b6f1260bee54f438e54f5e8d966ba6d2
ebrary

CHAPTER 1

How Modern Markets Differ from Those Past

b6f1260bee54f438e54f5e8d966ba6d2
ebrary

Structural change is not new to trading in financial instruments. If fact, it is the constancy of innovation that has helped drive the leadership of modern financial institutions. High-frequency trading (HFT) has stepped into the limelight over the past few years and delivered considerable operational improvements to the markets, most of which have resulted in lower volatility, higher market stability, better market transparency, and lower execution costs for traders and investors. This chapter of the book provides the overview of dramatic changes that precipitated in the securities markets over the past 50 years, and defines HFT and core strategies falling under the HFT umbrella.

b6f1260bee54f438e54f5e8d966ba6d2
ebrary

Over the past two decades, the demand for computer technology in consumer markets has led to significant drops in hardware prices across the board, as discussed in detail in Chapter 2 of this book. As a result, technology-enabled trading has become cost effective, and the ensuing investment into software has made trading platforms more accessible and more powerful. Additionally, the savings from lower errors in message transmission and data input, higher reliability of order execution, and continuity of business through computer code, deliver a business case for deepening financial firms' reliance on their technology systems. The escalating complexity of regulations also requires more advanced reporting capabilities that are becoming prohibitively expensive without substantial platforms. The lower cost base squeezes margins further, and this puts pressure on the traditional full-service model. Figures 1.1 and 1.2 illustrate the financial services landscape circa 1970s and today.

In the 1970s and earlier, the market participants were organizations and individuals now considered "traditional" players. As Figure 1.1 shows, on the portfolio management or "buy" side, the markets engaged

b6f1260bee54f438e54f5e8d966ba6d2
ebrary

Aldridge, Irene. Wiley Trading Ser. : High-Frequency Trading : A Practical Guide to Algorithmic Strategies and Trading Systems (2). Somerset, US: John Wiley & Sons, Incorporated, 2013. ProQuest ebrary. Web. 8 March 2017.
Copyright © 2013. John Wiley & Sons, Incorporated. All rights reserved.

b6f1260bee54f438e54f5e8d966ba6d2
ebrary

- Discretionary asset managers, including pension funds, mutual funds, and hedge funds.
- Retail flow, including individual "mom-and-pop" investors, and others with comparatively smaller capitalization.
- Manual speculators, individuals involved in intraday proprietary trading for their own account or for the account of their bank.

On the transaction facilitation, middle-men, or "sell" side, the markets supported

- Manual market makers (representatives of broker-dealers), taking short-term inventory risk, providing quotations to the buy side, and generally facilitating the buy-side trading for a fee.

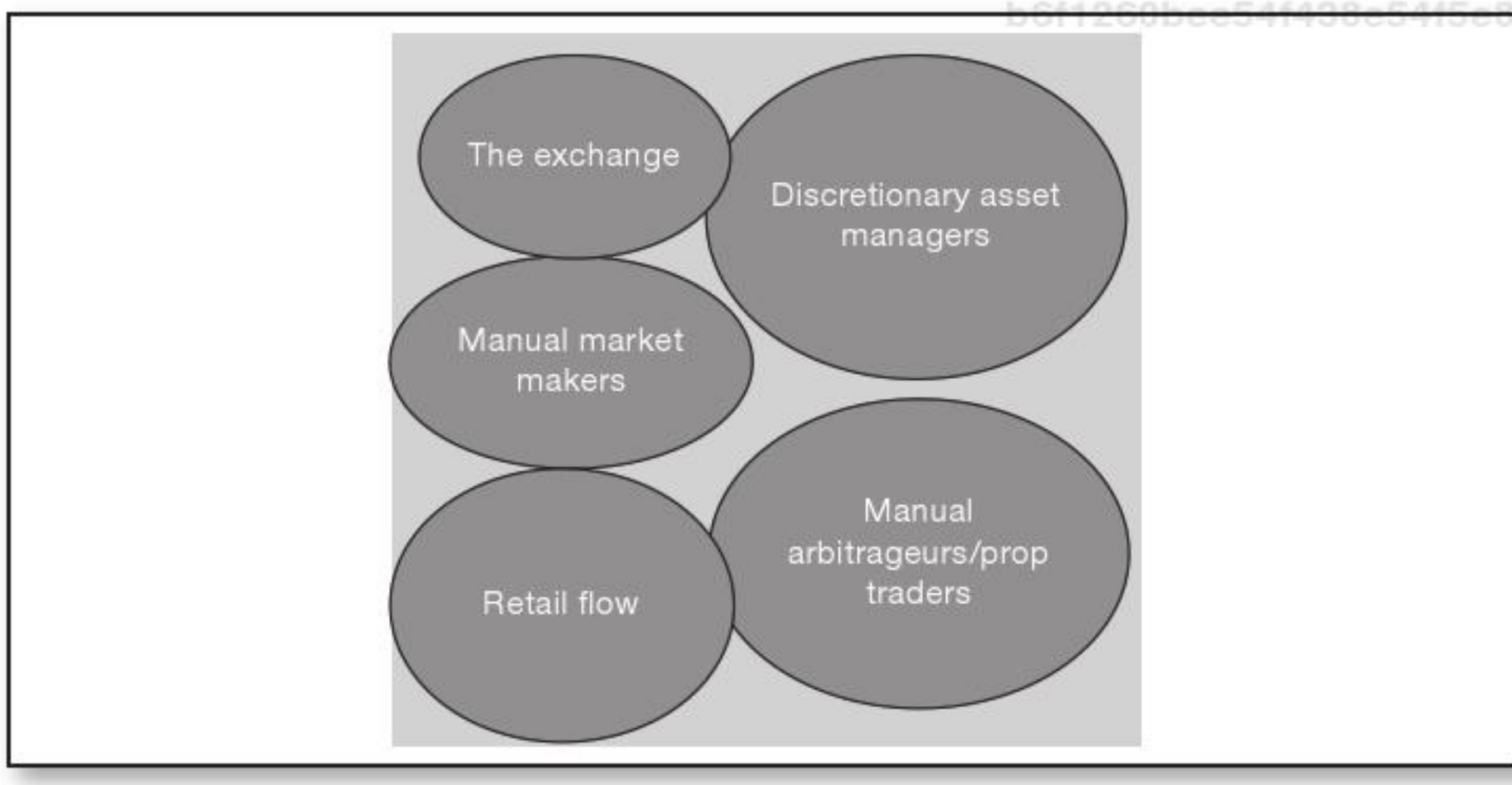


FIGURE 1.1 Financial Markets in the 1970s, before Electronization

6f1260bee54f438e54f5e8d966ba6d2
ebrary

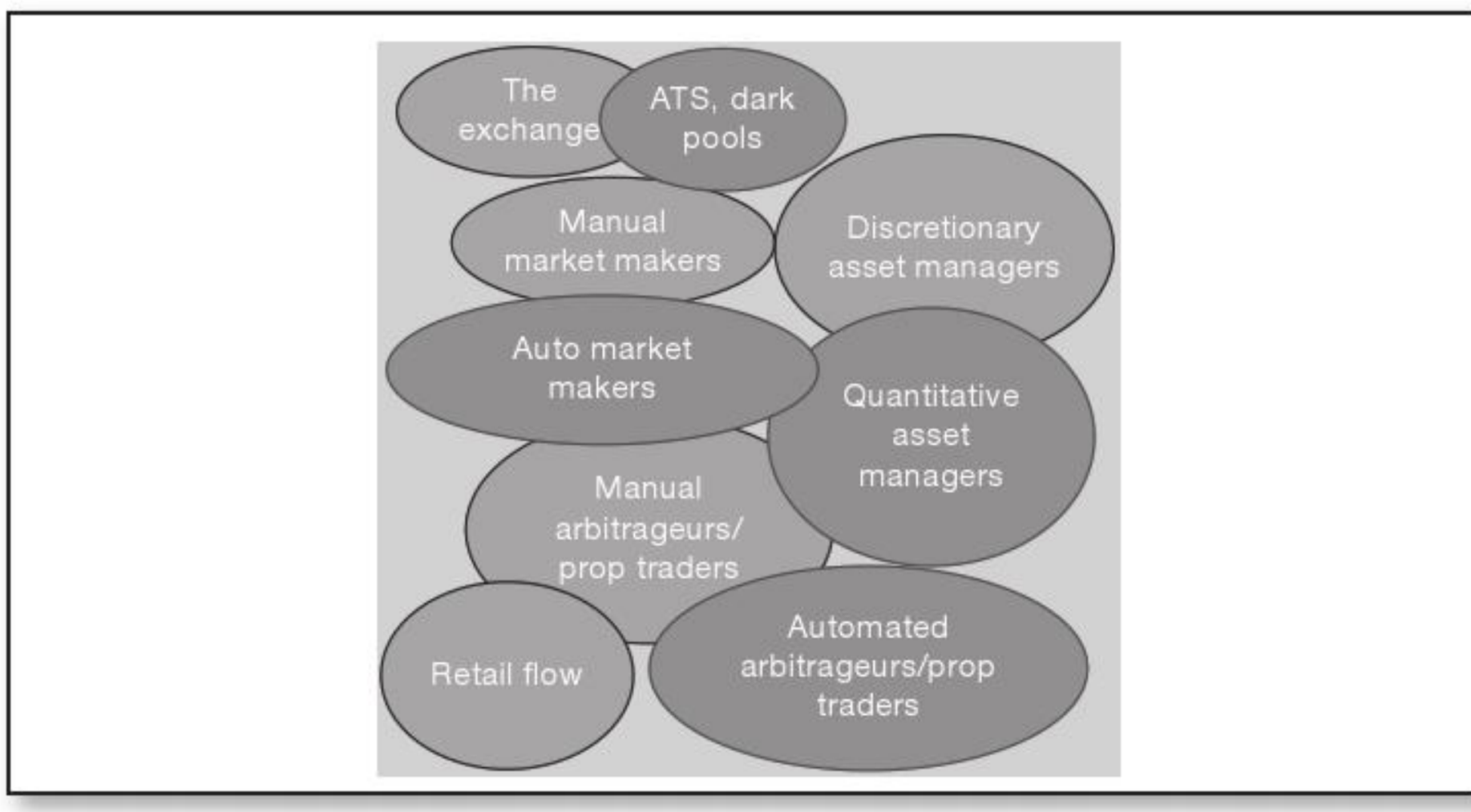


FIGURE 1.2 Financial Markets Today

b6f1260bee54f438e54f5e8d966ba6d2
ebrary

Aldridge, Irene. Wiley Trading Ser. : High-Frequency Trading : A Practical Guide to Algorithmic Strategies and Trading Systems (2). Somerset, US: John Wiley & Sons, Incorporated, 2013. ProQuest ebrary. Web. 8 March 2017.
Copyright © 2013. John Wiley & Sons, Incorporated. All rights reserved.

b6f1260bee54f438e54f5e8d966ba6d2
ebrary

- A single not-for-profit exchange in each asset class was established to curtail wild speculation by the exchange members and to lower transaction costs paid by the investors.[1]

The highly manual and therefore labor-intensive financial landscape of the 1970s was characterized by high transaction costs, leading to low turnover of securities; a high degree of error associated with manual processing of orders, and relatively high risk of trading, as traders predominantly relied on their experience and intuition as opposed to science in making their bets on the markets. Yet, the 1970s were also high margin businesses, with brokers receiving a large share of the spoils in the form of large commissions and, ultimately, the proverbial "fat-cat" bonuses to the tune of tens of millions of dollars.

Fast-forward to today's markets, illustrated in Figure 1.2: new entrants successfully compete using lean technology and science to hash out precise investing models, reshaping the markets in the process:

b6f1260bee54f438e54f5e8d966ba6d2
ebrary

- Quantitative money managers, such as mutual funds and hedge funds, are using the precise science of economics, finance, and the latest mathematical tools to chisel increasingly close forecasts of securities prices, improving profitability of their investments.
- Automated market makers, for example, broker-dealers and hedge funds, harness the latest technology, studies of market microstructure, and HFT to deliver low transaction costs, taking over market share from traditional broker-dealers.
- Automated arbitrageurs, such as statistical arbitrage hedge funds and proprietary traders, use quantitative algorithms, including high-frequency trading techniques, to deliver short-term trading profits.
- Multiple alternative trading venues, like new exchanges and dark pools, have sprung up to address market demand for affordable quality financial matching services.

b6f1260bee54f438e54f5e8d966ba6d2
ebrary

These innovations have changed the key characteristics of the markets, and largely for the better:

- The markets now enjoy vastly democratic access: due to proliferation of low-cost technology, anyone can trade in the markets and set quotes, a right formerly reserved to members of the exclusive connections-driven club of broker-dealers.
- Plummeting transaction costs keep money in investors' pockets; more on this later.
- Automated trading, order routing, and settlement deliver a new low degree of error.

[1] Most exchanges became not-for-profit only in the 1970s. From the time of their formation to the 1970s, however, the exchanges were very much for profit. In fact, the Buttonwood agreement of 1792 that laid foundation to the New York Stock Exchange, specified explicit profitability rules: no broker was allowed to charge less than 0.25 percent of the transaction volume, a staggering commission by today's standards.

b6f1260bee54f438e54f5e8d966ba6d2
ebrary

Aldridge, Irene. Wiley Trading Ser. : High-Frequency Trading : A Practical Guide to Algorithmic Strategies and Trading Systems (2). Somerset, US: John Wiley & Sons, Incorporated, 2013. ProQuest ebrary. Web. 8 March 2017.
Copyright © 2013. John Wiley & Sons, Incorporated. All rights reserved.

b6f1260bee54f438e54f5e8d966ba6d2
ebrary

The extreme competition among the new entrants and old incumbent market participants, however, has also resulted in reduced margins for broker-dealers, squeezing out technology-inefficient players.

The way trading is done has changed over time and these newer approaches affected the relative power of consumers and institutions. In the 1970s' marketplace, the trading process would often proceed as follows:

1. Brokers would deliver one-off trading ideas to their buy-side clients. The ideas were often disseminated via countless phone calls, were based on brokers' then-unique ability to observe markets in real time, and were generally required compensation in "soft-dollar" arrangements—if the customer decided to trade on the idea, he was expected to do so through the broker who produced the idea, and the customer would pay for the idea in the form of potentially higher broker commissions.
2. If and when the customer decided to trade on the idea, the customer would phone in the order to the broker or the broker's assistant. Such verbal orders frequently resulted in errors: the noise on the brokers' trading floors often impeded correct understanding of customer instructions.
3. After receiving a customer's order, the broker's next steps would depend on the size of the placed order: while large orders would be taken to the market right away (potentially in smaller parcels), smaller orders would sit on the broker's desk, waiting for other similar orders to fill up a "round lot"—the minimum order size executable on an exchange. Smaller customers were thus often at a disadvantage, waiting for execution of their orders while the favorable market price slipped away.
4. Once the order or several sequential orders comprised the order size acceptable to the broker, the broker would route the order to the appropriate exchange.
5. Next, human representatives of the exchange, known as "specialists," would match the order and send the trade acknowledgments back to the broker. It is well understood that the specialists often created preferential terms for some of their connections, at the expense of orders of others. Such behavior rewarded investment in connections and chummy networks, and resulted in exclusive Wall Street cliques capable of significant price discrimination for in-group versus out-of-group customers. Even though exchanges operated as not-for-profit organizations, influence peddling was common, and the markets were a long way away from anything resembling an equal playing field for all participants.
6. The broker notified the client of execution and collected his commissions and oversized bonuses. The brokers presided over the power of the markets and were compensated as kings.

Figure 1.3 illustrates the traditional investing process prevalent circa 1970s.

