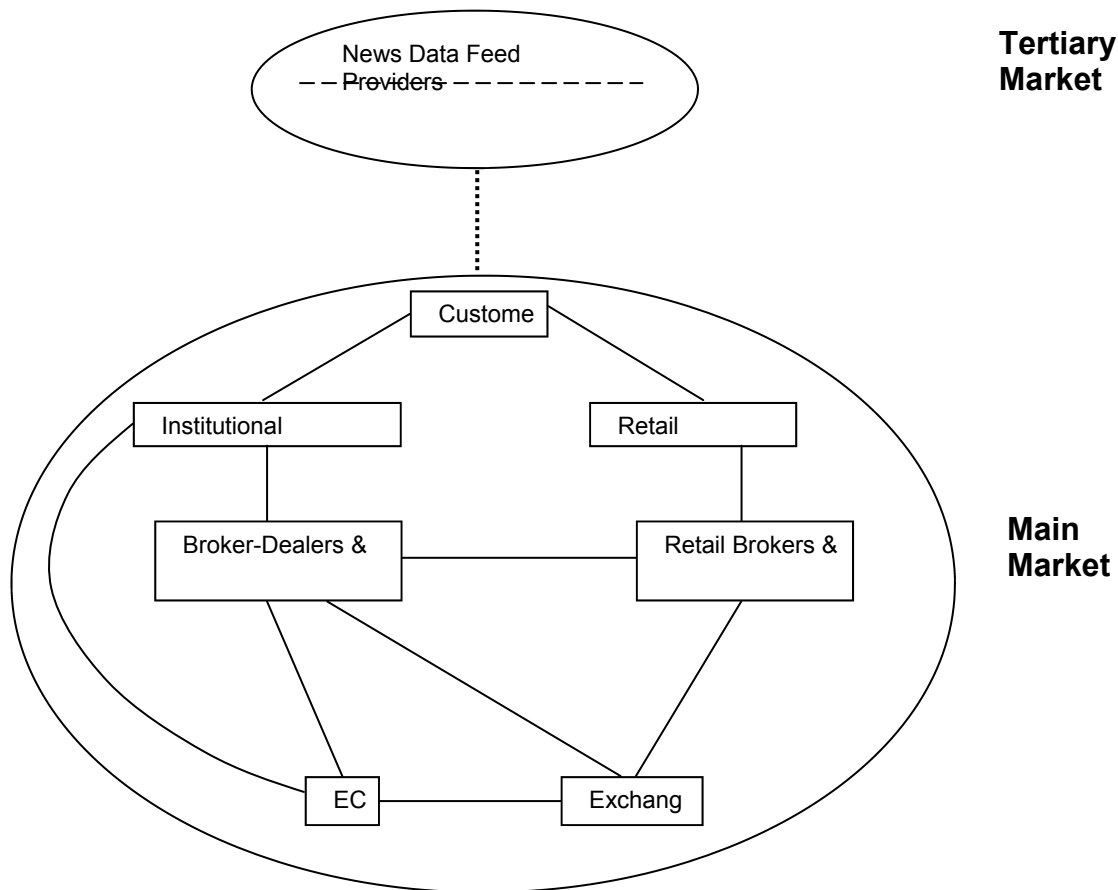
Fig 1.1 Market participants and their connectivity.

The relationship between these market participants are set out as a connectivity diagram shown in figure 1.1.

1.2 Structure of the Review

Automated trading has progressed and has gained increasing market share in those asset classes for which the markets are highly liquid and trading volumes are large. In section 2 of this report we consider briefly these asset classes; our review is, however, focused on equities as the automated news sentiment analysis is mostly developed for this asset class. A vast amount of literature has emerged on the topic of market microstructure and liquidity; the finance community, especially those concerned with trading are, very much involved in the development and understanding of the market mechanism which connect trading and liquidity. In section 3 we provide a summary of the relevant concepts of market microstructure and liquidity and these serve as a back drop for the rest of the report. In section 4 we first consider

the different trader types, namely, informed, uninformed and value traders; we also analyse automated trading and break it down to five major categories. In section 5 we provide an introduction and overview of news analytics in a summary form. News analytics is an emerging discipline. It has grown by borrowing research results from other disciplines, in particular, natural language processing, text mining, pattern classification, and econometric modelling. Its main focus is to automate the process of understanding news presented qualitatively in the form of textual narratives appearing in newswires, social media and financial blogs and turning these into quantified market sentiments. The market sentiment needs to be measured and managed by an automated process which combines data feeds and news feeds. In turn this process automates trading and risk control decisions. In section 6 we make the connection between earlier sections in respect of the informed traders and news analytics. In this context **news** is considered to be **an information event** which influences price formation, volatility of stock price as well as the liquidity of the market and that of a given stock. In short it impacts the market microstructure. There are now a growing number of research papers (see Mitra and Mitra, 2011[a]) which connect News analytics with (i) pricing and mispricing of stocks and discovering alpha, (ii) fund management and (iii) risk control. However, very few research papers or studies are available in open literature which connect news analytics with automated trading; the two major vendors of news analytics data and market sentiment (RavenPack, 2011 and Thomson Reuters, 2011, see appendix in Mitra and Mitra 2011[b]) due to client confidentiality only reveal limited information about the use of these data sets. In section 7 we consider the modelling and the information architecture by which automated analysis of news is connected to automated trading. In the final section of this review, that is, section 8 we give a summary discussion of the various findings and present our conclusions.

Consideration of asset classes for automated trading

In this section we consider the criterion which makes an asset class suitable for automated trading. These criteria are mainly associated with the market conditions of asset classes. Typically such market conditions include (i) sufficient market volatility and (ii) a high level of liquidity. This is so that firstly, changes in price are able to exceed transaction costs thereby making it possible to earn profits, and secondly, in order to make it feasible to move quickly in and out of positions in the market, which is a crucial criterion underpinning the strategies of high frequency trading. On top of this, the market needs to be electronically executable in order to facilitate the quick turnover of capital and to harness the speed of automated trading. Currently, only spot foreign exchange, equities, options and futures markets fulfil such conditions of automated execution.

Set against these considerations, we examine the suitability of computer trading of the following asset classes: (i) Equity markets, (ii) Foreign exchange markets, (iii) Commodity markets, (iv) Fixed income markets.

To start with, we study the relationship between trading frequency and liquidity for the above mentioned asset classes (see Fig 2.1); the daily trading volume has been used as a proxy for liquidity. We observe that asset classes which are traded at an optimal frequency of less than a day tend to be accompanied with higher levels of liquidity in that market. We use a simple working definition of liquidity: "the ability to convert assets into cash at the lowest possible transaction costs". Readily available liquidity becomes an attractive criterion for investors as they are able to trade without worrying about transaction costs eating away at their profits. We also note that the highly liquid asset classes are also those that are executed electronically and are traded on a more regular basis. These asset classes are therefore natural candidates for high frequency trading. We observe that a sweeping but steady transition has occurred with the conversion of over-the-counter markets into electronic markets to keep up with the trading strategies of investors.
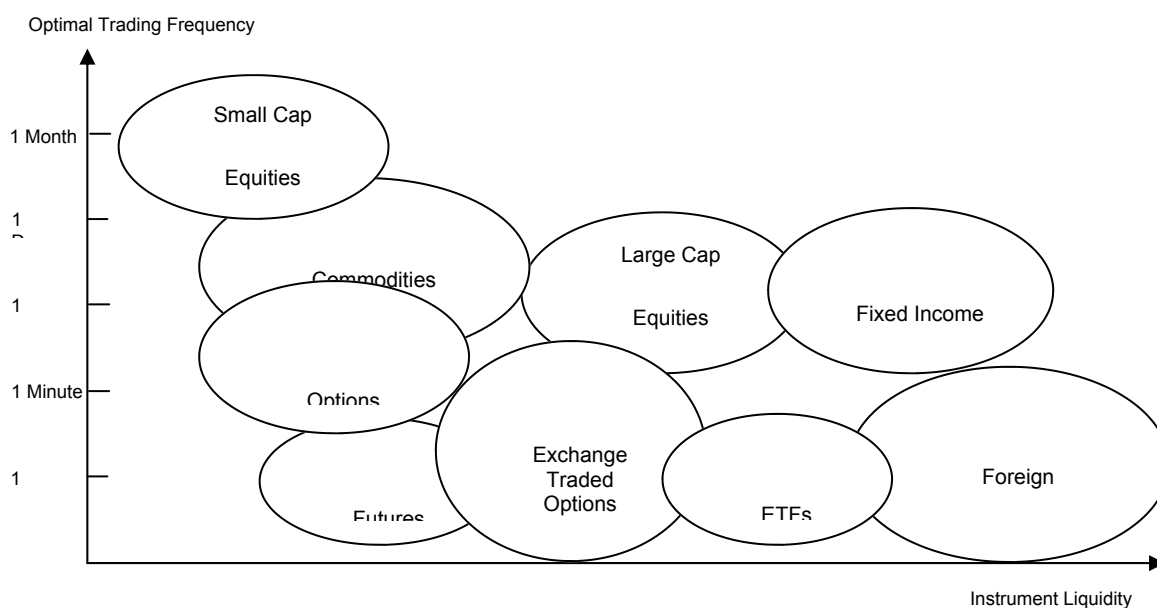
Figure 2.1   The trade-off between optimal trading frequency and liquidity for various trading instruments.

We assess each asset class individually, and analyze the properties which make them suitable for computer trading.

Equity markets

This is the most favoured asset class for automated trading because of the large size and volume of the market; this is supported by the market's breadth of listed stocks. It is also popular for its diversification properties in portfolio investment with its possible positions to long and short stocks. In addition to stocks which are traded in the equity markets, the market also includes exchange-traded funds (ETFs), warrants, certificates and structured products. In particular, hedge funds are especially active in trading index futures. According to research conducted by Aite Group, the asset class that is executed the most algorithmically is equities; for instance, by 2010 an estimated 50% or more of total volume of equities traded were handled by algorithms.
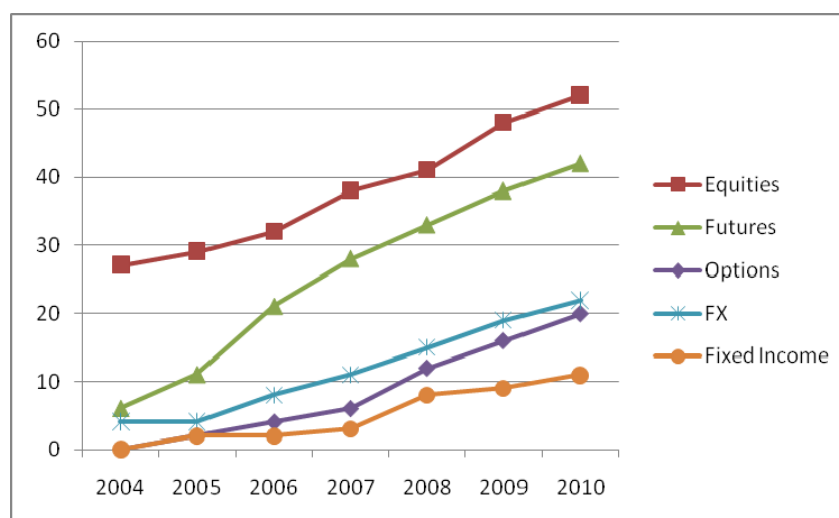
Figure 2.2 Progress in adoption of algorithmic execution by asset class from 2004 - 2010. *Source: Aite Group.*

Foreign exchange markets

The foreign exchange markets operate under a decentralised and unregulated mechanism whereby commercial banks, investment banks, hedge funds, proprietary trading funds, non-bank companies and non-U.S. investment banks all have access to the inter-dealer liquidity pools. However, due to this decentralisation, the foreign exchange markets lack volume measures and the rule of "one price". This has beneficial implications for automated traders as there are substantial arbitrage opportunities that can be identified by their automated strategies. However, there are only a limited number of contracts that may be found on the exchange, restricting the variety of financial instruments available for traders in the foreign exchange market, namely foreign exchange futures and selective options contracts. Over the years, there has been a swift transition from major trading in the spot foreign exchange markets to swaps.

Under the measure of liquidity as the average daily volume of each security, the foreign exchange market ranks as the most liquid market, followed by US Treasury securities. This volume figure is collected and published by the Bank for International Settlements, who conduct surveys of financial institutions every three years. There is no direct figure for traded volume to monitor developments in the foreign exchange market because of the decentralized structure for these markets.

Commodity markets

The financial products in the commodity markets that are liquid and electronically traded are commodity futures and options, to allow viable and profitable trading strategies in automated trading. Futures contracts in commodities tend to be smaller than the futures contracts in foreign exchange.

Fixed income markets

The fixed income markets include the interest rate market and the bond market, with securities traded in the form of either a spot, a future or a swap contract. The interest rate market trades short and long term deposits, and the bond market trades publicly issued debt obligations. The fixed income feature of these markets comes from the pre-specified or fixed income that is paid to their holders, which in turn is what automated traders focus their strategies on to take advantage of short-term price deviations and make a profit.

In the interest rate futures market, liquidity is measured by the bid-ask spread. A bid-ask spread on interest rate futures is on average one-tenth of the bid-ask spread on the underlying spot interest rate. The most liquid futures contract in the interest rate market is short-term interest rate futures. Swap products are the most populous interest rate category, yet most still trade over the counter.

The bond market contains an advantageous breadth of products; however, spot bonds are still mostly transacted over the counter. Bond futures contracts on the other hand are standardised by the exchange and are often electronic. The most liquid bond futures are associated with those bonds which are nearing their expiry dates compared to those with longer maturities.

Market microstructure and liquidity

3.1 Market microstructure

A financial market is a place where traders assemble to trade financial instruments. Such trades take place between willing buyers and willing sellers. The market place may be a physical market or an electronic trading platform or even a telephone market. The trading rules and trading systems used by a market define its market structure. Every market has procedures for matching buyers to sellers for trades to take place. In quote-driven markets dealers participate in every trade. On the other hand, in order-driven markets, buyers and sellers trade with each other without the intermediation of dealers. Garman (1976) coined the expression "market microstructure" to study the process of market making and inventory costs. Market microstructure deals with operational details of a 'trade' – the process of placement and handling of orders in the market place and their translation into trades and transaction prices. One of the most critical questions in market microstructure concerns the process by which new information is assimilated and price formation takes place. In a dealer-driven market, market makers, who stand willing to buy or sell securities on demand, provide liquidity to the market by quoting bid and ask prices. In a quote-driven market, limit orders provide liquidity. While the primary function of the market maker remains that of a supplier of immediacy, the market maker also takes an active role in price-setting, primarily with the objective of achieving a rapid inventory turnover and not accumulating significant positions on one side of the market. The implication of this model is that price may depart from expectations of value if the dealer is long or short relative to the desired (target) inventory, giving rise to transitory price movements during the day and possibly over longer periods (Madhavan, 2000). Market microstructure is concerned with how various frictions and departures from symmetric information affect the trading process (Madhavan, 2000). Microstructure challenges the relevance and validity of the random walk model.

The study in market microstructure started about four decades ago and it has attracted further attention in the past decade with the advent of computer-driven trading and availability of all trade and quote data in electronic form, leading to a new field of research called high frequency finance. Research in high frequency finance demonstrates that properties that define the behaviour of a financial market using low frequency data fail to explain the market behaviour observed in high frequency. Three events are cited (Francioni et al, 2008) as early triggers for the general interest in microstructure:
  (a) the U.S. Securities and Exchange Commission's Institutional Investor Report in 1971;
  (b) the passage by the U.S. Congress of the Securities Acts Amendment of 1975; and
  (c) the stock market crash in 1987

Market microstructure research typically examines the ways in which the working process of a market affects trading costs, prices, volume and trading behaviour. Madhavan (2000) classified research on microstructure into four broad categories:

(i)	price formation and price discovery;
(ii)	market structure and design issues;
(iii)	market transparency; and
(iv)	informational issues arising from the interface of market microstructure

The effect of market frictions (called microstructure noise) is generally studied by decomposing transaction price of a security into fundamental component and noise component. Ait-Sahalia and Yu (2009) related the two components to different observable measures of stock liquidity and found that more liquid stocks have lower (microstructure) noise.

The knowledge of market systems and structure is essential for a trader to decide in **which market** to trade and **when** to trade. Such knowledge would also facilitate a trader to assess the relative efficiency of the market and hence the arbitrage opportunities. In fact, the trading behaviour, and trading costs are affected by market microstructure. We turn next to market liquidity.

3.2 Market liquidity

Liquidity is an important stylized fact of financial markets. Echoing the description put forward by cognoscenti practitioners, O'Hara (O'Hara 1995) introduces the concept in the following way: 'liquidity , like pornography, is easily recognized but not so easily defined; we begin our analysis with a discussion of what liquidity means in an economic sense'. A market is termed liquid when traders can trade without significant adverse affect on price (Harris, 2005). Liquidity refers to the ability to convert stock into cash (or vice versa) at the lowest possible transaction cost. Transaction costs include both explicit costs (e.g. brokerage, taxes) and implicit costs (e.g. bid-ask spreads, market impact costs). More specifically Black (1971) pointed out the presence of several necessary conditions for a stock market to be liquid:

- (a) there are always bid-and-ask prices for the investor who wants to buy or sell small amounts of stock immediately;
- (b) the difference between the bid and ask prices (the spread) is always small;
- (c) an investor who is buying or selling a large amount of stock, in the absence of special information, can expect to do so over a long period of time, at a price not very different, on average, from the current market price; and
- (d) an investor can buy or sell a large block of stock immediately, but at a premium or discount that depends on the size of the block – the larger the block, the larger the premium or discount.