Fast-forward 40-something years ahead, and the balance of power has shifted. Customers have increased their expertise in quantitative analysis and are often better equipped for research than brokers. Brokers' area of expertise has decreased in scope from the all-encompassing sell-side research into securities behavior to a more narrow, albeit still important area of algorithmic execution designed to help clients

Aldridge, Irene. Wiley Trading Ser. : High-Frequency Trading : A Practical Guide to Algorithmic Strategies and Trading Systems (2). Somerset, US: John Wiley & Sons, Incorporated, 2013. ProQuest ebrary. Web. 8 March 2017.
Copyright © 2013. John Wiley & Sons, Incorporated. All rights reserved.

b6f1260bee54f438e54f5e8d966ba6d2
ebrary

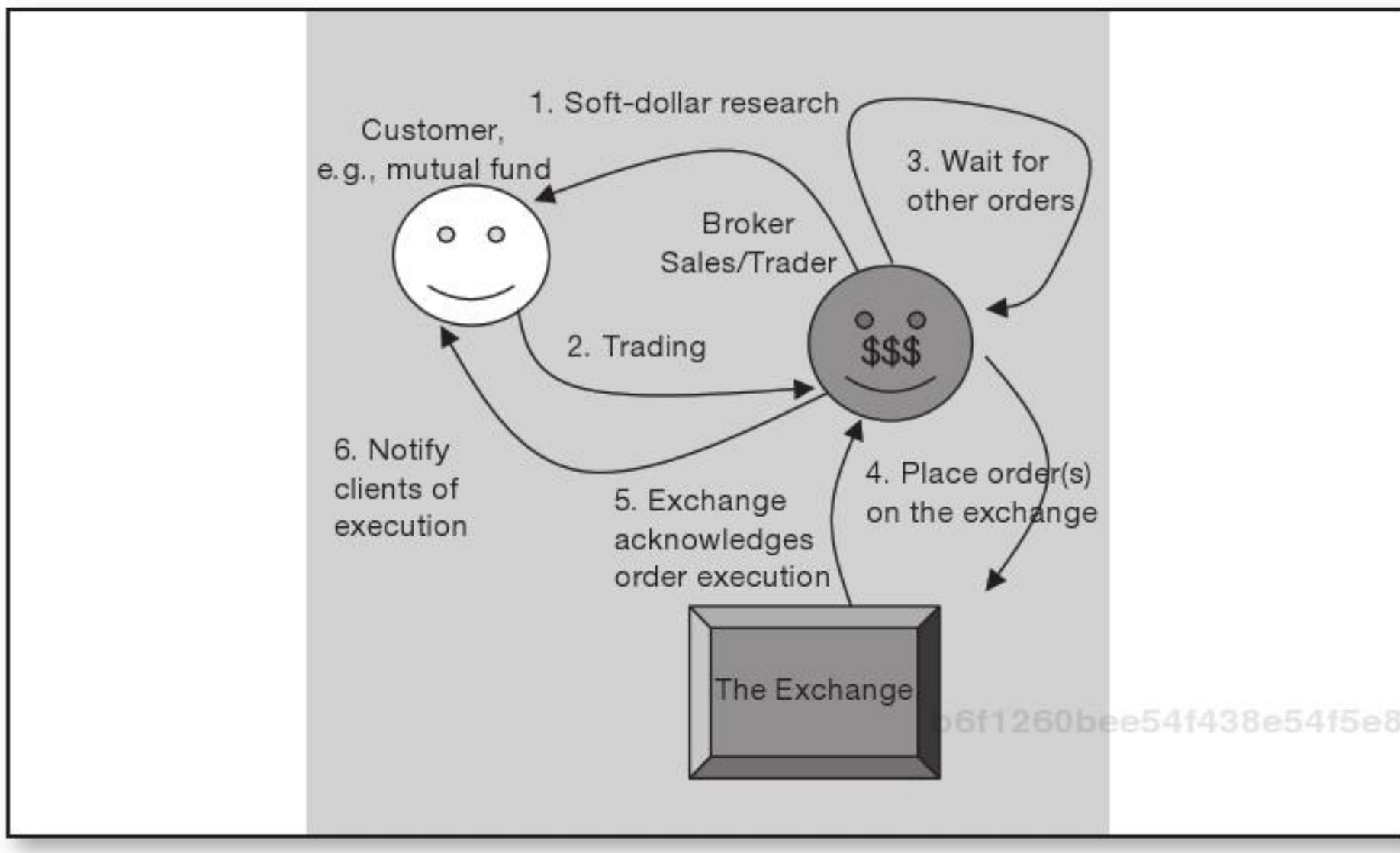


FIGURE 1.3 Broker-centric Investing Process Prevalent before Electronization

navigate the choppy intraday trading waters. With such buy-side investors, the market flow evolves according to the process in Figure 1.4:

1. Customers, not brokers, generate research based on forecasts of securities movements and their existing allocations, all within the quantitative portfolio management framework.
2. The customer places an order via electronic networks, greatly reducing errors and misunderstandings. The order instantaneously arrives on the broker's desktop.
3. The customer or the broker selects the broker's optimal execution algorithm designed to minimize the customer's execution costs and risk, speed up execution whenever possible, and minimize observability of the customer's trading actions.
4. Selected algorithm electronically parcels out the customer's order and routes the order slices to relevant exchanges and other trading venues.
5. Trading venues match the customer's order slices and acknowledge execution.
6. The broker sends the order acknowledgment back to the customer, and receives his considerably lower commission. (In 1997, the lowest broker commission on retail trades was offered by Merrill Lynch, and the commission was $70 per trade. Today, Interactive Brokers charges about $0.70 per trade, a 100-fold reduction in transaction costs available to clients.)

Some customers go even further and prefer to do away with broker service altogether, building their own execution algorithms, keeping a higher share of the profits. Plummeting costs of technology have enabled fast distribution of tick data to all interested parties, and now customers, not just brokers, can watch and time markets and generate short-term forecasts of market behavior. Customers taking the largely

Aldridge, Irene. Wiley Trading Ser. : High-Frequency Trading : A Practical Guide to Algorithmic Strategies and Trading Systems (2). Somerset, US: John Wiley & Sons, Incorporated, 2013. ProQuest ebrary. Web. 8 March 2017.
Copyright © 2013. John Wiley & Sons, Incorporated. All rights reserved.

b6f1260bee54f438e54f5e8d966ba6d2
ebrary

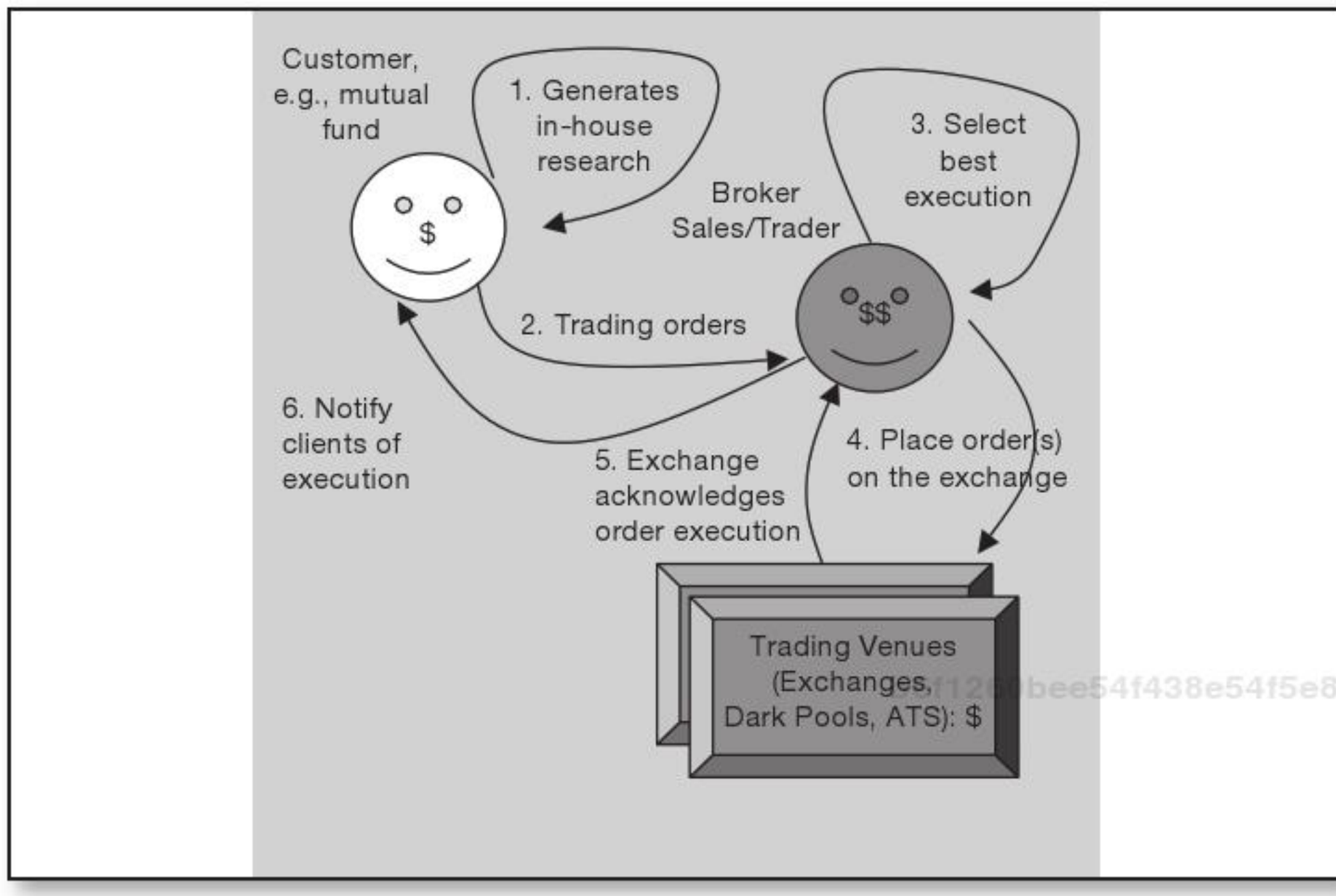


FIGURE 1.4 Modern Investing Process, Scenario 1: Brokers Provide Best Execution for Clients' Orders

broker-independent route are said to engage in "direct access" to the markets, and their execution process consists of the following steps:

1. The broker grants the customer a privilege to access the exchange directly for a negotiated per-trade or per-volume fee. To grant access, the broker may allow the customer to use the broker's own identification with a specific exchange. The customer's order routing systems then use the broker's identification in order messaging with the exchange.
2. Customer computer systems or human analysts generate a high- or low-frequency portfolio allocation decision that involves one or more trading orders.
3. Customer uses his own order splitting and routing algorithms to optimally place his orders directly with exchanges and other trading venues.
4. One or several exchanges and trading venues match the orders, acknowledg execution directly to client.
5. The broker receives settlement information and charges the client for the privilege of using the broker's direct access identification.

Figure 1.5 summarizes these steps.

Media, Modern Markets, and HFT

While the market-wide changes have disturbed the status quo on the broker-dealer side and squeezed many a broker out of business, the changes to the society at large have been mostly positive, depositing the saved dollars directly into investor pockets. Gone are the multimillion-dollar bonuses of many brokers taking phone orders

Aldridge, Irene. Wiley Trading Ser. : High-Frequency Trading : A Practical Guide to Algorithmic Strategies and Trading Systems (2). Somerset, US: John Wiley & Sons, Incorporated, 2013. ProQuest ebrary. Web. 8 March 2017.
Copyright © 2013. John Wiley & Sons, Incorporated. All rights reserved.

b6f1260bee54f438e54f5e8d966ba6d2
ebrary

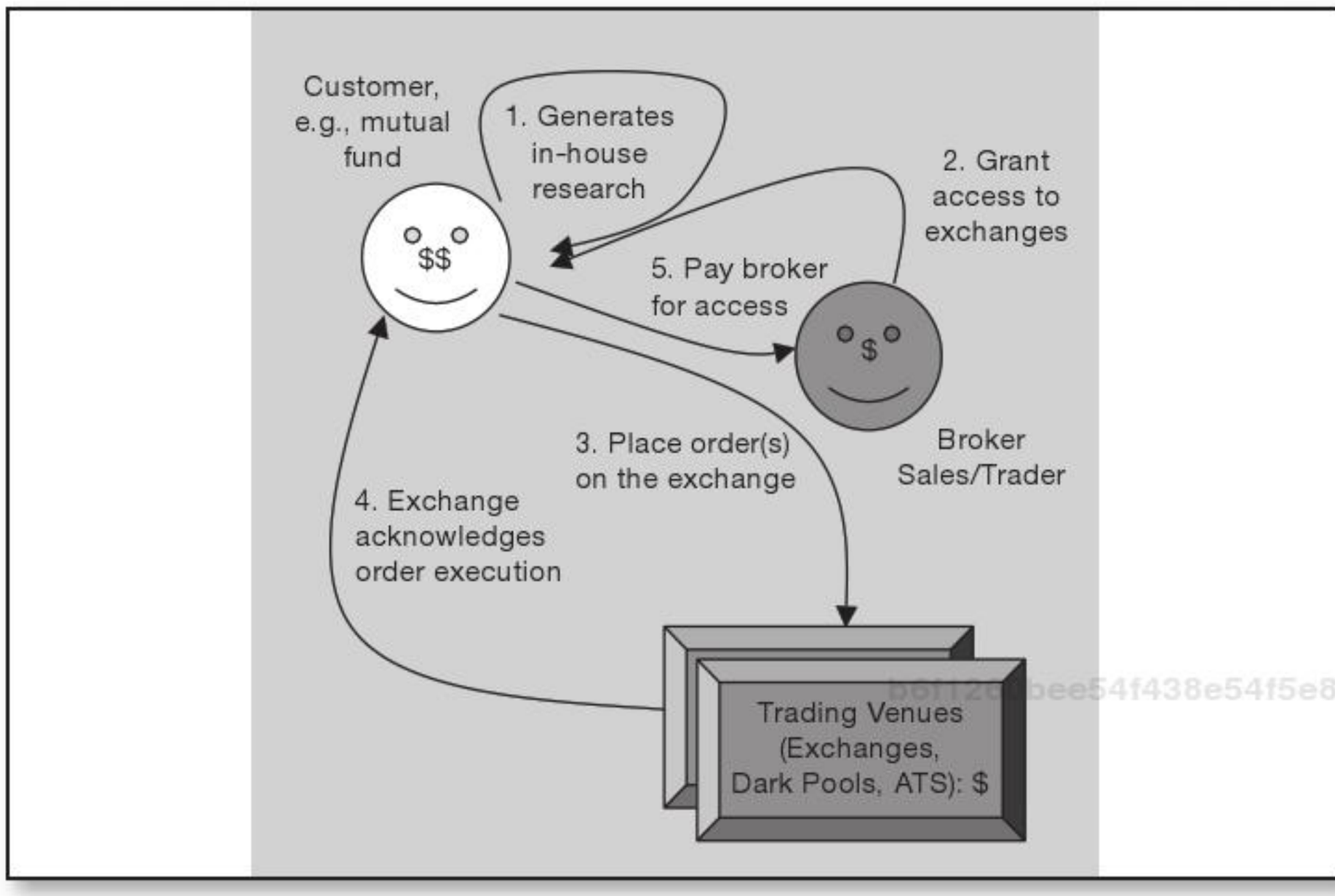


FIGURE 1.5 Modern Investing Process, Scenario 2: Clients Decide on Best Execution, Access Markets Directly

and watching markets on their computer screens. The money has been redirected to bank shareholders and end investors.

Clearly, not everyone is happy about such shifts in the industry, and the least happy bunch happens to be brokers losing their income to automation. Stripped of the ability to extract easy money out of investors' pockets, brokers have been the most vocal opponents of high-frequency trading. Brokers like Arnuk and Saluzzi (2012), for example, denounce automation, yet wax eloquent about those manual error-prone days when investors were not allowed on exchanges and brokers were the fat cats of the world.

Some brokers, whose lifestyle has been significantly reduced by technology, attempt to demonize HFT with an even more sinister goal in mind: they are seeking to lure in investors to their outfits still charging exorbitant transaction costs under the guise of protecting the poor investor lambs from the HFT predators. Investors should take time to compare costs of trading through a broker versus other available options. Chapters 5, 12, and 15 of this book provide specific information to enable low-frequency investors to estimate the risk of potentially adverse HFT, and take educated steps to managing said risk, without relying on self-serving hype of selected brokers who refuse to catch up on technical innovation, resorting to scare tactics at the expense of their clients instead. The remainder of this chapter is devoted to explaining the evolutionary nature of HFT and the definitions and overview of strategies that fall under the HFT umbrella.

HFT as Evolution of Trading Methodology

Brokers who speak loudly against the HFT tend to rely on technical analysis in making their decisions of when to enter or exit a position. Technical analysis was one of

Aldridge, Irene. Wiley Trading Ser. : High-Frequency Trading : A Practical Guide to Algorithmic Strategies and Trading Systems (2). Somerset, US: John Wiley & Sons, Incorporated, 2013. ProQuest ebrary. Web. 8 March 2017.
Copyright © 2013. John Wiley & Sons, Incorporated. All rights reserved.

b6f1260bee54f438e54f5e8d966ba6d2
ebrary

the earliest techniques that became popular with many traders and is, in many ways, a direct precursor to today's sophisticated econometrics and other HFT techniques.