Liquidity is easy to define but very difficult to measure. The various liquidity measures fall into two broad categories: trade-based measures and order-based measures (Aitken and Carole, 2003). Trade-based measures include trading value, trading volume, trading frequency, and the turnover ratio. These measures are mostly ex post measures. Order-driven measures are *tightness/width* (bid-ask spread), *depth* (ability of the market to process large volumes of trade without affecting current market price), and *resilience* (how long the market will take to return to its "normal" level after absorbing a large order). A commonly used measure of market depth is called Kyle's Lambda (Kyle, 1985):

$$\lambda = \frac{r_t}{NOF_t}$$

where r_t is the asset return and NOF_t is the net order flow over time. The parameter λ can be obtained by regressing asset return on net order flow.

Another measure of market depth is Hui-Heubel (HH) liquidity ratio (Hui and Heubel, 1984). This model was used to study asset liquidity on several major U.S equity markets, and relates trading volume to the change of asset price. Given the market activities observed over N unit time windows, the maximum price P_{Max}, minimum price P_{Min}, average unit closing price P, total

dollar trading volume V, and total number of outstanding quotes Q, the Hui-Heubel L_{HH} liquidity ratio is given as follows:

$$L_{HH} = \frac{(P_{max} - P_{min})/P_{min}}{V/Q * \overline{P}}$$

A higher HH ratio indicates higher price to volume sensitivity.

Resilience refers to the speed at which the price fluctuations resulting from trades are dissipated. Market-efficient coefficient (MEC) (Hasbrouck and Schwartz, 1988) uses the second moment of price movement to explain the effect of information impact on the market. If an asset is resilient, the asset price should have a more continuous movement and thus low volatility caused by trading. Market-efficient coefficient compares the short term volatility with its long term counterpart. Formally:

$$MEC = \frac{Var(R_{long})}{T \cdot Var(R_{short})}$$

where *T* is the number of short periods in each long period. A resilient asset should have a MEC ratio close to 1.

Literature also has precedence for another aspect of liquidity: *immediacy* - the speed at which trades can be arranged at a given cost. Illiquidity can be measured by the cost of immediate execution (Amihud and Mendelson, 1986). Thus, a natural measure of illiquidity is the spread between the bid and the ask prices. Later, Amihud (2002) modified the definition of illiquidity. The now-famous illiquidity measure is the daily ratio of absolute stock return to its dollar volume averaged over some period:

$$ILLIQ_{iy} = 1/D_{iy} \sum_{t=1}^{Diy} |R_{iyd}|/VOLD_{ivyd},$$

where R_{iyd} is the return on stock *i* on day *d* of year *y* and $VOLD_{iyd}$ is the respective daily volume in dollars. D_{iy} is the number of days for which data are available for stock *i* in year *y*.

The vast literature on liquidity studies the relationships of liquidity and the cost of liquidity with various stock performance measures, trading mechanisms, order-trader types and asset pricing. Acharya and Pederson (2005) present a simple theoretical model (liquidity-adjusted capital asset pricing model- LCAPM) that helps explain how liquidity risk and commonality in liquidity affect asset prices. The concept of commonality of liquidity was highlighted by Chordia et al. (2000) when the authors stated that liquidity is not just a stock-specific attribute given the evidence that the individual liquidity measures, like quoted spreads, quoted depth and effective spreads, co-move with each other. Later Hasbrouck and Seppi (2001) examined the extent and role of cross-firm common factors in returns, order flows, and market liquidity, using the analysis for the 30 Dow Jones stocks.

Asset prices are also affected by the activities and interactions of informed traders and noise traders. Informed traders make trading decisions based on exogenous information and true value of the asset. Noise traders do not rely on fundamental information to make any trade decision. Their trade decisions are purely based on market movements. Thus, noise traders are called trend followers.

Categorisation of trading activities

4.1 Trader types
Harris (1998) identifies three types of traders
(i)   liquidity traders also known as inventory traders (O'Hara 1995) or uninformed traders

(ii)   informed traders and
(iii)   value motivated traders.

The inventory traders are instrumental in providing liquidity; they make margins by simply keeping an inventory of stocks for the purpose of market making and realizing very small gains using limit orders through moving in and out of positions many times intra-day. Since the overall effect is to make the trading in the stock easier (less friction) they are also known as liquidity providers. These traders do not make use of any exogenous information about the stock other than its trading price and order volume. The informed traders in contrast assimilate all available information about a given stock and thereby reach some certainty about the market price of the stock. Such information may be acquired by subscription to (or purchased from) news sources; typically FT, Bloomberg, Dow Jones, or Reuters. They might have access to superior predictive analysis which enhances their information base. Value traders also apply predictive analytic models and use information to identify inefficiencies and mispricing of stocks in the market; this in turn provides them with buying or short selling opportunities. We note that the last two categories of traders make use of the value of information; such information is often extracted from anticipated announcements about the stock and is used in their predictive pricing models.

Automated trading
Automated trading in financial markets falls roughly into five categories:

(i)   Crossing Transactions
(ii)   Algorithmic Executions
(iii)   Statistical Arbitrage
(iv)   Electronic Liquidity Provision
(v)   Predatory Trading

Our first category, "crossing transactions" represents the situation where a financial market participant has decided to enter into a trade and seeks a counterparty to be the other side of the trade, without exposing the existence of the order to the general population of market participants. For example, an investor might choose to purchase 100,000 shares of stock X through a crossing network (e.g. POSIT) at today's exchange closing price. If there are other participants who wish to sell stock X at today's exchange closing price, the crossing network matches the buyers and sellers so as to maximize the amount of the security transacted. The advantage of crossing is that since both sides of the transaction have agreed in advance on an acceptable price which is either specified or formulaic in nature, the impact of the transactions on market prices is minimized. Crossing networks are used across various asset classes including less liquid instruments such as corporate bonds.

It should be noted that our four remaining categories of automated trading are often collectively referred to as "high frequency" trading. The second category of automated trading is "algorithmic execution". If a market participant wishes to exchange 1000 GBP for Euros, or buy 100 shares of a popular stock, modern financial markets are liquid enough that such an order can be executed instantaneously. On the other hand, if a market participant wishes to execute a very large order such as five million shares of particular equity Y there is almost zero probability that there exists a counterparty coincidentally wishing to sell five million shares of Y at the exact same moment, or even within a very short time window. One way of executing such a large order would be a principle bid trade with an investment bank, but such liquidity provision often comes at a high price. The alternative is an "algorithmic execution" where a large "parent" order is broken into many small "child" orders to be executed separately over several hours or even several days. In the case of our hypothetical five million share order, we might choose to purchase the shares over three trading days, breaking the large order into a large number of small orders (i.e. 200 shares on average) that would be executed throughout the three day period. Numerous analytical algorithms exist that can adjust the sizes of, and time between child orders to reflect changes in the asset price, general market conditions, or

the underlying investment strategy. Note that like crossing, automated execution is merely a process to implement a *known transaction* whose nature and timing has been decided by a completely external process. However, this class of 'algorithmic execution' will benefit from the inclusion of news analytics and predictive analysis of liquidity.

Our third category of automated trading is "statistical arbitrage". Unlike our first two categories, statistical arbitrage trading is based on *automation of the investment decision process.* A simple example of statistical arbitrage is "pairs trading". Let us assume we identify the relationship that "Shares of stock X trade at twice the price of shares of stock Z, plus or minus ten percent". If the price relation between X and Z goes outside the ten percent band, we would *automatically* buy one security and short sell the other accordingly. If we expand the set of assets that are eligible for trading to dozens or hundreds, and simultaneously increase the complexity of the decision rules, and update our metrics of market conditions on a real time basis, we have a statistical arbitrage strategy of the modern day. The most obvious next step in improving our hypothetical pairs trade would be insert a step in the process that automatically checks for news reports that would indicate that the change in the monitored price relationship had occurred as a result of a clear fundamental cause, as opposed to random price movements such that we would expect the price relationship to revert to historic norms. Pairs trading may also benefit by taking into consideration market sentiment as determined by news.

The fourth form of automated trading is electronic liquidity provision. This form of automated trading is really a direct descendent of traditional over-the-counter market making, where a financial entity has no particular views on which securities are overpriced or underpriced. The electronic liquidity provider is *automatically* willing to buy or sell any security within its eligible universe at some spread away from the current market price upon counterparty request. Electronic liquidity providers differ from traditional market makers in that they often do not openly identify the set of assets in which they will trade. In addition, they will often place limit orders away from the market price for many thousands of securities simultaneously, and engage in millions of small transactions per trading day. Under the regulatory schemes of most countries such liquidity providers are treated as normal market participants, and hence are not subject to regulations or exchange rules that often govern market making activities. Many institutional investors believe that due to the lack of regulation automated liquidity providers may simply withdraw from the market during crises, reducing liquidity at critical moments.

The final form of automated trading we address is "predatory trading". In such activities, a financial entity typically places thousands of simultaneous orders into a market while expecting to actually execute only a tiny fraction of the orders. This "place and cancel" process has two purposes. The first is an information gathering process. By observing which orders execute, the predatory trader expects to gain knowledge of the trading intentions of larger market participants such as institutional asset managers. Such asymmetric information can then be used to advantage in the placement of subsequent trades. A second and even more ambitious form of predatory trading is to place orders so as to artificially create abnormal trading volume or price trends in a particular security so as to purposefully mislead other traders and thereby gain advantage. Under the regulatory schemes of many countries there are general prohibitions against "market manipulation", but little if any action has been taken against predatory trading on this basis.

 Some practitioners believe (Arnuk and Saluzzi, 2008) automated trading puts the manual trading of retail investors, as well as institutional investors in considerable disadvantage from a perspective of price discovery and liquidity. A number of financial analytics/consulting companies typically Quantitative Services Group LLC, Greenwich Associates, Themis Trading LLC (particular mention should be made of insightful white papers posted by Arnuk and Saluzzi (2008) and (2009)) have produced useful white papers on this topic. (Please see web references in the reference section.)

Automated news analysis and market sentiment

Introduction and overview

A short review of news analytics focusing on its applications in finance is given in this section; it is an abridged version of the review chapter in the Hand Book compiled by one of the authors (Mitra and Mitra, 2011[a]). In particular, we review the multiple facets of current research and some of the major applications.

It is widely recognized news plays a key role in financial markets. The sources and volumes of news continue to grow. New technologies that enable automatic or semi-automatic news collection, extraction, aggregation and categorization are emerging. Further machine-learning techniques are used to process the textual input of news stories to determine quantitative sentiment scores. We consider the various types of news available and how these are processed to form inputs to financial models. We consider applications of news, for prediction of abnormal returns, for trading strategies, for diagnostic applications as well as the use of news for risk control. There is a strong yet complex relationship between market sentiment and news. The arrival of news continually updates an investor's understanding and knowledge of the market and influences investor sentiment. There is a growing body of research literature that argues media influences investor sentiment, hence asset prices, asset price volatility and risk (Tetlock, 2007; Da, Engleberg, and Gao, 2009; diBartolomeo and Warrick, 2005; Barber and Odean; Dzielinski, Rieger, and Talpsepp; Mitra, Mitra, and diBartolomeo, 2009, (chapter 7, chapter 11, chapter 13, Mitra and Mitra 2011[a])). Traders and other market participants digest news rapidly, revising and rebalancing their asset positions accordingly. Most traders have access to newswires at their desks. As markets react rapidly to news, effective models which incorporate news data are highly sought after. This is not only for trading and fund management, but also for risk control. Major news events can have a significant impact on the market environment and investor sentiment, resulting in rapid changes to the risk structure and risk characteristics of traded assets. Though the relevance of news is widely acknowledged, how to incorporate this effectively, in quantitative models and more generally within the investment decision-making process, is a very open question. In considering how news impacts markets, Barber and Odean note ''significant news will often affect investors' beliefs and portfolio goals heterogeneously, resulting in more investors trading than is usual'' (high trading volume). It is well known that volume increases on days with information releases (Bamber, Barron and Stober, 1997). It is natural to expect that the application of these news data will lead to improved analysis (such as predictions of returns and volatility). However, extracting this information in a form that can be applied to the investment decision-making process is extremely challenging. News has always been a key source of investment information. The volumes and sources of news are growing rapidly. In increasingly competitive markets investors and traders need to select and analyse the relevant news, from the vast amounts available to them, in order to make ''good'' and timely decisions. A human's (or even a group of humans') ability to process this news is limited. As computational capacity grows, technologies are emerging which allow us to extract, aggregate and categorize large volumes of news effectively. Such technology might be applied for quantitative model construction for both high-frequency trading and low-frequency fund rebalancing. Automated news analysis can form a key component driving algorithmic trading desks' strategies and execution, and the traders who use this technology can shorten the time it takes them to react to 'breaking news stories' (that is, reduce latency times).

News Analytics (NA) technology can also be used to aid traditional non-quantitative fund managers in monitoring the market sentiment for particular stocks, companies, brands and sectors. These technologies are deployed to automate filtering, monitoring and aggregation of news, in addition to helping free managers from the minutiae of repetitive analysis, such that they are able to better target their reading and research. NA technologies also reduce the

burden of routine monitoring for fundamental managers. The basic idea behind these NA technologies is to automate human thinking and reasoning. Traders, speculators and private investors anticipate the direction of asset returns as well as the size and the level of uncertainty (volatility) before making an investment decision. They carefully read recent economic and financial news to gain a picture of the current situation. Using their knowledge of how markets behaved in the past under different situations, people will implicitly match the current situation with those situations in the past most similar to the current one. News analytics seeks to introduce technology to automate or semi-automate this approach. By automating the judgement process, the human decision maker can act on a larger, hence more diversified, collection of assets. These decisions are also taken more promptly (reducing latency). Automation or semi-automation of the human judgement process widens the limits of the investment process. Leinweber (2009) refers to this process as intelligence amplification (IA).

As shown in Figure 5.1 news data are an additional source of information that can be harnessed to enhance (traditional) investment analysis. Yet it is important to recognize that NA in finance is a multi-disciplinary field which draws on financial economics, financial engineering, behavioural finance and artificial intelligence (in particular, natural language processing).
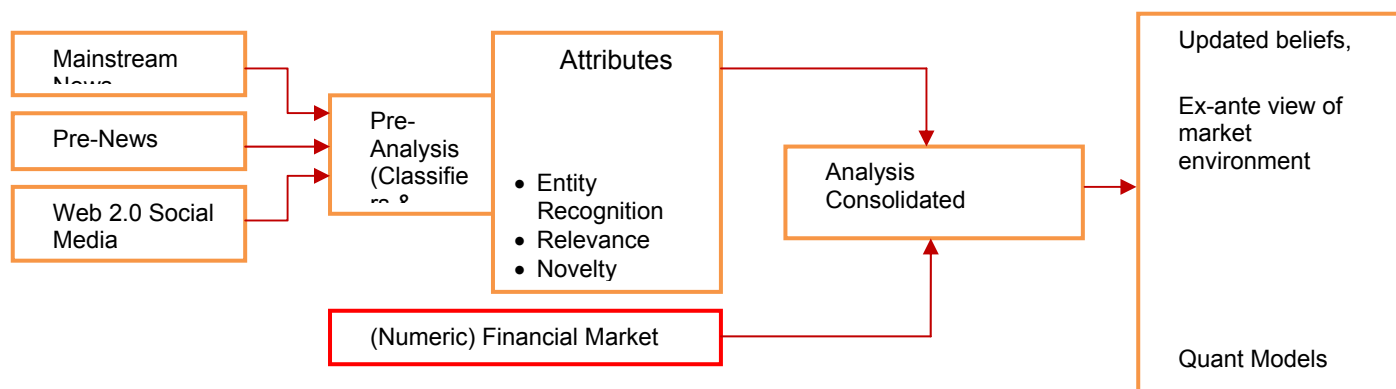

Figure 5.1   An outline of information flow and modelling architecture

5.2 News data sources

In this section we consider the different sources of news and information flows which can be applied for updating (quantitative) investor beliefs and knowledge. Leinweber (2009) distinguishes the following broad classifications of news (informational flows).

1. News This refers to mainstream media and comprises the news stories produced by reputable sources. These are broadcast via newspapers, radio and television. They are also delivered to traders' desks on newswire services. Online versions of newspapers are also progressively growing in volume and number.

2. Pre-news This refers to the source data that reporters research before they write news articles. It comes from primary information sources such as Securities and Exchange Commission reports and filings, court documents and government agencies. It also includes scheduled announcements such as macroeconomic news, industry statistics, company earnings reports and other corporate news.