Technical analysts came in vogue in the early 1910s and sought to identify recurring patterns in security prices. Many techniques used in technical analysis measure current price levels relative to the rolling moving average of the price, or a combination of the moving average and standard deviation of the price. For example, a technical analysis technique known as moving average convergence divergence (MACD) uses three exponential moving averages to generate trading signals. Advanced technical analysts would look at security prices in conjunction with current market events or general market conditions to obtain a fuller idea of where the prices may be moving next.

Technical analysis prospered through the first half of the twentieth century, when trading technology was in its telegraph and pneumatic-tube stages and the trading complexity of major securities was considerably lower than it is today. The inability to transmit information quickly limited the number of shares that changed hands, curtailed the pace at which information was incorporated into prices, and allowed charts to display latent supply and demand of securities. The previous day's trades appeared in the next morning's newspaper and were often sufficient for technical analysts to successfully infer future movement of the prices based on published information. In post-WWII decades, when trading technology began to develop considerably, technical analysis developed into a self-fulfilling prophecy.

b6f1260bee54f438e54f5e8d966ba6d2
ebrary

If, for example, enough people believed that a "head-and-shoulders" pattern would be followed by a steep sell-off in a particular instrument, all the believers would place sell orders following a head-and-shoulders pattern, thus indeed realizing the prediction. Subsequently, institutional investors have moved to high-frequency econometric modeling using powerful computer technology, trading away technical patterns. By now, technical analysis at low frequencies, such as daily or weekly intervals, is marginalized to work only for the smallest, least liquid securities, which are traded at very low frequencies—once or twice per day or even per week.

b6f1260bee54f438e54f5e8d966ba6d2
ebrary

Some technical analysis techniques, such as momentum or Bollinger bands, have been successfully adopted and extended by modern-day quants in all investing frequencies. It has long been shown that human investors tend to pour money into strategies that worked in recent months. As a result, strategies working in the past month are also likely to work the following month, forming a tradable momentum that can be detected using simple technical moving-average-based indicators, as well as more complex quantitative tools. Similarly, Bollinger bands detect deviation of prices the prespecified number of standard deviations away from the mean. The concept of statistical arbitrage extended Bollinger band principle to detect, for example, deviation of price differences from their long-running means. In this trading exercise, commonly known as pairs trading, traders identify the overpriced and underpriced financial instruments when the price of one instrument exceeds the price of another by the prespecified number of standard deviations of price difference changes. More generally, quants use Bollinger band ideas to pinpoint mean-reverting processes and trade financial instruments with the expectation that the measured average quantity will stay stable, or "stationary" in the language of statistics.

b6f1260bee54f438e54f5e8d966ba6d2
ebrary

Another important investing and trading technique, known as fundamental analysis, originated in equities in the 1930s when traders noticed that future cash flows,

Aldridge, Irene. Wiley Trading Ser. : High-Frequency Trading : A Practical Guide to Algorithmic Strategies and Trading Systems (2). Somerset, US: John Wiley & Sons, Incorporated, 2013. ProQuest ebrary. Web. 8 March 2017.
Copyright © 2013. John Wiley & Sons, Incorporated. All rights reserved.

b6f1260bee54f438e54f5e8d966ba6d2
ebrary

such as dividends, affected market price levels. The cash flows were then discounted back to the present to obtain the fair present market value of the security. Graham and Dodd (1934) were the earliest purveyors of the methodology and their approach is still popular. Over the years, the term *fundamental analysis* expanded to include pricing of securities with no obvious cash flows based on expected economic variables. For example, fundamental determination of exchange rates today implies equilibrium valuation of the rates based on macroeconomic theories.

Fundamental analysis developed through much of the twentieth century. Today, fundamental analysis refers to trading on the expectation that the prices will move to the level predicted by supply-and-demand relationships, the fundamentals of economic theory. In equities, microeconomic models apply; equity prices are still most often determined as present values of future cash flows. In foreign exchange, macroeconomic models are most prevalent; the models specify expected price levels using information about inflation, trade balances of different countries, and other macroeconomic variables. Derivatives are traded fundamentally through advanced econometric models that incorporate statistical properties of price movements of underlying instruments. Fundamental commodities trading analyzes and matches available supply and demand.

b6f1260bee54f438e54f5e8d966ba6d2
ebrary

Various facets of fundamental analysis are inputs into many high-frequency trading models, alongside market microstructure. For example, event arbitrage consists of trading the momentum response accompanying the price adjustment of the security in response to new fundamental information. The date and time of the occurrence of the news event is typically known in advance, and the content of the news is usually revealed at the time of the news announcement. In high-frequency event arbitrage, fundamental analysis can be used to forecast the fundamental value of the economic variable to be announced, in order to further refine the high-frequency process.

Like selected technical models, some fundamental models were adopted by quants who extended the precision of their models, and often dramatically sped up calculation of the relevant values. Fair values of equities following an earnings announcement were recomputed on the fly, enabling quants to reap the profits, at the expense of fundamental traders practicing longhand analysis in Excel spreadsheets.

b6f1260bee54f438e54f5e8d966ba6d2
ebrary

Speed, in fact, became the most obvious aspect of quant competition. Whoever was able to run a quant model the fastest was the first to identify and trade on a market inefficiency and was the one to capture the biggest gain. To increase trading speed, traders began to rely on fast computers to make and execute trading decisions. Technological progress enabled exchanges to adapt to the new technology-driven culture and offer docking convenient for trading. Computerized trading became known as *systematic trading* after the computer systems that processed run-time data and made and executed buy-and-sell decisions.

High-frequency trading developed in the 1990s in response to advances in computer technology and the adoption of the new technology by the exchanges. From the original rudimentary order processing to the current state-of-the-art all-inclusive trading systems, HFT has evolved into a billion-dollar industry.

To ensure optimal execution of systematic trading, algorithms were designed to mimic established execution strategies of traditional traders. To this day, the term

b6f1260bee54f438e54f5e8d966ba6d2
ebrary

Aldridge, Irene. Wiley Trading Ser. : High-Frequency Trading : A Practical Guide to Algorithmic Strategies and Trading Systems (2). Somerset, US: John Wiley & Sons, Incorporated, 2013. ProQuest ebrary. Web. 8 March 2017.
Copyright © 2013. John Wiley & Sons, Incorporated. All rights reserved.

b6f1260bee54f438e54f5e8d966ba6d2
ebrary

algorithmic trading usually refers to the automated "best execution" process—that is, the optimization of buy-and-sell decisions once these buy-and-sell decisions were made by another part of the systematic trading process or by a human portfolio manager. Algorithmic trading may determine how to process an order given current market conditions: whether to execute the order aggressively (on a price close to the market price) or passively (on a limit price far removed from the current market price), in one trade or split into several smaller "packets." As mentioned previously, algorithmic trading does not usually make portfolio allocation decisions; the decisions about when to buy or sell which securities are assumed to be exogenous.

The advances in computer technology over the past decades have enabled fully automated HFT, fueling the profitability of trading desks and generating interest in pushing the technology even further. Trading desks seized upon cost savings realized from replacing expensive trader head count with less expensive trading algorithms along with other advanced computer technology. Immediacy and accuracy of execution and lack of hesitation offered by machines as compared with human traders has also played a significant role in banks' decisions to switch away from traditional trading to systematic operations. Lack of overnight positions has translated into immediate savings due to reduction in overnight position carry costs, a particular issue in crisis-driven tight lending conditions or high-interest environments.

b6f1260bee54f438e54f5e8d966ba6d2
ebrary

Banks also developed and adopted high-frequency functionality in response to demand from buy-side investors. Institutional investors, in turn, have been encouraged to practice high-frequency trading by the influx of capital following shorter lock-ups and daily disclosure to investors. Both institutional and retail investors found that investment products based on quantitative intraday trading have little correlation with traditional buy-and-hold strategies, adding pure return, or alpha, to their portfolios.

Under the Dodd-Frank Act, banks were forced to close many of the proprietary trading operations, but not HFT. In certain banks, the formerly prop-trading HFT is alive and well in the market-making function, where it is now run with client rather than bank capital and is often referred to as *prehedging*.

b6f1260bee54f438e54f5e8d966ba6d2
ebrary

As computer technology develops further and drops in price, high-frequency systems are bound to take on an even more active role. Special care should be taken, however, to distinguish HFT from electronic trading, algorithmic trading, and systematic trading. Figure 1.6 illustrates a schematic difference between high-frequency, systematic, and traditional long-term investing styles.

Systematic trading refers to computer-driven trading decisions that may be held a month or a day or a minute and therefore may or may not be high frequency. An example of systematic trading is a computer program that runs daily, weekly, or even monthly; accepts daily closing prices; outputs portfolio allocation matrices; and places buy-and-sell orders. Such a system is not a high-frequency system.

Another term often mentioned in conjunction but not synonymous with HFT is *electronic trading*. Electronic trading refers to the ability to transmit the orders electronically as opposed to telephone, mail, or in person. Since most orders in today's financial markets are transmitted via computer networks, the term *electronic trading* is rapidly becoming obsolete.

b6f1260bee54f438e54f5e8d966ba6d2
ebrary

Aldridge, Irene. Wiley Trading Ser. : High-Frequency Trading : A Practical Guide to Algorithmic Strategies and Trading Systems (2). Somerset, US: John Wiley & Sons, Incorporated, 2013. ProQuest ebrary. Web. 8 March 2017.
Copyright © 2013. John Wiley & Sons, Incorporated. All rights reserved.

b6f1260bee54f438e54f5e8d966ba6d2
ebrary

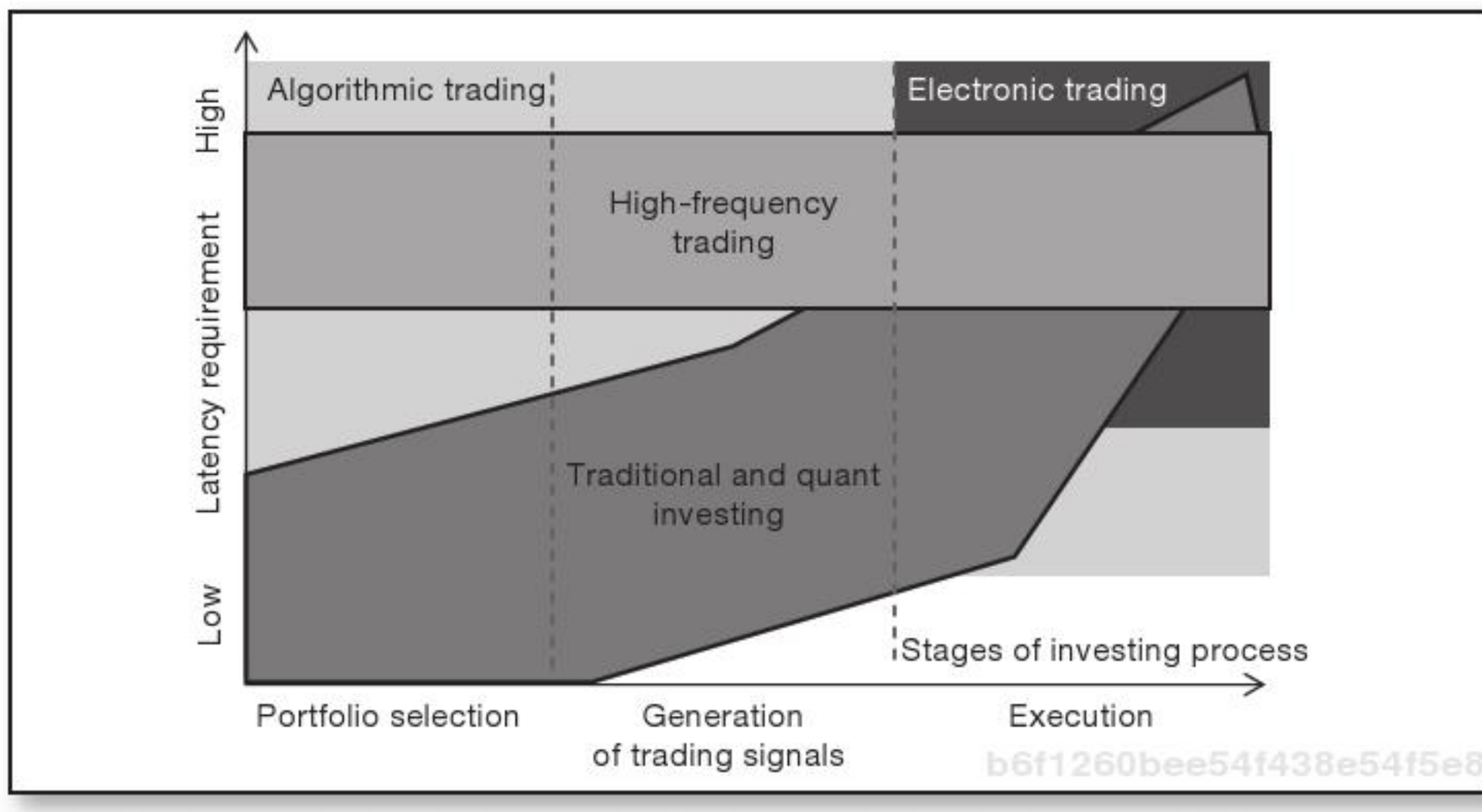


FIGURE 1.6 HFT versus Algorithmic (Systematic) Trading and Traditional Long-Term Investing

Algorithmic trading is more complex than electronic trading and can refer to a variety of algorithms spanning order-execution processes as well as high-frequency portfolio allocation decisions. The execution algorithms are designed to optimize trading execution once the buy-and-sell decisions have been made elsewhere. Algorithmic execution makes decisions about the best way to route the order to the exchange, the best point in time to execute a submitted order if the order is not required to be executed immediately, and the best sequence of sizes in which the order should be optimally processed. Algorithms generating HFT signals make portfolio allocation decisions and decisions to enter or close a position in a particular security. For example, algorithmic execution may determine that a received order to buy 1 million shares of IBM is best handled using increments of 100-share lots to prevent a sudden run-up in the price. The decision fed to the execution algorithm, however, may or may not be high frequency. An algorithm deployed to generate HFT signals, however, would generate the decision to buy the 1 million shares of IBM. The high-frequency signals would then be passed on to the execution algorithm that would determine the optimal timing and routing of the order.

Successful implementation of HFT requires both types of algorithms: those generating HFT signals and those optimizing execution of trading decisions. This book covers both groups of algorithms: those designed for generation of trading signals (Chapters 8 through 11) and those for order execution designed to conceal information within (Chapter 15). Chapter 14 of the book also includes latest algorithms for managing risk of HFT operations.

The intent of algorithmic execution is illustrated by the results of a survey conducted by Automated Trader in 2012. Figure 1.7 shows the full spectrum of responses from the survey. In addition to the previously mentioned factors related to adoption of algorithmic trading, such as performance management and reporting, both buy-side and sell-side managers also reported their use of the algorithms to be driven by trading decision and portfolio management needs.