3. Web 2.0 and social media These are blogs and websites that broadcast ''news'' and are less reputable than news and pre-news sources. The quality of these varies significantly. Some may be blogs associated with highly reputable news providers and reporters (for example, the blog of BBC's Robert Peston). At the other end of the scale some blogs may lack any substance and may be entirely fueled by rumour. Social media websites fall at the lowest end of the reputation scale. Barriers to entry are extremely low and the ability to publish ''information'' easy. These can be dangerously inaccurate sources of information. At a minimum they may help us identify future volatility. Individual investors pay relatively more attention to the second two sources of news than institutional investors. Information from the web may be less reliable than mainstream news. However, there may be ''collective intelligence'' information to be gleaned. That is, if a large group of people have no ulterior motives, then their collective opinion may be useful (Leinweber, 2009, Ch. 10).

There are services which facilitate retrieval of news data from the web. For example, Google Trends is a free but limited service which provides an historical weekly time series of the popularity of any given search term. This search engine reports the proportion of positive, negative and neutral stories returned for a given search. The Securities and Exchange Commission (SEC) provides a lot of useful pre-news. It covers all publicly traded companies (in the US). The Electronic Data Gathering, Analysis and Retrieval (EDGAR) system was introduced in 1996 giving basic access to filings via the web (see http://www.sec.gov/edgar.shtml). Premium access gave tools for analysis of filing information and priority earlier access to the data. In 2002 filing information was released to the public in real time. Filings remain unstructured text files without semantic web and XML output, though the SEC are in the process of upgrading their information dissemination. High-end resellers electronically dissect and sell on relevant component parts of filings. Managers are obliged to disclose a significant amount of information about a company via SEC filings. This information is naturally valuable to investors. Leinweber introduces the term ''molecular search: the idea of looking for patterns and changes in groups of documents.'' Such analysis/information is scrutinized by researchers/analysts to identify unusual corporate activity and potential investment opportunities. However, mining the large volume of filings, to find relationships, is challenging. Engleberg and Sankaraguruswamy (2007) note the EDGAR database has 605 different forms and there were 4,249,586 filings between 1994 and 2006. Connotate provides services which allow customized automated collection of SEC filing information for customers (fund managers and traders). Engleberg and Sankaraguruswamy (2007) consider how to use a web crawler to mine SEC filing information through EDGAR.

Financial news can be split into regular synchronous, that is, anticipated announcements (scheduled or expected news) and event-driven asynchronous news items (unscheduled or unexpected news). Mainstream news, rumours, and social media normally arrive asynchronously in an unstructured textual form. A substantial portion of pre-news arrives at pre-scheduled times and generally in a structured form. Scheduled (news) announcements often have a well-defined numerical and textual content and may be classified as structured data. These include macroeconomic announcements and earnings announcements. Macroeconomic news, particularly economic indicators from the major economies, is widely used in automated trading. It has an impact in the largest and most liquid markets, such as foreign exchange, government debt and futures markets. Firms often execute large and rapid trading strategies. These news events are normally well documented, thus thorough back testing of strategies is feasible. Since indicators are released on a precise schedule, market participants can be well prepared to deal with them. These strategies often lead to firms fighting to be first to the market; speed and accuracy are the major determinants of success. However, the technology requirements to capitalize on events are substantial. Content publishers often specialize in a few data items and hence trading firms often multisource their data. Thomson Reuters, Dow Jones, and Market News International are a few leading content service providers in this space. Earnings are a key driving force behind stock prices. Scheduled earnings announcement information is also widely anticipated and used within trading

strategies. The pace of response to announcements has accelerated greatly in recent years (see Leinweber, 2009, p. 104–105). Wall Street Horizon and Media Sentiment (see Munz, 2010) provide services in this space. These technologies allow traders to respond quickly and effectively to earnings announcements.

Event-driven asynchronous news streams in unexpectedly over time. These news items usually arrive as textual, unstructured, qualitative data. They are characterized as being non-numeric and difficult to process quickly and quantitatively. Unlike analysis based on quantified market data, textual news data contain information about the effect of an event and the possible causes of an event. However, to be applied in trading systems and quantitative models they need to be converted to a quantitative input time-series. This could be a simple binary series where the occurrence of a particular event or the publication of a news article about a particular topic is indicated by a one and the absence of the event by a zero. Alternatively, we can try to quantify other aspects of news over time. For example, we could measure news flow (volume of news) or we could determine scores (measures) based on the language sentiment of text or determine scores (measures) based on the market's response to particular language. It is important to have access to historical data for effective model development and back testing. Commercial news data vendors normally provide large historical archives for this purpose. The details of historic news data for global equities provided by RavenPack and Thomson Reuters NewsScope are summarized in Section 1.A (the appendix on p. 25 Mitra and Mitra, 2011[b]).

5.3 Pre-analysis of news data: creating meta data

Collecting, cleaning and analysing news data is challenging. Major news providers collect and translate headlines and text from a wide range of worldwide sources. For example, the Factiva database provided by Dow Jones holds data from 400 sources ranging from electronic newswires, newspapers and magazines.

We note there are differences in the volume of news data available for different companies. Larger companies (with more liquid stock) tend to have higher news coverage/news flow. Moniz, Brar, and Davis (2009) observe that the top quintile accounts for 40% of all news articles and the bottom quintile for only 5%. Cahan, Jussa, and Luo (2009) also find news coverage is higher for larger cap companies.

Classification of news items is important. Major newswire providers tag incoming news stories. A reporter entering a story on to the news systems will often manually tag it with relevant codes. Further, machine-learning algorithms may also be applied to identify relevant tags for a story. These tags turn the unstructured stories into a basic machine readable form. The tags are often stored in XML format. They reveal the story's topic areas and other important metadata. For example, they may include information about which company a story is about. Tagged stories held by major newswire providers are also accurately time-stamped. The SEC is pushing to have companies file their reports using XBRL (eXtensible Business Reporting Language). Rich Site Summary (RSS) feeds (an XML format for web content) allow customized, automated analysis of news events from multiple online sources. Tagged news stories provide us with hundreds of different types of events, so that we can effectively use these stories. We need to distinguish what types of news are relevant for a given model (application). Further, the market may react differently to different types of news. For example, Moniz, Brar, and Davis (2009) find the market seems to react more strongly to corporate earnings-related news than corporate strategic news. They postulate that it is harder to quantify and incorporate strategic news into valuation models, hence it is harder for the market to react appropriately to such news.

Machine-readable XML news feeds can turn news events into exploitable trading signals since they can be used relatively easily to back-test and execute event study-based strategies (see Kothari and Warner, 2005; Campbell, Lo, and MacKinlay, 1996 for in-depth reviews of event

study methodology). Leinweber (Chapter 6, Mitra and Mitra 2011[a]) uses Thomson Reuters tagged news data to investigate several news-based event strategies. Elementized news feeds mean the variety of event data available is increasing significantly. News providers also provide archives of historic tagged news which can be used for back-testing and strategy validation. News event algorithmic trading is reported to be gaining acceptance in industry (Schmerken, 2006).

To apply news effectively in asset management and trading decisions we need to be able to identify news which is both relevant and current. This is particularly true for intraday applications, where algorithms need to respond quickly to accurate information. We need to be able to identify an "information event"; that is, we need to be able to distinguish those stories which are reporting on old news (previously reported stories) from genuinely "new" news. As would be expected, Moniz, Brar, and Davis (2009) find markets react strongly when "new" news is released. Tetlock, Saar-Tsechansky, and Macskassy (2008) undertake an event study which illustrates the impact of news on cumulative abnormal returns (CARs).

Method and types of score (meta data)

TIMESTAMP_UTC: The date/time (yyyy-mm-dd hh :mm: ss.sss) at which the news item was received by RavenPack servers in Coordinated Universal Time (UTC).

COMPANY: This field includes a company identifier in the format ISO_CODE/TICKER. The ISO_CODE is based on the company's original country of incorporation and TICKER on a local exchange ticker or symbol. If the company detected is a privately held company, there will be no ISO_CODE/TICKER information, COMPANY_ID.

ISIN: An International Securities Identification Number (ISIN) to identify the company referenced in a story. The ISINs used are accurate at the time of story publication. Only one ISIN is used to identify a company, regardless of the number of securities traded for any particular company. The ISIN used will be the primary ISIN for the company at the time of the story.

COMPANY_ID: A unique and permanent company identifier assigned by RavenPack. Every company tracked is assigned a unique identifier comprised of six alphanumeric characters. The RP_COMPANY_ID field consistently identifies companies throughout the historical archive. RavenPack's company detection algorithms find only references to companies by information that is accurate at the time of story publication (point-in-time sensitive).

RELEVANCE: A score between 0 and 100 that indicates how strongly related the company is to the underlying news story, with higher values indicating greater relevance. For any news story that mentions a company, RavenPack provides a relevance score. A score of 0 means the company was passively mentioned while a score of 100 means the company was predominant in the news story. Values above 75 are considered significantly relevant. Specifically, a value of 100 indicates that the company identified plays a key role in the news story and is considered highly relevant (context aware).

CATEGORIES: An element or "tag" representing a company-specific news announcement or formal event. Relevant stories about companies are classified in a set of **predefined event categories** following the RavenPack taxonomy. When applicable, the role played by the company in the story is also detected and tagged. RavenPack automatically detects key news events and identifies the role played by the company. Both the topic and the company's role in the news story are tagged and categorized. For example, in a news story with the headline "IBM Completes Acquisition of Telelogic AB" the category field includes the tag acquisition-acquirer (since IBM is involved in an acquisition and is the acquirer company). Telelogic would

receive the tag acquisition/acquire in its corresponding record since the company is also involved in the acquisition but as the acquired company.

ESS—EVENT SENTIMENT SCORE: A granular score between 0 and 100 that represents the news sentiment for a given company by measuring various proxies sampled from the news. The score is determined by systematically matching stories typically categorized by financial experts as having short-term positive or negative share price impact. The strength of the score is derived from training sets where financial experts classified company-specific events and agreed these events generally convey positive or negative sentiment and to what degree. Their ratings are encapsulated in an algorithm that generates a score range between 0 and 100 where higher values indicate more positive sentiment while values below 50 show negative sentiment.

ENS—EVENT NOVELTY SCORE: A score between 0 and 100 that represents how ''new'' or novel a news story is within a 24-hour time window. The first story reporting a categorized event about one or more companies is considered to be the most novel and receives a score of 100. Subsequent stories within the 24-hour time window about the same event for the same companies receive lower scores.

News Analytics and market sentiment : Impact on Liquidity

News influences and formulates sentiment; sentiments move markets. The crash of 1987 was one such sentiment forming event in the recent past. Since 2003 equity markets have grown steadily, but at the end of 2007 it started to decline and there was a dip in the sentiment. Over January 2008 market sentiment deteriorated further, driven by a few key events. In the US, George Bush announced a stimulus plan for the economy and the Federal Reserve made cuts in the interest rate by 75 basis points, the largest since 1984. In Europe, Societe Generale was hit by the scandal of the rogue trader Jerome Kerviel. In September-October 2008 further events in the finance sector impacted the market: Lehman Brothers filed for bankruptcy, Bank of America announced the purchase of Merrill Lynch, the Federal Reserve announced the rescue of AIG, under the guidance of the UK Government Lloyds Bank took over HBOS. These news events had a devastating **impact on market liquidity**.

Market sentiment influences: price, volatility, liquidity

Financial markets are characterised by two leading measures: (i) stock price returns and (ii) stock price volatility. In the context of trading, a third aspect, namely, (iii) liquidity is seen to be equally important. There is a strong relationship between news flows and volatility. To the extent that a broad market or a particular security becomes more volatile, it can be expected that liquidity providers will demand greater compensation for risk by widening bid/asked spreads. This is confirmed in a recent research reported by Gross-Klussmann et al. (2011) who conclude that by capturing dynamics and cross-dependencies in the vector autoregressive modelling framework they find the strongest effect in volatility and cumulative trading volumes. Bid-ask spreads, trade sizes and market depth may not directly react to news; but they do so indirectly through the cross dependencies to volume and volatility, and the resulting spillover effects. There is a strong distinction between "news" and "announcements" in terms of liquidity. If information comes to the financial markets as an "announcement" (e.g. the scheduled publication of an economic statistic, or a company's period results), market participants will have anticipated the announcement and formulated action plans conditional on the revealed content. Since everyone is prepared for the announcement market participants can act quickly and liquidity is maintained. On the other hand, if a "news" item (fully unanticipated) is revealed to financial market participants, they will need some time to assess the meaning of the announcement and formulate appropriate actions. During such periods of contemplation, traders are unwilling to trade and liquidity dries up. If the news item is of extreme importance (e.g. 9/11), it may take several days for conditions to return to normal. Regulators and

exchanges respond to such liquidity "holes" by suspending trading for short periods in particular securities or markets. There is a vast literature on the impact of anticipated earnings announcements; in contrast there are very few studies on the intraday firm-specific news. Berry and Howe (1994) in a study links intraday market activity to an aggregated news flow measure, that is, the number of news items. Kalev et al. (2004) and Kalev et al. (2011) report a positive relationship between the arrival of intraday news and the volatility of a given stock, the market index and the index futures respectively. Mitchell and Mullherin (1994) and Ranaldo (2008) consider the impact of news on intraday trading activities.

News enhanced predictive analysis models

The Hand Book compiled by one of the authors Mitra (2011) reports studies which cover stock returns and volatility in response to news; however, none of these studies are either in the context of high frequency or consider the impact on liquidity. We therefore turn to the study by Gross-Klussmann et al. (2011) as they consider the impact of intra-day news flow. These authors consider an interesting research problem: 'are there significant and theory-consistent market reactions in high-frequency returns, volatility and liquidity to the intra-day news flow?' The authors set out to answer this question by applying a predictive analysis model (in this case an event study model) and use the news data feed provided by Thomson Reuters News analytics sentiment engine. These authors conclude that the release of a news item significantly increases bid-ask spreads but does not necessarily affect market depth. Hence, liquidity suppliers predominantly react to news by revising quotes and not by order volumes. This is well supported by asymmetric information based market microstructure theory (Easley and O'Hara, 1992) where specialists try to overcompensate for possible information asymmetries. Although there are no designated market makers in an electronic market, the underlying mechanism is similar to that of a non-electronic market: Liquidity suppliers reduce their order aggressiveness in order to avoid being picked off (that is, selected adversely) by better informed traders. For earnings announcements, such effects are also reported by Krinsky and Lee (1996). Overall, the authors find that the dynamic analysis strongly confirms the unconditional effects discussed above, and that volatility and trading volume are most sensitive to news arrival.

We generalise this approach and propose a modelling framework which closely follows the paradigm of event studies and is shown in Figure 6.1
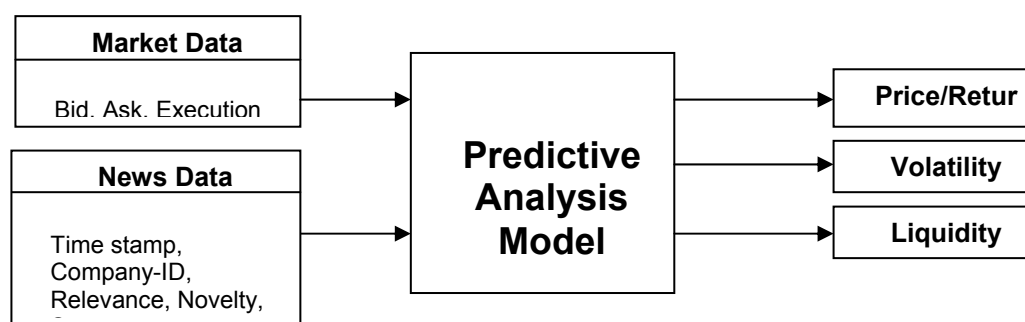

Figure 6.1 Architecture of predictive analysis model

The **input** to the Predictive analysis model is made up of

(i) **Market data** (bid, ask, execution price , time bucket)
(ii) **News data** suitably pre-analysed and turned into **meta data**
(time stamp, company-ID, relevance, novelty, sentiment score , event category...)

The **output** is designed to determine **state** of the stock/market (returns, volatility, liquidity).

News analytics and its application to trading

The automated sentiment scores (computed by using natural language processing, text mining and AI classifiers see section 5) are finding applications in investment decisions and trading. Two major content vendors of news analytics data, namely, (i) Thomson Reuters and (ii) RavenPack provide web posting of white papers and case studies; see A Team (2010) and RavenPack (2011), respectively. In this section we consider the growing influence of news analytics to investment management and manual trading as well as automated, that is, computer mediated algorithmic trading. In the discussions and conclusions presented in section 8 we provide a critical evaluation of the issues surrounding the interaction between manual and automated trading.