Aldridge, Irene. Wiley Trading Ser. : High-Frequency Trading : A Practical Guide to Algorithmic Strategies and Trading Systems (2). Somerset, US: John Wiley & Sons, Incorporated, 2013. ProQuest ebrary. Web. 8 March 2017.
Copyright © 2013. John Wiley & Sons, Incorporated. All rights reserved.

b6f1260bee54f438e54f5e8d966ba6d2
ebrary

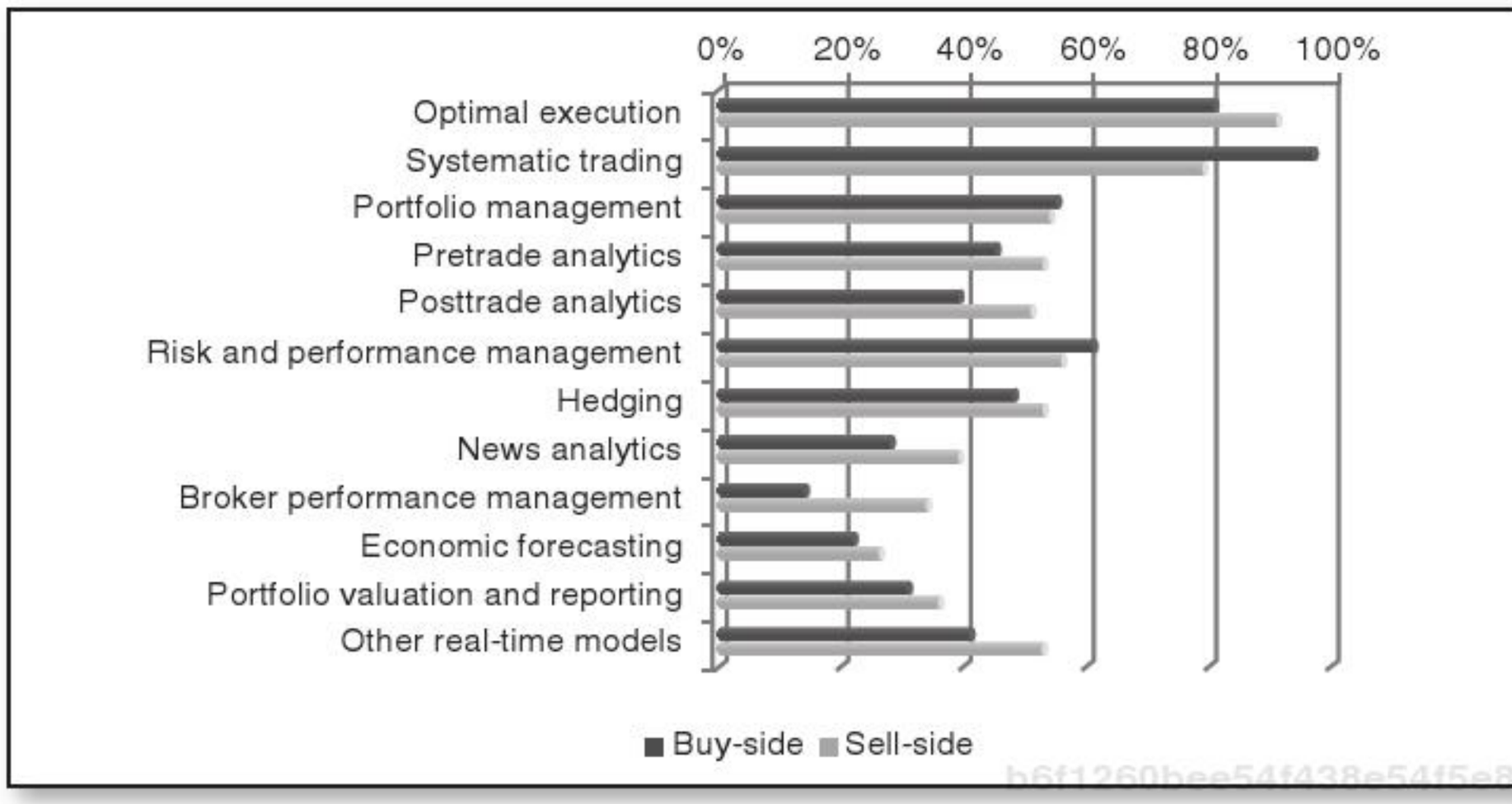


FIGURE 1.7 Reasons for Using Algorithms in Trading
Source: Automated Trader Survey, 2012

True HFT systems make a full range of decisions, from identification of underpriced or overpriced securities through optimal portfolio allocation to best execution. The distinguishing characteristic of HFT is the short position holding times, one day or shorter in duration, usually with no positions held overnight. Because of their rapid execution nature, most HFT systems are fully systematic and are also examples of systematic and algorithmic trading. All systematic and algorithmic trading platforms, however, are not high frequency.

Ability to execute an order algorithmically is a prerequisite for HFT in a given financial instrument. As discussed in Chapter 3, some markets are not yet suitable for HFT, inasmuch as most trading in those markets is performed over the counter (OTC). According to research conducted by Aite Group, equities are the most algorithmically executed asset class, with over 50 percent of the total volume of equities expected to be handled by algorithms by 2010. As Figure 1.8 shows, equities are closely followed by futures. Advances in algorithmic execution of foreign exchange, options, and fixed income, however, have

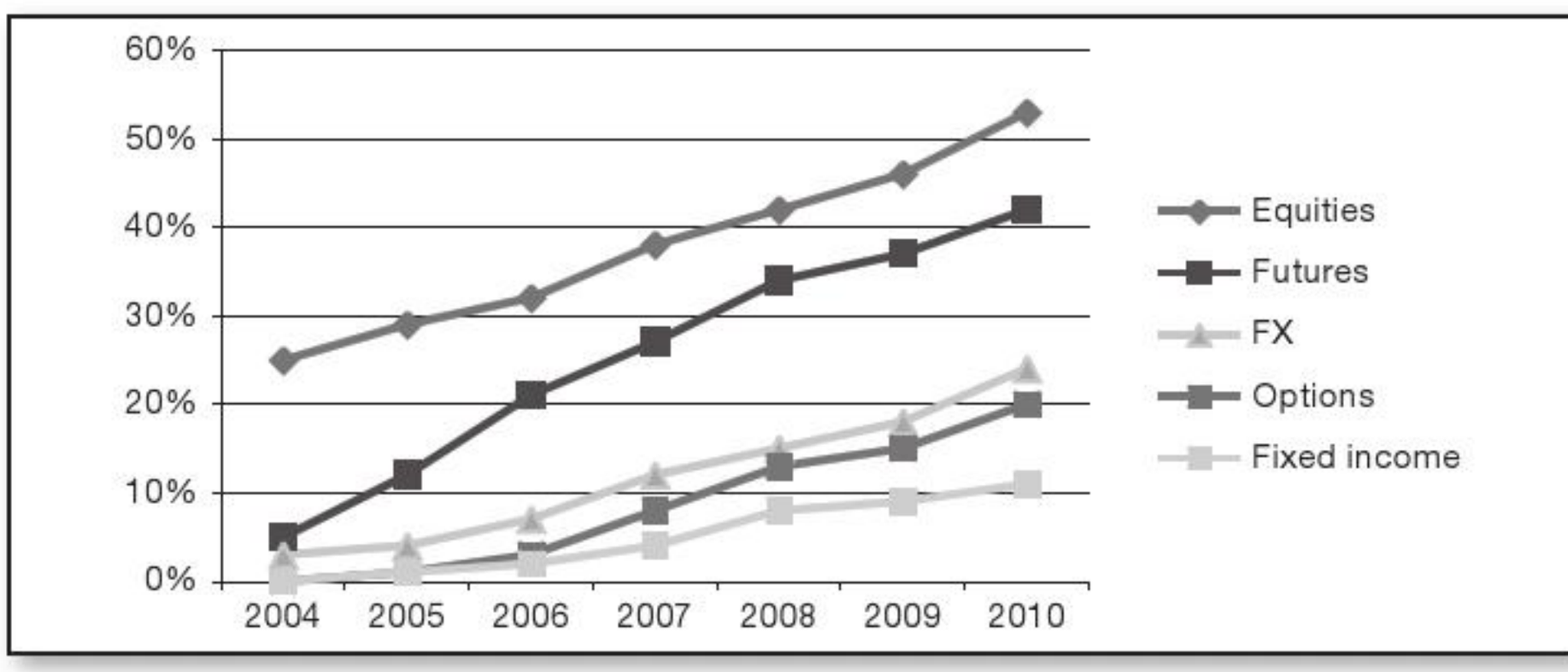


FIGURE 1.8 Adoption of Algorithmic Execution by Asset Class
Source: Aite Group

b6f1260bee54f438e54f5e8d966ba6d2
ebrary

Aldridge, Irene. Wiley Trading Ser. : High-Frequency Trading : A Practical Guide to Algorithmic Strategies and Trading Systems (2). Somerset, US: John Wiley & Sons, Incorporated, 2013. ProQuest ebrary. Web. 8 March 2017.
Copyright © 2013. John Wiley & Sons, Incorporated. All rights reserved.

b6f1260bee54f438e54f5e8d966ba6d2
ebrary

been less visible. As illustrated in Figure 1.8, the lag of fixed-income instruments can be explained by the relative tardiness of electronic trading development for them, given that many of them are traded OTC and are difficult to synchronize as a result.

While research dedicated to the performance of HFT is scarce relative to data on long-term buy-and-hold strategies, anecdotal evidence suggests that most computer-driven strategies are high-frequency strategies. Systematic and algorithmic trading naturally lends itself to trading applications demanding high speed and precision of execution, as well as high-frequency analysis of volumes of tick data. Systematic trading, in turn, has been shown to outperform human-led trading along several key metrics. Aldridge (2009b), for example, shows that systematic funds consistently outperform traditional trading operations when performance is measured by Jensen's alpha (Jensen, 1968), a metric of returns designed to measure the unique skill of trading by abstracting performance from broad market influences. Aldridge (2009b) also shows that the systematic funds outperform nonsystematic funds in raw returns in times of crisis. That finding can be attributed to the lack of emotion inherent in systematic trading strategies as compared with emotion-driven human traders.

b6f1260bee54f438e54f5e8d966ba6d2
ebrary

Furthermore, computers are superior to humans in such basic tasks as information gathering and lightning-fast analysis of a multitude of quotes and news. Physiologically, the human eye cannot capture more than 50 data points per second, as evidenced by an entirely different industry—cinematography. In modern movies, the human eye is exposed to only 24 frames per second, which appear seamless to most moviegoers. Even then, the majority of images displayed on sequential frames involve continuously moving objects. In comparison, modern financial information incorporates drastically bouncing quotes, the number of which can easily exceed 1,000 per second for just one financial instrument. Detecting inter-instrument information spillovers involves processing data for multiple assets and asset classes, as discussed in Chapter 15. Where efficient processing of high volumes of information is key to profitable trading, technology-averse humans have little chance of succeeding. HFT takes over.

b6f1260bee54f438e54f5e8d966ba6d2
ebrary

What Is High-Frequency Trading?

High-frequency trading is an umbrella term comprising several groups of strategies. Given the breadth of HFT, various market participants have somewhat divergent opinions of what HFT actually stands for. This section discusses common definitions of HFT:

- *A definition of HFT that includes all activity utilizing fast algorithmic execution.* For example, the Technology Subcommittee of the U.S. Commodity Futures Trading Commission (CFTC), tasked with compiling a working definition of HFT, came back with the following draft definition in June 2012:

  High-frequency trading is a form of automated trading that employs:

  - Algorithms for decision making, order initiation, generation, routing, or execution, for each individual transaction without human direction;

b6f1260bee54f438e54f5e8d966ba6d2
ebrary

Aldridge, Irene. Wiley Trading Ser. : High-Frequency Trading : A Practical Guide to Algorithmic Strategies and Trading Systems (2). Somerset, US: John Wiley & Sons, Incorporated, 2013. ProQuest ebrary. Web. 8 March 2017.
Copyright © 2013. John Wiley & Sons, Incorporated. All rights reserved.

b6f1260bee54f438e54f5e8d966ba6d2
ebrary

- low-latency technology that is designed to minimize response times, including proximity and co-location services;
- high speed connections to markets for order entry; *and*
- high message rates (orders, quotes or cancellations).

Such definition captures many high-frequency traders, yet also includes 95 percent of investors using algorithmic technology to execute their orders. Even a "mom-and-pop" retail investor entrusting his broker to execute his order in the most efficient algorithmic manner becomes a high-frequency trader under the definition proposed by the CFTC's subcommittee on HFT. Not surprisingly, this definition generated strong dissent from many members of the subcommittee itself.

- *The definition of HFT as a latency-sensitive subset of algorithmic trading.* Gomber, Arndt, Lutat, and Uhle (2011) proposed to define HFT as shown in Figure 1.9. Under such definition, HFTs are algo traders "on steroids," utilizing super-fast technology to speed up algorithmic processes and drive models in supersonic time. Interestingly, also under this definition, the HFTs do not engage in portfolio construction or management, but generate trading signals, validating models, and execute trades in any one security.

b6f1260bee54f438e54f5e8d966ba6d2
ebrary

b6f1260bee54f438e54f5e8d966ba6d2
ebrary

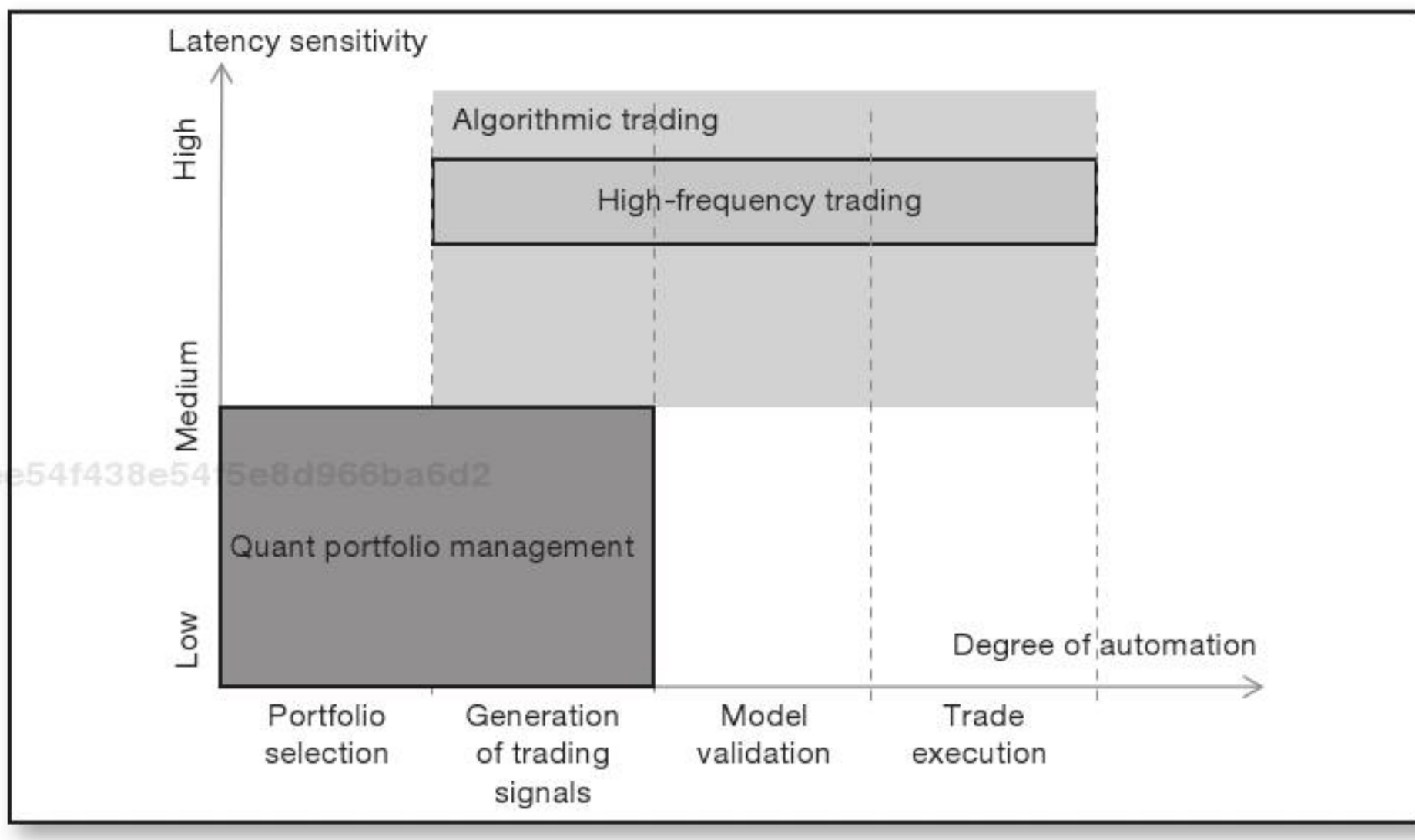


FIGURE 1.9 HFT vs. Algorithmic Trading and Quant Portfolio Management
Source: Gomber, Arndt, Lutat and Uhle (2011)

- *The definition of HFT based on holding period of capital throughput.* A survey of hedge-fund managers, conducted by FINalternatives in 2009, generated the following definition of HFT:

  High-frequency trading comprises

  - Systematic,
  - Quant-based models

b6f1260bee54f438e54f5e8d966ba6d2
ebrary

Aldridge, Irene. Wiley Trading Ser. : High-Frequency Trading : A Practical Guide to Algorithmic Strategies and Trading Systems (2). Somerset, US: John Wiley & Sons, Incorporated, 2013. ProQuest ebrary. Web. 8 March 2017.
Copyright © 2013. John Wiley & Sons, Incorporated. All rights reserved.

b6f1260bee54f438e54f5e8d966ba6d2
ebrary

- With holding periods from a fraction of a second to one day (no positions held overnight).

The survey was based on nearly 300 responses from hedge fund managers who subscribe to FINalternatives (close to 10,000 questionnaires were sent out). It is also worth noting that at the time, a prominent multibillion-dollar Greenwich-based hedge fund launched a high-frequency fund with an average position holding period of three days, a far departure from submicrosecond frequencies often mentioned in connection with HFT. The fund was later retracted.

- *The definition of HFT based on their observed market activity.* Kirilenko, Kyle, Samadi, and Tuzun (2011) identify high-frequency traders as market participants that generate high market volume, all the while holding low inventory. The researchers use the definition to distinguish HFT from other market participants:

b6f1260bee54f438e54f5e8d966ba6d2
ebrary

  - Intermediaries, characterized by low inventory, but not high trading volume.
  - Fundamental buyers, who are consistent net buyers intraday.
  - Fundamental sellers, who are consistent net sellers within a given day.
  - Small traders, generating low volume.
  - Opportunistic traders, loosely defined as all other traders, not fitting the definition of HFT or other categories above.

  Such definition may rely on somewhat arbitrary cutoffs of low inventory and high volume.

  The definition of HFT based on behavior unattainable by human market participants. A common definition used by brokers to segment their clients into HFT and non-HFT, this definition calls for attribution of trading activity of each specific account into human feasible and human infeasible. For example, an account generating 200 orders per second would be deemed HFT, as would an account that consistently succeeds at locking in a penny gain day-in and day-out.

b6f1260bee54f438e54f5e8d966ba6d2
ebrary

  This book considers all of the definitions discussed here.

While a concrete definition of HFT has proven to be a challenge, most market participants are comfortable with the range of strategies deployed by HFT, summarized in Figure 1.10.

What Do High-Frequency Traders Do?

Despite the disagreements about the precise definition of HFT, most market participants agree that HFT strategies fall into the following four broad classes:

1. Arbitrage
2. Directional event-based trading
3. Automated market making
4. Liquidity detection

b6f1260bee54f438e54f5e8d966ba6d2
ebrary

Aldridge, Irene. Wiley Trading Ser. : High-Frequency Trading : A Practical Guide to Algorithmic Strategies and Trading Systems (2). Somerset, US: John Wiley & Sons, Incorporated, 2013. ProQuest ebrary. Web. 8 March 2017.
Copyright © 2013. John Wiley & Sons, Incorporated. All rights reserved.

b6f1260bee54f438e54f5e8d966ba6d2
ebrary

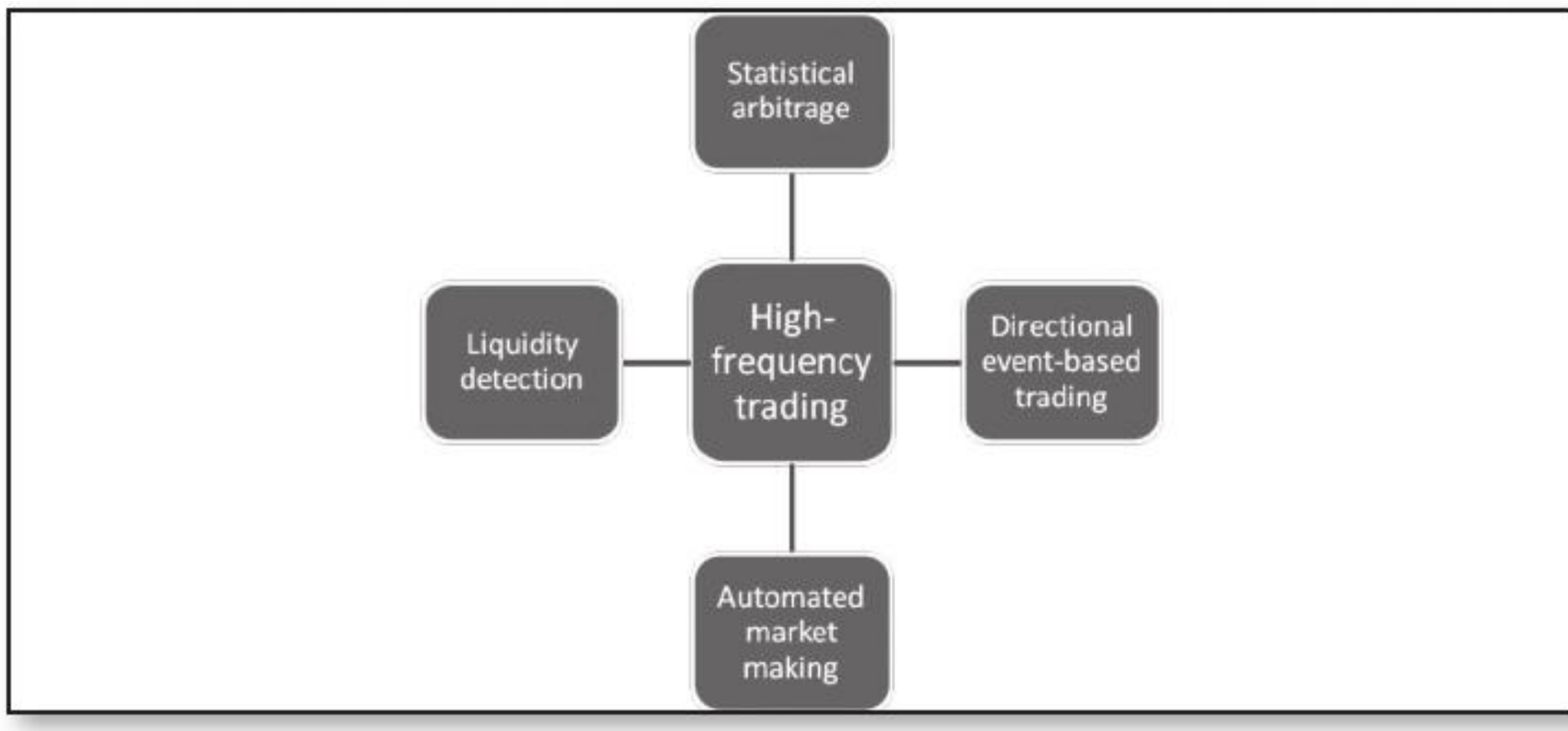


FIGURE 1.10 Major Categories of HFT Strategies

b6f1260bee54f438e54f5e8d966ba6d2
ebrary

Arbitrage strategies trade away price deviations from long-running equilibria or relative asset mispricing, and can include multiple asset classes, as well as multiple exchanges. Many HF arbitrage strategies detect price discrepancies in multiple securities, as discussed in Chapter 8. Several strategies arbitrage prices of the same asset trading on different exchanges, are known as *latency arbitrage strategies,* and are discussed in Chapter 12. Most arbitrage strategies are based on assumptions of mean-reversion of asset prices.

Statistical arbitrage models comprise a range of models, including cross-asset models, where financial securities have strong statistical relationships. All the models included in the book are deep-rooted in economic theories, ruling out spurious statistical relationships often developed using plain data mining and also known as the Spaghetti Principle of Modeling (if one throws a plate of spaghetti filled with data against the wall of statistics, something may stick; what sticks, however, may not have any sound reason for sticking and is likely to fall apart in production). For example, bonds and interest rate futures have been shown to possess considerable dependencies, and their values, as a result, tend to move in tandem. When prices of bonds or interest rate futures deviate from their long-running price equilibrium for no obvious reason, statistical arbitrage trading may be feasible via buying the instrument with a lower-than-expected price relative to the other instrument(s), and selling the instrument with a higher-than-expected price relative to the other instrument(s). Chapter 8 details many economic models, as well as the model estimation techniques and known results.

b6f1260bee54f438e54f5e8d966ba6d2
ebrary

Directional strategies identify short-term trend or momentum. This class of high-frequency strategies includes event-driven strategies, discussed in Chapter 9, other strategies based on predictable short-term price movements, discussed in Chapter 11, as well as the controversial ignition strategies, discussed in Chapter 12. Event arbitrage models show the methodology as well as performance of trading on predictable and recurrent effects of news. Various types of news used in event arbitrage are showcased in Chapter 9, which also includes references to the latest relevant studies as well as specific practical examples.

b6f1260bee54f438e54f5e8d966ba6d2
ebrary

Aldridge, Irene. Wiley Trading Ser. : High-Frequency Trading : A Practical Guide to Algorithmic Strategies and Trading Systems (2). Somerset, US: John Wiley & Sons, Incorporated, 2013. ProQuest ebrary. Web. 8 March 2017.
Copyright © 2013. John Wiley & Sons, Incorporated. All rights reserved.

b6f1260bee54f438e54f5e8d966ba6d2
ebrary

Automated market-making strategies comprise perhaps the most traditional trading strategies, encompassing automated market making, a cost-effective and accurate alternative to human broker-dealers, discussed in detail in Chapter 10. The category of automated market making and liquidity provision includes both inventory-driven and information-driven approaches. Inventory-driven methods tend to focus on joint minimization of the inventory risk and market risk, ensuring that the positions are within a trader's risk tolerance limits given market conditions, and hedged where appropriate. Information-driven market-making models are built with the goal of minimizing the risk of adverse selection, the risk of taking an opposite position to a better-informed party. To minimize the number of such losing positions, high-frequency traders can deploy a wide range of models that help forecast short-term directionality of markets, track the number of well-informed market players in the market waters, and even help forecast impending lumps and shortages of liquidity, covered in Chapter 11. These techniques allow traders to choose the quantities and levels of aggressiveness of their orders based on expectations of surplus or dearth of liquidity.

b6f1260bee54f438e54f5e8d966ba6d2
ebrary

Perhaps least palatable to the low-frequency investors are liquidity detection strategies, like pinging (also known as sniffing and sniping), quote stuffing, and spoofing, addressed in Chapter 12. While this book focuses on explaining sound HFT strategies, the book attempts to draw a balanced perspective and include the methodology behind controversial HFT as well. "Pinging" has been shown to exist on selected venues (pinging was detected in dark pools). The nature of other strategies like "ignition strategies" have been mostly speculative, and no credible evidence of strategy existence has been produced to date. Still, the hypothetical strategies like ignition strategies have been included for completeness, accompanied by a brief analysis of their feasibility, properties, and impact on the broader markets.

How Many High-Frequency Traders Are There?

b6f1260bee54f438e54f5e8d966ba6d2
ebrary

The number of high-frequency traders largely depends on the definition of HFT used. As mentioned earlier, under the CFTC draft definition proposed in June 2012, 19 out of every 20, or 95 percent of all investors and traders would qualify as HFT. Kirilenko, Kyle, Samadi, and Tuzun (2011), define HFTs as traders who produce large trading volume while holding little inventory, and find that HFTs account for about 30 percent of volume in the Standard & Poor's (S&P) 500 E-Mini markets. Aldridge (2012a) estimates that HFTs comprise just 25 to 30 percent in EUR/USD foreign exchange futures and that in the most liquid exchange-traded fund, the S&P 500 SPDR (NYSE:SPY), high-frequency traders on average represent fewer than 20 percent of market participants.

Major Players in the HFT Space

Many HFT participants prefer to stay out of the limelight, all the while generating considerable profits. The most well-known HFT outfits include Getco, Renaissance Capital, and DE Shaw. Lesser-known but still very profitable players dedicated to HFT include specialist firms like IV Capital, DKR Fusion, and WorldQuant.

b6f1260bee54f438e54f5e8d966ba6d2
ebrary

Aldridge, Irene. Wiley Trading Ser. : High-Frequency Trading : A Practical Guide to Algorithmic Strategies and Trading Systems (2). Somerset, US: John Wiley & Sons, Incorporated, 2013. ProQuest ebrary. Web. 8 March 2017.
Copyright © 2013. John Wiley & Sons, Incorporated. All rights reserved.

b6f1260bee54f438e54f5e8d966ba6d2
ebrary

The line between HFT and other forms of trading, however, can be blurred. As mentioned earlier, HFT, and specifically automated market making, are becoming staples on most trading desks in all the major banks. And the advantages of such developments are easy to see: the new automated market-making "robots" are considerably more accurate, inexpensive, and reliable than their human counterparts. Likewise, HFT can seamlessly blend in with activities of statistical arbitrage. In Canada, for example, banks often list most of their HFT in the statistical arbitrage category in the banks' annual reports.

Organization of This Book

This book is written with the explicit goal of providing the latest, yet applied and ready-to-implement information to management and employees that are interested in starting or enhancing their high-frequency trading operations, individuals and institutions seeking to protect their and their clients' trading activity against high-frequency traders, as well as casual observers, seeking to better understand modern financial markets.

b6f1260bee54f438e54f5e8d966ba6d2
ebrary

Chapters 2 through 5 of the book explain the present-day frontiers in financial markets. Chapter 2 describes technological evolution that has enabled algorithmic and high-frequency trading. Chapters 3 through 5 lay the foundation of analysis via description of modern microstructure, high-frequency data, and trading costs.

Chapters 6 and 7 delve into the economics of high-frequency trading. Chapter 6 describes methodologies for evaluating performance and capacity of HFT strategies, and Chapter 7 outlines the business case of HFT.

Chapters 8 through 12 and 14 through 16 are devoted to actual implementation of HFT. Chapters 8 through 12 dissect core models of today's high-frequency strategies. Chapter 14 focuses on risk measurement and management of high-frequency trading as well as portfolio construction. Chapters 15 and 16 discuss the nuts-and-bolts in implementation of HFT systems, as well as best practices in running and monitoring HFT systems.

b6f1260bee54f438e54f5e8d966ba6d2
ebrary

Chapters 13 and 15 focus on regulation of HFT and mitigation of HFT externalities. Chapter 13 presents a summary of current regulatory thought on HFT, discusses models for detection of HFT market manipulation, as well as mathematics of foreseeing market-wide events like flash crashes. Chapter 15 of the book offers solutions for low-frequency traders concerned about the impact of HFT on modern markets. Chapter 15 discusses the latest order slicing techniques and their respective ability to avoid information-prying HFTs, and may also prove useful to high-frequency traders seeking to further expand capacity of their trading systems.

Summary

- High-frequency trading is an organic evolution of trading technology.
- The technological evolution of financial markets created ability to replace human-driven intermediation function with cost-effective technology, returning broker compensation to end investors and bank shareholders.

b6f1260bee54f438e54f5e8d966ba6d2
ebrary

Aldridge, Irene. Wiley Trading Ser. : High-Frequency Trading : A Practical Guide to Algorithmic Strategies and Trading Systems (2). Somerset, US: John Wiley & Sons, Incorporated, 2013. ProQuest ebrary. Web. 8 March 2017.
Copyright © 2013. John Wiley & Sons, Incorporated. All rights reserved.

b6f1260bee54f438e54f5e8d966ba6d2
ebrary

- High-frequency trading strategies are well defined, and most of them are beneficial to the markets.

End-of-Chapter Questions

1. Describe the major groups of today's market participants. What role do they play? How do they interact?
2. What are the core groups of strategies deployed by high-frequency traders?
3. How do high-frequency trading strategies relate to other trading strategies, such as technical analysis, fundamental analysis, and quant strategies?
4. What are the major changes that have occurred in the financial markets over the past 40 years?
5. What is algorithmic trading?
6. How do end investors benefit from high-frequency trading?

b6f1260bee54f438e54f5e8d966ba6d2
ebrary

b6f1260bee54f438e54f5e8d966ba6d2
ebrary

b6f1260bee54f438e54f5e8d966ba6d2
ebrary

Aldridge, Irene. Wiley Trading Ser. : High-Frequency Trading : A Practical Guide to Algorithmic Strategies and Trading Systems (2). Somerset, US: John Wiley & Sons, Incorporated, 2013. ProQuest ebrary. Web. 8 March 2017.
Copyright © 2013. John Wiley & Sons, Incorporated. All rights reserved.

b6f1260bee54f438e54f5e8d966ba6d2
ebrary

b6f1260bee54f438e54f5e8d966ba6d2
ebrary

b6f1260bee54f438e54f5e8d966ba6d2
ebrary

b6f1260bee54f438e54f5e8d966ba6d2
ebrary

Aldridge, Irene. Wiley Trading Ser. : High-Frequency Trading : A Practical Guide to Algorithmic Strategies and Trading Systems (2). Somerset, US: John Wiley & Sons, Incorporated, 2013. ProQuest ebrary. Web. 8 March 2017.
Copyright © 2013. John Wiley & Sons, Incorporated. All rights reserved.

b6f1260bee54f438e54f5e8d966ba6d2
ebrary

CHAPTER 2

Technological Innovations, Systems, and HFT

b6f1260bee54f438e54f5e8d966ba6d2
ebrary

Technological innovation leaves the most persistent mark on the operations of financial markets. While the introduction of new financial instruments, such as EUR/USD in 1999, created large-scale one-time disruptions in market routines, technological changes have a subtle and continuous impact on the markets. Over the years, technology has improved the way news is disseminated, the quality of financial analysis, and the speed of communication. The adoption of technology in financial services, however, was greatly aided by the ever-plummeting costs of technological improvements. This chapter examines the key developments that have occurred in technology over the past several decades in the context of enabling modern financial landscape.

b6f1260bee54f438e54f5e8d966ba6d2
ebrary

A Brief History of Hardware

Trading was institutionalized during the Roman Empire, when the first exchange in currency in designated locations can be noted (benches or "bancas" were the direct precursors of today's banks). Gradual change guided the operations of trading firms until the technological revolution of the twentieth century enabled the current state of trading with rapid exchanges of information. As Figure 2.1 illustrates, over the past 100 years or so the computational speed available to traders has increased exponentially, while the cost of computing has been falling steadily since the 1980s, after reaching its peak.