7.1   Trading by institutional investors and retail investors

Barber and Odean in their landmark paper (Barber and Odean, 2011) report the buying behaviour of individual (retail) investors as well as those of professional money managers. The study is based on substantial data (78,000 households' investment activities between 1991 and 1996) collected from a leading brokerage house. The authors observe that retail investors show a propensity to buy attention grabbing stocks (impact of news stories). They conclude that this is more driven by emotional behaviour of the investor than based on a rational analysis of investment opportunities. By and large such trades lead to losses for the retail investors. The institutional investors in contrast tend to make better use of information (flowing from news), in particular they use predictive analysis tools thus enhancing their fundamental analysis. Leinweber and Sisk (2011) describe a study in which they use pure news signals as indicators for buy signals. Through portfolio simulation of test data over the period 2006–2009 they find evidence of exploitable alpha using news analytics. The quantitative research team at Macquarie Securities (see Moniz et al, 2011) report on an empirical study where they show how news flow can be exploited in existing momentum strategies by updating earnings forecast ahead of analysts' revisions after news announcements. Cahan et al (2010) who use Thomson Reuters news data report similar results; these studies have many similarities given that Cahan and the team moved from Macquarie securities to Deutsche Bank in 2009. Macquarie securities and Deutsche Bank offer these news enhanced quant analysis services to their institutional clients. Other examples of applying NA in investment management decisions such as identifying sentiment reversal of stocks (see Kittrell, 2011) are to be found in Mitra and Mitra (2011).

7.2   News analytics applied to automated trading

The topic of automated algorithmic trading is treated as a 'black art' by its practitioners, that is, hedge funds and proprietary trading desks. As we stated in the introduction, section 1, even the content vendors are unwilling to reveal information about organizations which utilize NA in algorithmic trading. Given a trade order the execution by a strategy such as volume weighted average price (VWAP) is designed to minimize the market impact (see Kissell and Glantz, 2003). Almgren and Chriss (2000) in a landmark paper discuss the concept and the model for optimal execution strategies. In these models for execution the implicit assumption is that for a stock there is no price spike which often follows some anticipated news (announcements) or an unexpected news event. Aldridge (2010) in her book introduces the following categories of automated arbitrage trading strategies, namely, event arbitrage, statistical arbitrage including liquidity arbitrage. Of these the first: event arbitrage is based on the response of the market to an

information event, that is, a macro-economic announcement or a strategic news release. Event arbitrage strategies follow a three-stage development process:

    (i)   identification of the dates and times of past events in historical data

    (ii) computation of historical price changes at desired frequencies pertaining to securities

          of interest and the events identified in step-1 above

    (iii) estimation of expected price responses based on historical price behaviour surrounding the past events

The event arbitrage strategy is based on events surrounding news release about economic activity, market disruption or anything else that impact the market price. A tenet of efficient market hypothesis is that price adjusts to new information as soon as this becomes available. In practice market participants form expectations well ahead of the release of the announcements and the associated figures. For the FX market the study by Almeida, Goodhart and Payne (1998) find that for USD/DEM new announcements pertaining to the US employment and trade balance were significant predictors of the exchange rates. For a discussion of statistical arbitrage including liquidity arbitrage we refer the readers to Aldridge (2010). Taking into consideration the above remarks we have encapsulated the information flow and computational modelling architecture for news enhanced algorithmic trading as shown in Figure 7.1.
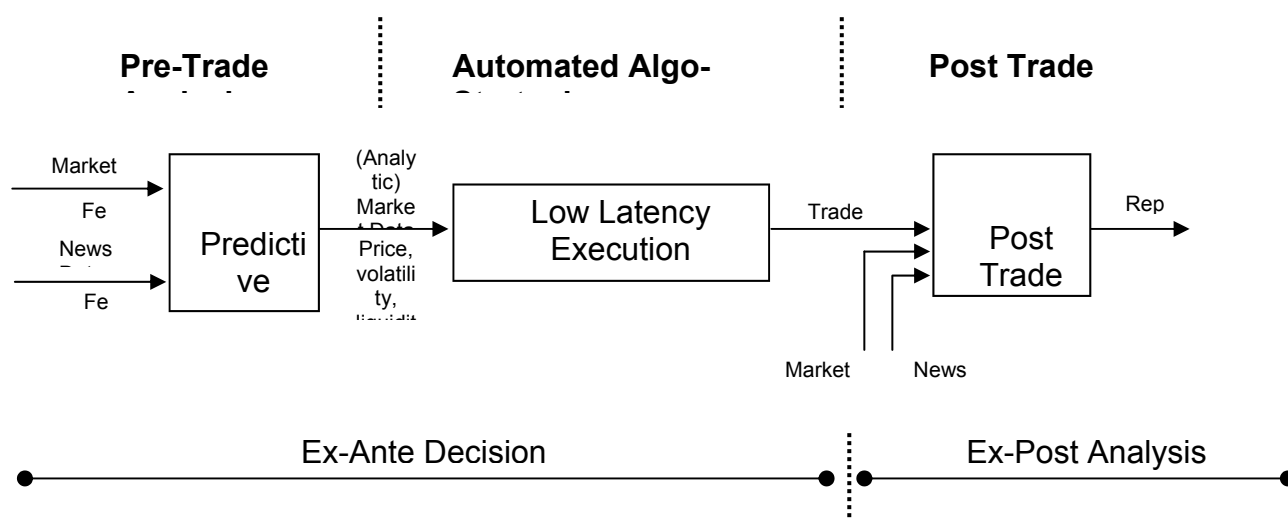

Figure 7.1 Information flow and computational architecture for automated trading

In the pre-trade analysis the predictive analysis tool brings together and consolidates market data feed and the news data feed. The output of the model goes into automated algorithm trading tools; these are normally low latency automatic trading algorithms (algos). Finally the outputs of these algorithms take the form of automatic execution orders. Whereas pre-trade analysis and the algos constitute ex-ante automatic decision tool, the results are evaluated using a paradigm of ex-post analysis. We finally note that Brown (2011) suggest use of news analytics to 'circuit breakers and wolf detection' in automated trading strategies thereby enhancing the robustness and reliability of such systems.

Discussions

As the saying goes the genie is out of the bottle and cannot be put back. Automated trading is here to stay and increasingly dominate the financial markets; this can be seen from the trends illustrated in Figure 2.2. In this report we have first examined the asset classes which are suitable for automated trading and conclude these to be primarily Equity including ETFs and index futures, FX, and to a lesser extent commodities and fixed income instruments. We have then considered in a summary form market microstructure and liquidity and their role in price formation. We have examined the role of different market participants in trading and types of automated trading activities. Set against this back drop we have explored how automated analysis of informational contents of anticipated news events as well as non anticipated extraordinary news events impact both 'manual' and automated trading activities. Both automated algorithmic trading and news analytics are recently developed technologies. The interactions of these technologies are uncharted and rely upon artificial intelligence, information and communication technologies as well as behavioural finance. Computer mediated automated trading continue to grow in many venues where equities, futures, options and foreign exchange are traded. The research and adoption of automated analysis of newsfeed for trading leads to the enhancement of performance, yet should there be a positive feedback effect due to inclusion of news sentiment, this will lead to increase in market instablility; thus uncontrolled automated trading may become one of the drivers of extreme behaviour such as ' flash crash '. The challenge for the regulatory authorities is to understand the combined impact of these technologies and postulate regulations which can control volatility, improve the provision for liquidity and generally stabilize the market behaviour.

References:

A Team (2010) *Machine Readable News and Algorithmic Trading*. Thomson Reuters and Market News International, White Paper.

Acharya, V.V. and Pedersen, L.H. (2005) *Asset pricing with liquidity risk.* Journal of Financial Economics 77(2), 375–410.

Aitken, M. and Comerton-Forde, C. (2003) *How should liquidity be measured?* Pacific-Basin Finance Journal 11, 45–59.

Ait-Sahala, Y. and Yu, J. (2009) *High Frequency Market Microstructure Noise Estimates and Liquidity Measures.* The Annals of Applied Statistics 3(1), 422–457.

Aldridge, I. (2010) *High Frequency Trading: A Practical Guide to Algorithmic Strategies and Trading Systems*. John Wiley & Sons, New Jersey.

Almeida, A., Goodhart, C. and Payne, R. (1998) *The effect of macro-economic news on high frequency exchange rate behaviour.* Journal of Financial and Quantitative Analysis 33, 1–47.

Almgren, R. and Chriss, N. (2000) *Optimal Execution of Portfolio Transactions*. Journal of Risk 12, 5–39.

Amihud, Y. and Mendelson, H. (1986) *Asset Pricing and the Bid-Ask Spread*. Journal of Financial Econometrics 17, 223–249.

Amihud, Y. (2002) *Illiquidity and Stock returns: cross-section and time-series effects.* Journal of Financial Markets 5, 31–56.