The price decline in computer technology has been spectacular over the past 20 years. A computer system with 2 gigabytes of memory (RAM), 300 gigabytes of

b6f1260bee54f438e54f5e8d966ba6d2
ebrary

Aldridge, Irene. Wiley Trading Ser. : High-Frequency Trading : A Practical Guide to Algorithmic Strategies and Trading Systems (2). Somerset, US: John Wiley & Sons, Incorporated, 2013. ProQuest ebrary. Web. 8 March 2017.
Copyright © 2013. John Wiley & Sons, Incorporated. All rights reserved.

b6f1260bee54f438e54f5e8d966ba6d2
ebrary

hard drive space, and a 2-gigahertz processor cost several million dollars in 1995, and was big enough to require its own room. In 2012, a computer with identical specifications not only fits in a standard desktop case, it can also be found for as little as $400 in any neighborhood Best Buy or other computer store.

The decline in the cost of computing can be largely traced to the efficiency of scale in production of computer chips overseas. The demand for the increasingly accessible and cheaper technology has, surprisingly, been driven not by the financial services practitioners, but rather by more casual users of computer technology with considerably thinner wallets. Over the past two decades, the latter demanders for cost-efficient technology happened to be video game players, whose sheer scale and desire for lifelike graphics has fueled the surge in mass production and plummeting prices of fast technology. Financial firms have reaped the benefits of innovation and cost efficiencies created by the activity of the video gaming industry.

b6f1260bee54f438e54f5e8d966ba6d2
ebrary

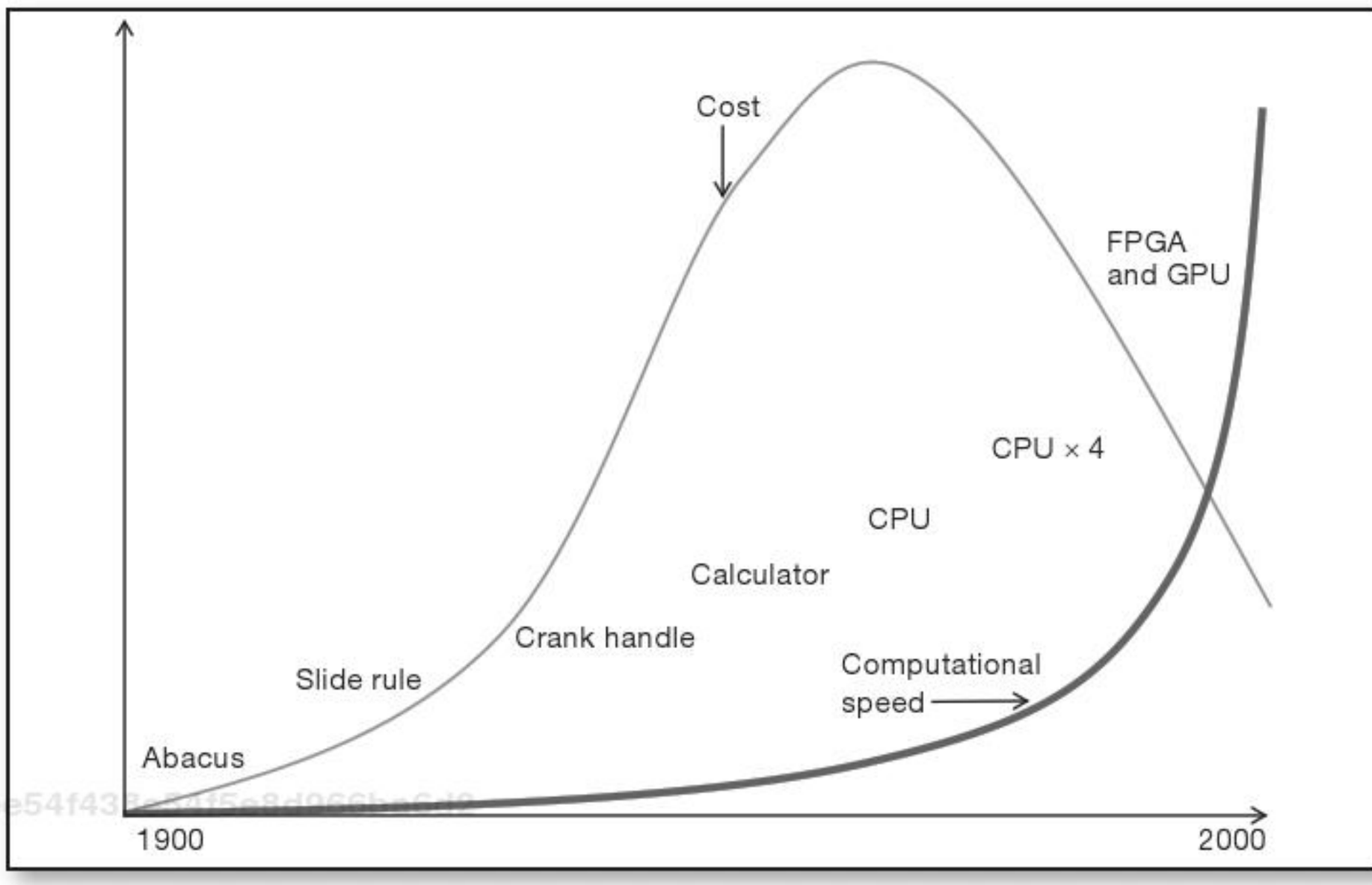


FIGURE 2.1 Evolution of Speed and Costs of Technology over the Twentieth Century

As shown in Figure 2.1, today's advanced technologies comprise multicore central processing units (CPUs), field programmable gate arrays (FPGAs), graphics processing units (GPUs), and the so-called massively parallel architecture chips. A CPU is the brain of the computer and decides how to store information in memory. Multicore CPUs use a shared memory for fast inter-CPU communication, while each individual CPU schedules tasks and performs computations on a given process branch or "thread." Sample architecture of a multicore CPU is shown in Figure 2.2. At the time this book was written, a multicore CPU could cost $100 and higher.

Unlike CPUs, where the majority of the space on the chip is occupied by memory and scheduler functions, the space on a sample GPU is largely devoted to the computational operations, performed in the so-called arithmetic logic units (ALUs). To further maximize efficiency of each chip, process threads are executed in parallel batches of identical size. These batches of threads are called *warps*. To minimize

Aldridge, Irene. Wiley Trading Ser. : High-Frequency Trading : A Practical Guide to Algorithmic Strategies and Trading Systems (2). Somerset, US: John Wiley & Sons, Incorporated, 2013. ProQuest ebrary. Web. 8 March 2017.
Copyright © 2013. John Wiley & Sons, Incorporated. All rights reserved.

b6f1260bee54f438e54f5e8d966ba6d2
ebrary

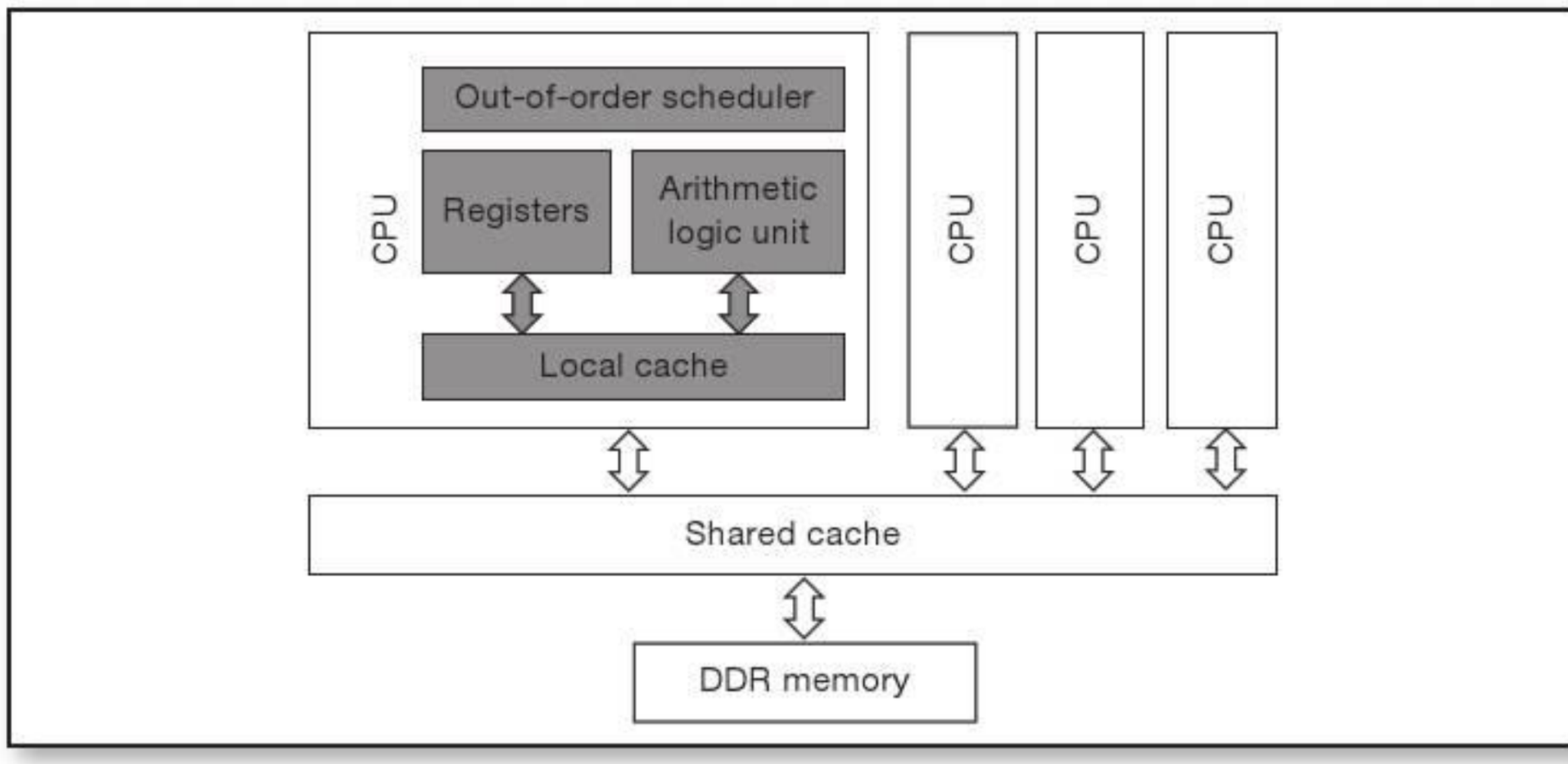


FIGURE 2.2 Architecture of a Sample Multicore CPU
Source: Thomas, Howes and Luk (2009)

b6f1260bee54f438e54f5e8d966ba6d2
ebrary

latency, however, care should be taken to ensure that the threads of the process are similar in terms of the number of loops and conditional exits. In other words, programming expertise is required to ensure that GPUs are deployed with maximum efficiency. Figure 2.3 illustrates sample architecture of the GPU. A popular model of a GPU is Nvidia GTX series, which can retail for $100 to $700 per card.

FPGAs are an entirely different class of chips that do not have any fixed instruction set architecture. Instead, an FPGA provides a blank slate of bitwise functional units that can be programmed to create any desired circuit or processor. Some FPGAs

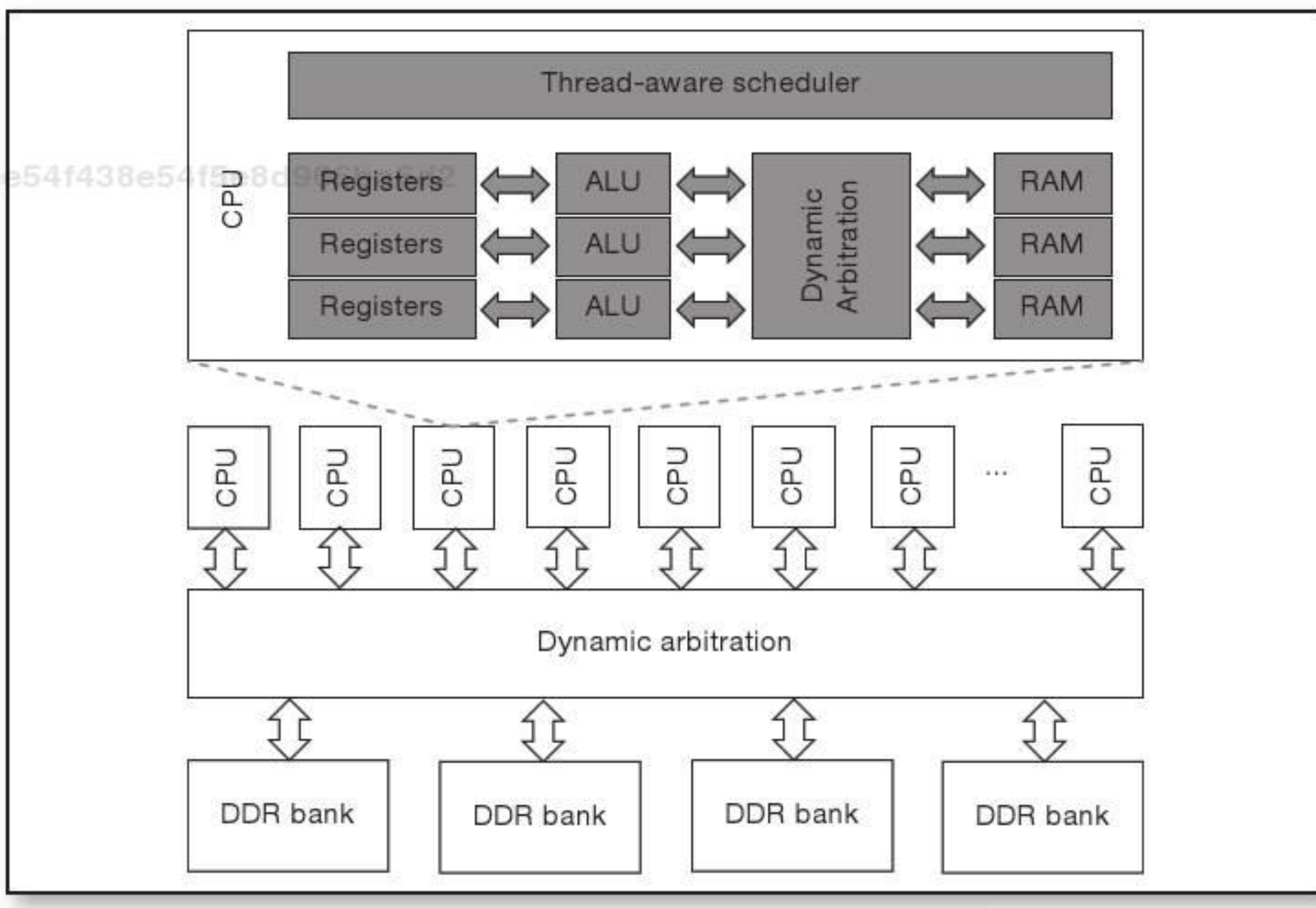


FIGURE 2.3 Architecture of a Sample GPU
Source: Thomas, Howes and Luk (2009)

b6f1260bee54f438e54f5e8d966ba6d2
ebrary

Aldridge, Irene. Wiley Trading Ser. : High-Frequency Trading : A Practical Guide to Algorithmic Strategies and Trading Systems (2). Somerset, US: John Wiley & Sons, Incorporated, 2013. ProQuest ebrary. Web. 8 March 2017.
Copyright © 2013. John Wiley & Sons, Incorporated. All rights reserved.

b6f1260bee54f438e54f5e8d966ba6d2
ebrary

contain a number of dedicated functional units, such as multipliers and memory blocks. Most of the area of an FPGA, however, is dedicated to the routing infrastructure the run-time connectivity of the FPGA's functional units. Figure 2.4 shows the architecture of a sample FPGA chip.