Arnuk, S.L. and Saluzzi, J (2008) *Toxic equity trading order flow on Wall Street: the real force behind the explosion in volume and volatility* Link: http://www.themistrading.com/article_files/0000/0524/Toxic_Equity_Trading_on_Wall_Street_--_FINAL_2__12.17.08.pdf

Arnuk, S.L. and Saluzzi, J. (2009) *Latency Arbitrage: the real power behind predatory high frequency trading.* Link:http://www.themistrading.com/article_files/0000/0519/THEMIS_TRADING_White_Paper_--_Latency_Arbitrage_--_December_4__2009.pdf

Bamber, L.S., Barron, O.E. and Stober, T.L. (1997) *Trading volume and different aspects of disagreement coincident with earnings announcements.* The Accounting Review 72, 575–597.

Berry, T.D. and Howe, K.M. (1994) *Public information arrival.* The Journal of Finance 49(4), 1331–1346.

Black, F. (1971) *Towards a fully automated exchange, part I*. Financial Analysts Journal 27, 29–34.

Brown, R. (2011) *Incorporating News Analytics into Algorithmic Trading Strategies: Increasing the Signal-to-Noise Ratio.* In The Handbook of News Analytics in Finance, Chapter 14, John Wiley & Sons. (See reference 49.)

Cahan, R., Jussa, J. and Luo, Y. (2009) *Breaking news: how to use sentiment to pick stocks.* MacQuarie Research Report.

Cahan, R., Jussa, J. and Luo, Y. (2010) *Beyond the Headlines: Using News flow to Predict Stock Returns*, Deutsche Bank Quantitative Strategy Report, July 2010.

Campbell, J.Y., Lo, A.W. and MacKinlay, A.C. (1996) *The econometrics of financial markets.* Event Study Analysis, Chapter 4, Princeton University Press, Princeton, NJ.

Chordia, T., Roll, R. and Subrahmanyam, A. (2000) *Commonality in liquidity*. Journal of Financial Economics 56(1), 3–28.

Da, Z., Engleberg, J. and Gao, P. (2009) *In search of attention.* Working Paper, SSRN. Link: http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1364209

diBartolomeo, D. and Warrick, S. (2005) *Making covariance based portfolio risk models sensitive to the rate at which markets reflect new information.* In J. Knight and S. Satchell (Eds.), Linear Factor Models, Elsevier Finance.

Dzielinski, M., Rieger, M.O. and Talpsepp, T. (2011) *Volatility, asymmetry, news and private investors.* In The Handbook of News Analytics in Finance, Chapter 11, John Wiley & Sons. (See reference 49.)

Easley, D. and O'Hara, M. (1992) *Time and the process of security price adjustment.* Journal of Finance 47, 577–605.

Engleberg, J. and Sankaraguruswamy, S. (2007) *How to gather data using a web crawler: an application using SAS to research EDGAR.* Link: http://papers.ssrn.com/sol3/papers.cfm?abstractid=1015021&r

Francioni, R., Hazarika, S., Reck, M. and Schwartz, R.A. (2008) *Equity Market Microstructure: Taking Stock of What We Know*. Journal of Portfolio Management

Garman, M. (1976) *Market Microstructure*. Journal of Financial Economics 3, 257–275.

Goodhart, C.A.E. and O'Hara, M. (1997) *High Frequency Data in Financial Markets: Issues and Applications.* Journal of Empirical Finance 4, 73–114.

Greenwich Associates (2009) *High-frequency trading: lack of data means regulators should move slowly.*

   Link:http://www.greenwich.com/WMA/in_the_news/news_details/1,1637,1851,00.html?

Gross-Klussmann, A. and Hautsch, N (2011) *When machines read the news: using automated text analytics to quantify high frequency news-implied market reactions.* Journal of Empirical Finance 18, 321–340.

Haiss, P. (2010) *Bank Herding and Incentive Systems as Catalysts for the Financial Crisis.* The IUP Journal of Behavioural Finance 7(Nos. 1 &2), 30–58.

Harris, L. (2005) *Trading & Strategies*. Oxford University Press.

Hasbrouck, J. and Schwartz, R.A. (1988) *Liquidity and execution cost in equity markets.* The Journal of Portfolio Management 14, 10–16.

Hasbrouck, J. and Seppi, D.J. (2001) *Common factors in prices, order flows, and liquidity.* Journal of Financial Economics 59(3), 383–411.

Hui, B. and Heubel, B. (1984) *Comparative liquidity advantages among major U.S. stock markets.* Technical Report, DRI Financial Information Group Study Series No. 84081.

Kahneman, D. and Tversky, A. (1979) *Prospect Theory: An Analysis of Decision under Risk.* Econometrica 47(2), 263–292.

Kahneman, D. (2002) *Maps of bounded rationality: The [2002] Sveriges Riksbank Prize [Lecture] in Economic Sciences.*

   Link: http://nobelprize.org/nobel_prizes/economics/laureates/2002/kahneman-lecture.html

Kalev, P.S., Liu, W.M., Pham, P.K. and Jarnecic, E. (2004) *Public information arrival and volatility of intraday stock returns.* Journal of Banking and Finance 28(6), 1441–1467.

Kalev, P.S. and Duong, H.N. (2011) *Firm-specific news arrival and the volatility of intraday stock index and futures returns*. In The Handbook of News Analytics in Finance, Chapter 12, John Wiley & Sons. (See reference 49.)

Kissell, R. and Glantz, M. (2003) *Optimal Trading Strategies*. American Management Association, AMACOM.

Kittrell, J. (2011) *Sentiment reversals as buy signals*. In The Handbook of News Analytics in Finance, Chapter 9, John Wiley & Sons. (See reference 49.)

Krinsky, I. and Lee, J. (1996) *Earnings announcements and the components of the bid-ask spread.* Journal of Finance 51(4), 1523–1535.

Kothari, S.P. and Warner, J.B. (2005) *Econometrics of event studies.* In B. Espen Eckbo (Ed.), Handbook of Empirical Corporate Finance, Elsevier Finance.

Kyle, A. (1985) *Continuous auction and insider trading.* Econometrica 53, 1315–35.

Leinweber, D. (2009) *Nerds on Wall Street.* John Wiley & Sons.

Leinweber, D. and Sisk, J. (2011) *Relating news analytics to stock returns.* In The Handbook of News Analytics in Finance, Chapter 6, John Wiley & Sons. (See reference 49.)

Madhavan, A. (2000) *Market Microstructure: A Survey.* Journal of Financial Markets 3, 205–258.

Mitchell, M.L. and Mulherin, J.H. (1994) *The impact of public information on the stock market.* Journal of Finance 49, 923–950.

Mitra, L., Mitra, G. and diBartolomeo, D. (2009) *Equity portfolio risk (volatility) estimation using market information and sentiment.* Quantitative Finance 9(8), 887–895.

Mitra, L. and Mitra, G. (Editors) (2011)[a] *The Handbook of News Analytics in Finance.* John Wiley & Sons.

Mitra, L. and Mitra, G. (2011)[b] *Applications of news analytics in finance: a review.* The Handbook of News Analytics in Finance, Chapter 1, John Wiley & Sons. (See reference 49.)

Moniz, A., Brar, G., and Davies, C. (2009) *Have I got news for you.* MacQuarie Research Report.

Moniz, A., Brar, G., Davies, C. and Strudwick, A. (2011) *The impact of news flow on asset returns: an empirical study.* In The Handbook of News Analytics in Finance, Chapter 8, John Wiley & Sons.

Munz, M. (2010) *US markets: earnings news release - an inside look.* Paper presented at CARISMA Annual Conference. Link: http://www.optirisk-systems.com/papers/MarianMunz.pdf

O'Hara, M. (1995) *Market Microstructure Theory*. Blackwell Publishing, Malden, Massachussetts.

Quantitative Services Group LLC (2009) *QSG® study proves higher trading costs incurred for VWAP algorithms vs. arrival price algorithms, high frequency trading contributing factor.* Link: http://www.qsg.com/PDFReader.aspx?PUBID=722

Ranaldo, A. (2008) *Intraday market dynamics around public information disclosures.* In Stock Market Liquidity, Chapter 11, John Wiley & Sons, New Jersey.

RavenPack white papers (2011) Link:

  http://www.ravenpack.com/research/resources.html

Schmerken, I. (2006) *Trading off the news*. Wall Street and Technology. Link: http://www.wallstreetandtech.com/technology-risk-management/showArticle.jhtml

Shefrin, H. (2008) *A Behavioural Approach to Asset Pricing.* Academic Press.

Shiller, R. (2000) *Irrational Exuberance.* Princeton University Press.

Tetlock, P.C. (2007) *Giving content to investor sentiment: the role of media in the stock market.* Journal of Finance 62, 1139–1168.

Tetlock, P.C., Saar-Tsechansky, M. and Macskassy, S. (2008) *More than words: Quantifying language to measure firms' fundamentals.* Journal of Finance 63(3), 1437–1467.