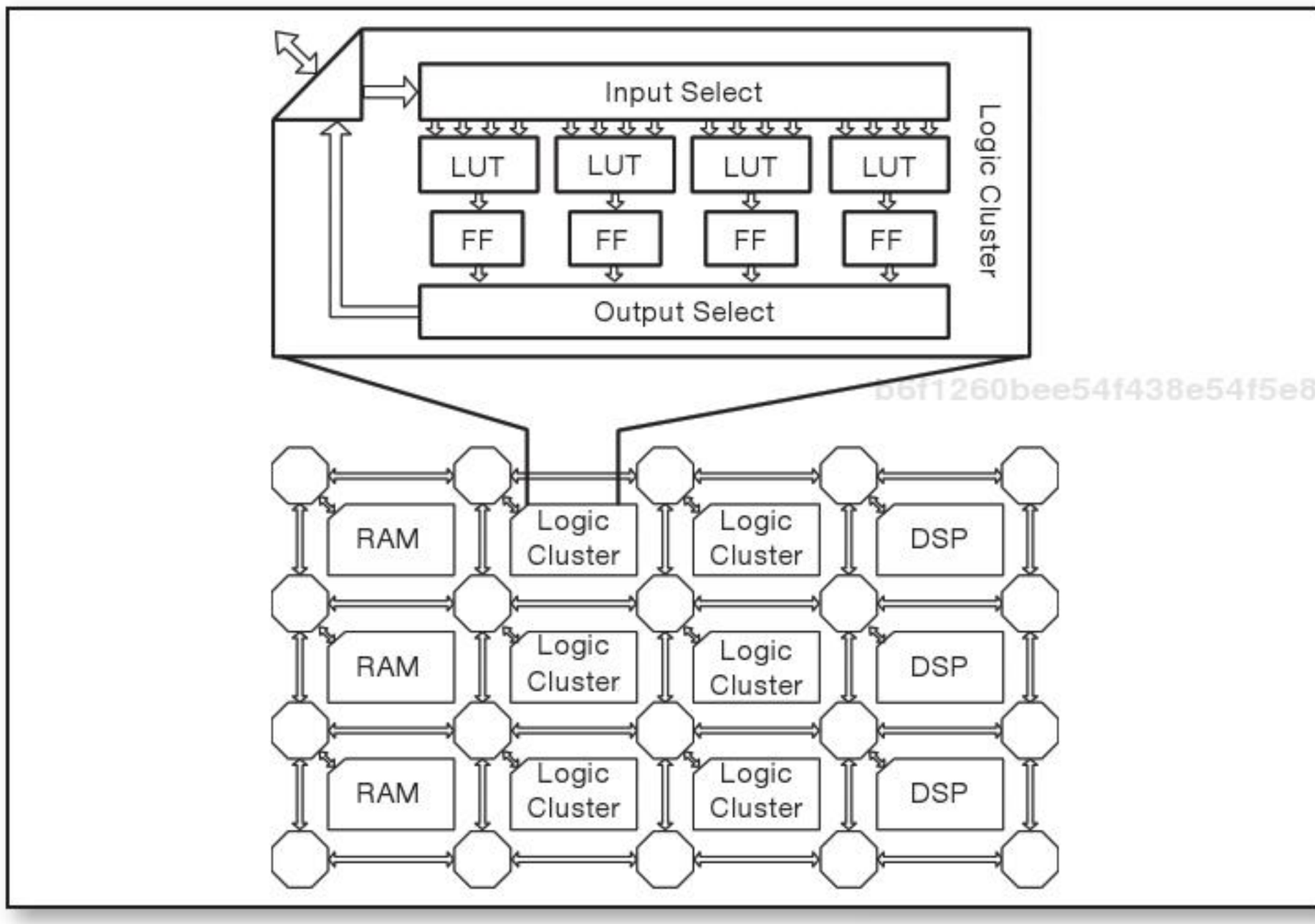


FIGURE 2.4 Architecture of a Sample FPGA Chip
Source: Thomas, Howes and Luk (2009)

The main distinction of FPGAs is that the programming code is written directly onto the chip from the outset. FPGAs are programmed using special programming languages, such as Verilog or VHDL. The languages are similar to C programming language and are easy to learn. A special FPGA programming device translates Verilog or VHDL into Assembly language understood by the FPGA chips. In the absence of FPGAs, trading programs need to be compiled and translated to the computer chips like CPUs during program run time, requiring additional computer operations and eating into the latency. The process of programming an FPGA is rather straightforward and inexpensive. While there exists a significant variation in costs of blank FPGA chips and Verilog or VHDL compilers and simulators, quality inexpensive options are commonly available, once again produced to satisfy demand of video gamers. A blank FPGA chip may cost anywhere from $4 to $5,000. The Verilog software and simulators may be free ("open-source") or $20,000. The software is then downloaded onto the chip, using the process specific to the chip manufacturer. Programming of FPGA chips is often taught in undergraduate electrical engineering programs, and tends to be easy to learn. However, achieving a state-of-the-art FPGA system may require arranging FPGAs in a format known as *massively parallel processor array configuration,* demanding advanced understanding of hardware and software optimization.

Aldridge, Irene. Wiley Trading Ser. : High-Frequency Trading : A Practical Guide to Algorithmic Strategies and Trading Systems (2). Somerset, US: John Wiley & Sons, Incorporated, 2013. ProQuest ebrary. Web. 8 March 2017.
Copyright © 2013. John Wiley & Sons, Incorporated. All rights reserved.

b6f1260bee54f438e54f5e8d966ba6d2
ebrary

Performance-wise, FPGAs tend to be superior to GPUs and CPUs, particularly when used to simultaneously process a limited number of time series. Figure 2.5 shows a graphical comparison of efficiency of key hardware models. The horizontal axis of the figure shows the "input" size, or the number of independent variables simultaneously fed into the algorithm. The vertical axis shows the number of computer "cycles" required to perform an operation involving the given number of inputs. As Figure 2.5 illustrates, an FPGA posts best results when the number of inputs is less than 2,000. When the number of inputs exceeds this threshold, the speed of an FPGA becomes comparable to that of a GPU.

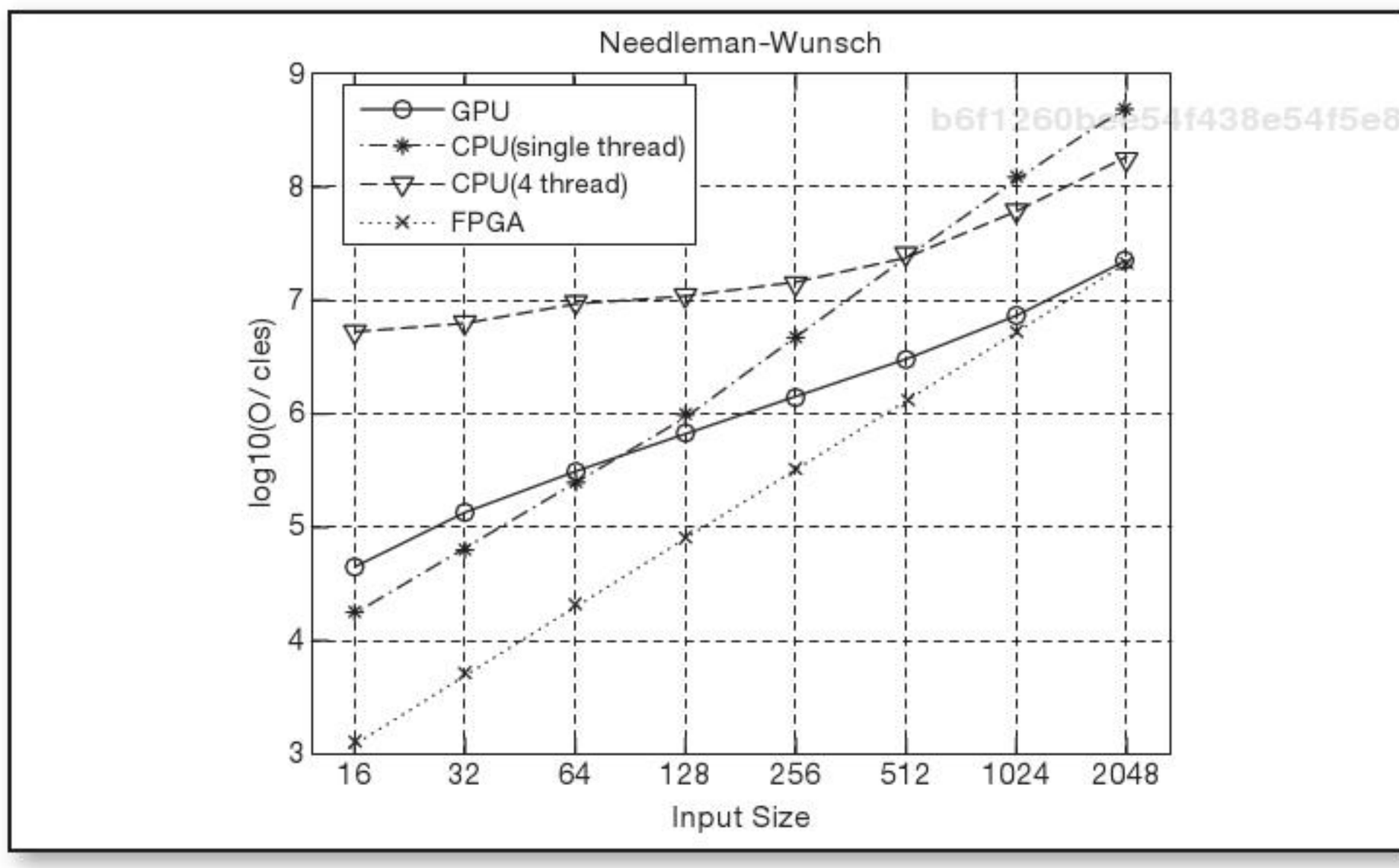


FIGURE 2.5 Comparative Performance of FPGA, GPU, Single CPU, and Quad CPU Architectures
Source: Thomas, Howes and Luk (2009)

b6f1260bee54f438e54f5e8d966ba6d2
ebrary

The choice of a chip itself is not the single determinant of the speed of the computer program. The speed of each computer cycle is determined by the so-called oscillator crystal within each machine and, most important, organization of the program's algorithm.

Messaging

Hardware is just one of many components of computer technology necessary for achieving successful trading. Another crucial component is messaging, enabling communication among hardware and software modules of various market participants. Just as speed is important in hardware, it is also important in messaging. In fact, it is the speed of messaging that presents a hurdle or a bottleneck for trading communication.

b6f1260bee54f438e54f5e8d966ba6d2
ebrary

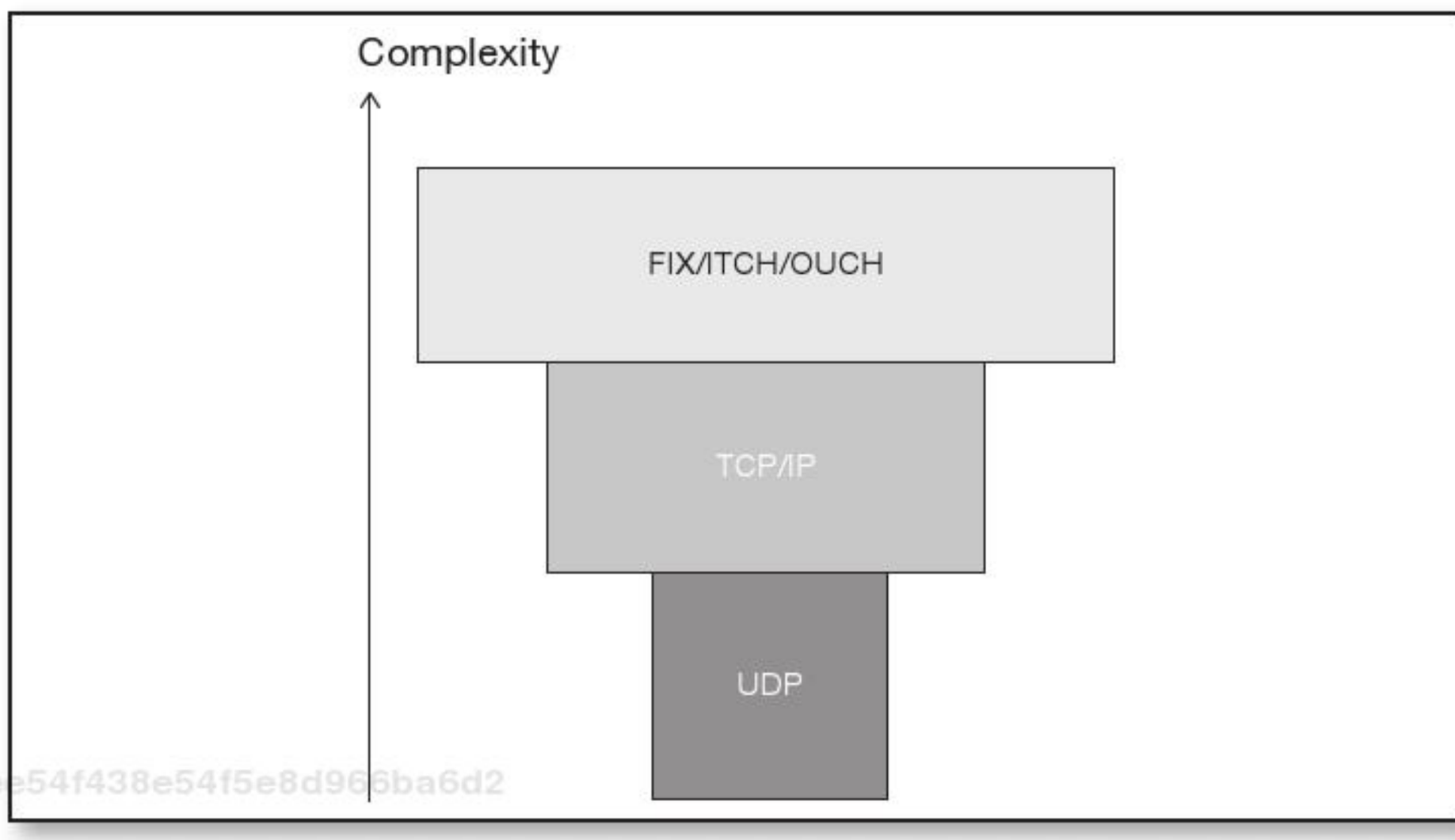
Aldridge, Irene. Wiley Trading Ser. : High-Frequency Trading : A Practical Guide to Algorithmic Strategies and Trading Systems (2). Somerset, US: John Wiley & Sons, Incorporated, 2013. ProQuest ebrary. Web. 8 March 2017.
Copyright © 2013. John Wiley & Sons, Incorporated. All rights reserved.

b6f1260bee54f438e54f5e8d966ba6d2
ebrary

Messaging Protocols

Trading messaging is comprised of three levels of protocols, shown in Figure 2.6. The most basic level of communication enables data streaming and is known as the User Datagram Protocol (UDP). UDP is the "bare bones" data communication protocol, lean in its implementation, and utilizing the fewest number of bytes and messages to identify and deliver the stream of data. As a result, UDP is very fast, but does not guarantee delivery of sent data. UDP is the same technology as the one used to stream games and movies over the Internet, where loss of one packet here and there does not significantly impact the viewer's experience. In trading, UDP is sometimes used to transmit quotes, the data that are refreshed continuously, and are, therefore, not very sensitive to lost information. If a particular quote sent from an exchange fails to reach a trader, the resulting impact may be deemed minimal: a new revised quote is already on the way, retiring the lost quote upon hitting the trader's account.

b6f1260bee54f438e54f5e8d966ba6d2
ebrary



b6f1260bee54f438e54f5e8d966ba6d2
ebrary

FIGURE 2.6 Three Levels of Complexity of Communication Protocols Used in Trading

The integrity of the quote process, however, can matter in trading model development. A trading algorithm developer may rely on the quote stream idiosyncrasies to generate predictive signals about impending market movements. If the structure of the historical quote stream used in model development differs significantly from that of the quote stream encountered by the trader "in production" environment, the calculated forecasts may cease working. Care should be taken to ensure that the data used in simulation and back-test of the algorithms is structurally compatible to the data received in production environment. At a minimum, the algorithm designer should ascertain that the frequency of quotes received in production matches that in the historical data used in the back-test. More complicated data tests can also be performed. For example, a rolling autocorrelation metric can be computed on the two sets of data, and the distribution of the resulting metrics should be comparable for successful algorithm design and implementation.

b6f1260bee54f438e54f5e8d966ba6d2
ebrary

Aldridge, Irene. Wiley Trading Ser. : High-Frequency Trading : A Practical Guide to Algorithmic Strategies and Trading Systems (2). Somerset, US: John Wiley & Sons, Incorporated, 2013. ProQuest ebrary. Web. 8 March 2017.
Copyright © 2013. John Wiley & Sons, Incorporated. All rights reserved.

b6f1260bee54f438e54f5e8d966ba6d2
ebrary

The next level of complexity in communication protocols is Transmission Control Protocol/Internet Protocol (TCP/IP). TCP/IP is another standard Internet communication protocol, presently used in most e-mail and Web-browsing communication. Unlike the UDP, where individual packets of information do not carry any identifying monikers, all packets of a TCP/IP transmission are sequentially numbered, the total number of bytes within each packet is counted, and undelivered or corrupt data is re-sent. As a result, TCP/IP provides a more secure framework for information delivery, and is used to transmit orders, order acknowledgments, execution acknowledgments, order cancellations, and similarly important information. As a trade-off, TCP/IP tends to be three times slower than UDP. Figure 2.7 summarizes common usage of UDP, TCP/IP and FIX in trading communication.

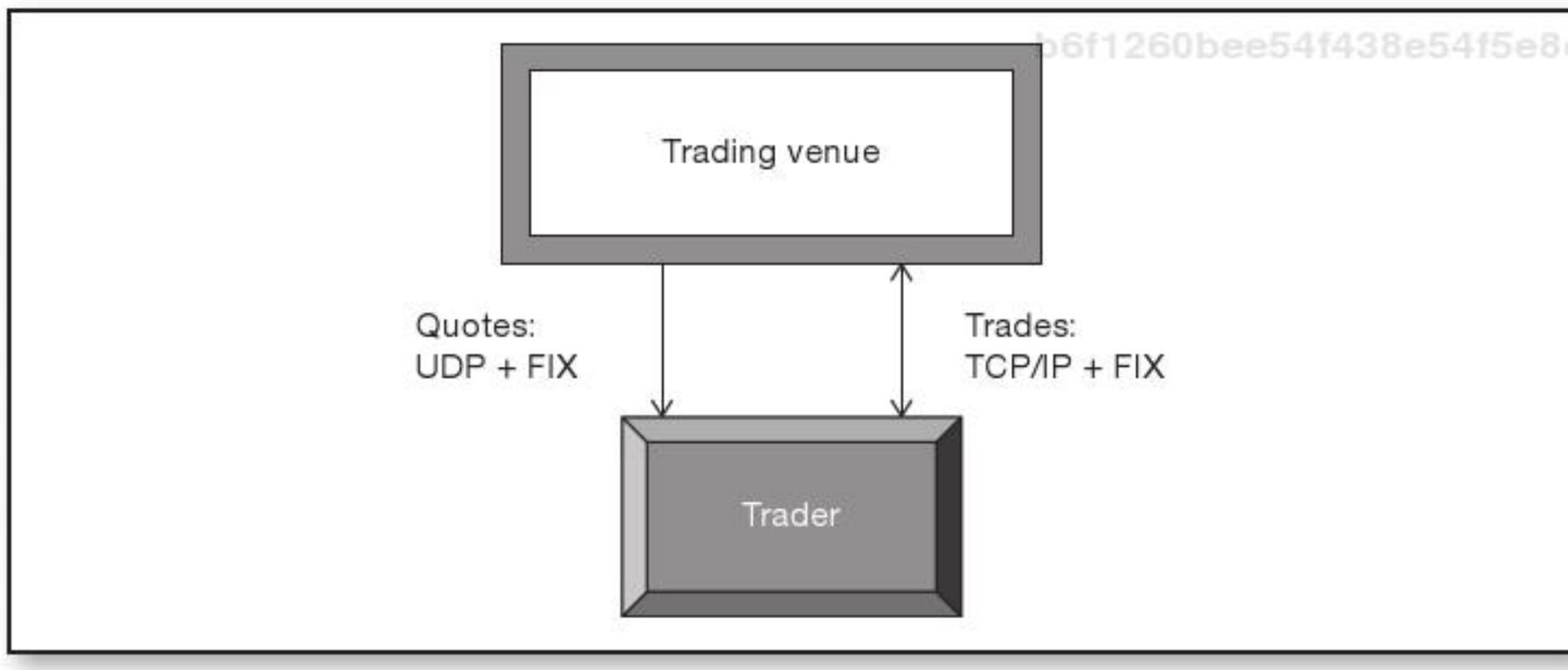


FIGURE 2.7 Common Use of Protocols in Trading Communication

Both the UDP and TCP/IP, however, require an additional layer of communication to standardize messages of the trading process. Protocols like Financial Information eXchange (FIX), ITCH, OUCH, and FAST are used on top of UDP and TCP to transmit data in a standardized machine-readable format. FIX protocol is a free XML-based text specification for quote, order, trade and related message transmission. The FIX protocol comprises data field definitions, enumerations, and various components, forming messages. Each message is then populated with the user-generated data. Each field of the message, including the version of FIX used, the time stamp, and other information, is separated from the following field by binary 1.

```
8=FIX.4.2 | 9=309 | 35=S | 34=5015 | 52=20070731-15:25:20 |
131=1185895365 | 301=0 | 55=USD/CAD | 167=FOR | 15=USD |
132=1.065450 | 133=1.065850 | 134=5000000.0 | 135=5000000.0 |
647=2000001.0 | 648=2000001.0 | 188=1.06545 | 190=1.06585 |
60=20070731-15:25:20 | 40=H | 64=20070801 | 10=178
```

FIGURE 2.8 Sample FIX Message

Figure 2.8 illustrates a sample FIX message, transmitting a quote for USD/CAD exchange rate. The shown quote contains the following information:

Aldridge, Irene. Wiley Trading Ser. : High-Frequency Trading : A Practical Guide to Algorithmic Strategies and Trading Systems (2). Somerset, US: John Wiley & Sons, Incorporated, 2013. ProQuest ebrary. Web. 8 March 2017.
Copyright © 2013. John Wiley & Sons, Incorporated. All rights reserved.

b6f1260bee54f438e54f5e8d966ba6d2
ebrary

- Version of FIX, "FIX.4.2" (field number 8)
- The time stamp of the message, "20070731-15:25:20" (field number 52)
- Security identifier, "USD/CAD" (field number 55)
- Security type, "FOR" for foreign exchange (field number 167)
- Base currency, "USD" for U.S.$ (field number 15)
- Best bid and best ask (fields 132 and 133, respectively)
- Sizes at the best bid and best ask (fields 134 and 135)

Transmission speed of communication messages depends on several factors:

- Size of message
- Connection bandwidth
- TCP/IP and UDP "window" sizes, specifying how many bytes market participants are willing to send and receive at per message "slice." Once the system of one market participant sends out a message, the message is sliced into the individual parcels of a specified window length, a message header is attached to each parcel, and the messages are sent out on their route. The UDP message header typically identifies the destination, and consists of just 8 bytes. The TCP/IP message header includes the sender and destination identifications, parcel sequence number, and the total number of parcels comprising the message, among other variables. The standard TCP/IP header is 20 bytes. The FIX header can be much more elaborate, and is often in excess of 100 bytes.

b6f1260bee54f438e54f5e8d966ba6d2
ebrary

While FIX is widely used, it is slow in comparison with Nasdaq's protocols known as ITCH and OUCH. The binary nature of ITCH and OUCH ensures that the messages arrive in the machine-readable format, using no processing time to convert them from text to binary and back. In addition to the binary format, ITCH and OUCH messages have a fixed message length, enabling faster transmission. OUCH is the order entry protocol, while ITCH is the outbound quote and trade-data dissemination specification. Yet ITCH and OUCH support only a limited number of messages. OUCH provides the platform for:

b6f1260bee54f438e54f5e8d966ba6d2
ebrary

- Order entry.
- Replacement and cancellations.
- Receipt of execution acknowledgements.

ITCH is built for fast and lean quote and past trade data dissemination, and is able to transmit:

- Order-level data.
- Trade messages.
- Net order imbalance data.

b6f1260bee54f438e54f5e8d966ba6d2
ebrary

Aldridge, Irene. Wiley Trading Ser. : High-Frequency Trading : A Practical Guide to Algorithmic Strategies and Trading Systems (2). Somerset, US: John Wiley & Sons, Incorporated, 2013. ProQuest ebrary. Web. 8 March 2017.
Copyright © 2013. John Wiley & Sons, Incorporated. All rights reserved.

b6f1260bee54f438e54f5e8d966ba6d2
ebrary

- Administrative messages.
- Event controls, such as start of day, end of day, and emergency market halt/resume.

For more complex messages, ITCH- and OUCH-enabled market participants are often required to use FIX.

Core Message Architecture

Tick information is transmitted among market participants using one or several quote messaging specifications. FIX, ITCH, OUCH, and FAST are just a few message languages enabling transmission of critical trading information. Despite their complicated acronyms, messaging is built around strikingly simple architecture, as illustrated in Figure 2.9.

b6f1260bee54f438e54f5e8d966ba6d2
ebrary

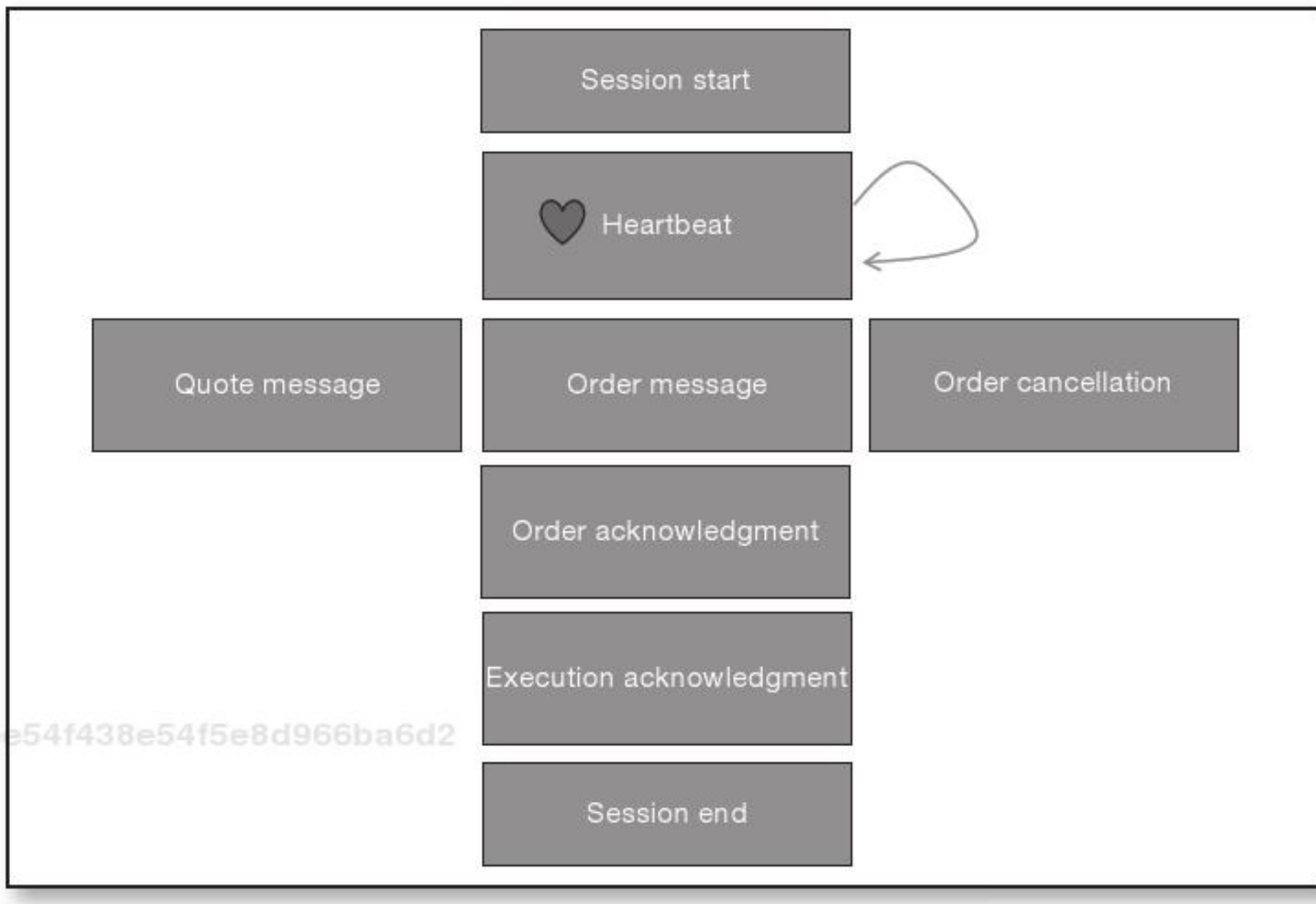


FIGURE 2.9 Core Message Architecture in Trading

b6f1260bee54f438e54f5e8d966ba6d2
ebrary

As shown in Figure 2.9, every stream of quote and trade communication includes the following key messages:

1. *Session start* is the message sent in the beginning of every communication session, sometimes only once a day. The session start message notifies relevant market participants that the entity is open for trading and desires to establish a communication stream.
2. *Heartbeat* is a recurrent message that notifies the participant's communication parties that the participant is online, in a state of good technological health, and open for business. Parties that fail to receive their communication partners'

b6f1260bee54f438e54f5e8d966ba6d2
ebrary

Aldridge, Irene. Wiley Trading Ser. : High-Frequency Trading : A Practical Guide to Algorithmic Strategies and Trading Systems (2). Somerset, US: John Wiley & Sons, Incorporated, 2013. ProQuest ebrary. Web. 8 March 2017.
Copyright © 2013. John Wiley & Sons, Incorporated. All rights reserved.

b6f1260bee54f438e54f5e8d966ba6d2
ebrary

heartbeat messages for a preconfigured period of time often shut down the communication channel. The communication channel may then be reinstated using the "session start" sequence.

3. *Quote* message is a message carrying quote information, such as best bid and ask prices and sizes. Level II data, like the depth of the order book behind the best bid and ask quotes, may also be transmitted using quote messages.
4. *Order* message is used to transmit actual order information. A typical order message includes a buy or sell identifier, an order type—a market, limit or other specification, order size, and a desired execution price and validity period (day, good-till-canceled) in the case of limit orders.
5. *Order cancellation* message includes the unique identifier of the previously placed order that now needs to be canceled.
6. *Order acknowledgment* and *order cancellation acknowledgment* messages include confirmations of order placement or order cancellation, respectively.
7. *Execution acknowledgment* messages specify the details of execution: time of execution, obtained price, and execute quantity.
8. *Session end* message informs parties that a given trading entity has stopped trading and quoting for the day.

b6f1260bee54f438e54f5e8d966ba6d2
ebrary

The resulting trade messaging flow comprises intuitive methodology to deliver effective, reliable, and traceable communication. Most trading outfits log their daily communication for easy reconciliation and fast identification of potential issues, such as network connectivity problems, algorithm errors, and the like.

Speed and Security

Neither TCP/IP nor UDP incorporate encryption. In other words, most TCP/IP and UDP messages are sent via Internet networks in plain text. FIX provides optional encryption at a considerable latency. While ITCH and OUCH send messages in a binary format, most Nasdaq OMX messages are still sent unencrypted over the Internet networks.

b6f1260bee54f438e54f5e8d966ba6d2
ebrary

What kind of risk do market participants face by sending unencrypted messages? To answer this question, one needs to consider the current layout and flow of the Internet traffic. Today, most Internet traffic in the world flows through about 80 "core" nodes. These nodes, such as major Internet service providers (ISPs) like Verizon, have some security measures in place, limiting the incidence of spying activity at these nodes. At the same time, nodes can be quite congested, slowing down messaging traffic without consideration to its urgency.

If the core nodes were to fail, 70 percent of Internet traffic would flow through peer-to-peer networks, redundant backup structures in which the traffic would hop from one local user to another, in a distributed fashion. While the peer-to-peer network configuration can allow the network participants to observe each other's traffic and read unencrypted messages in full, the peer-to-peer communication is sufficiently randomized to prevent any peer party to accumulate the entire message flow. Still, peer-to-peer networks may be vulnerable to malicious intent, and the resulting potential hijacking of the order flow could destroy markets and cause tremendous losses for all market participants.

b6f1260bee54f438e54f5e8d966ba6d2
ebrary

Aldridge, Irene. Wiley Trading Ser. : High-Frequency Trading : A Practical Guide to Algorithmic Strategies and Trading Systems (2). Somerset, US: John Wiley & Sons, Incorporated, 2013. ProQuest ebrary. Web. 8 March 2017.
Copyright © 2013. John Wiley & Sons, Incorporated. All rights reserved.

b6f1260bee54f438e54f5e8d966ba6d2
ebrary

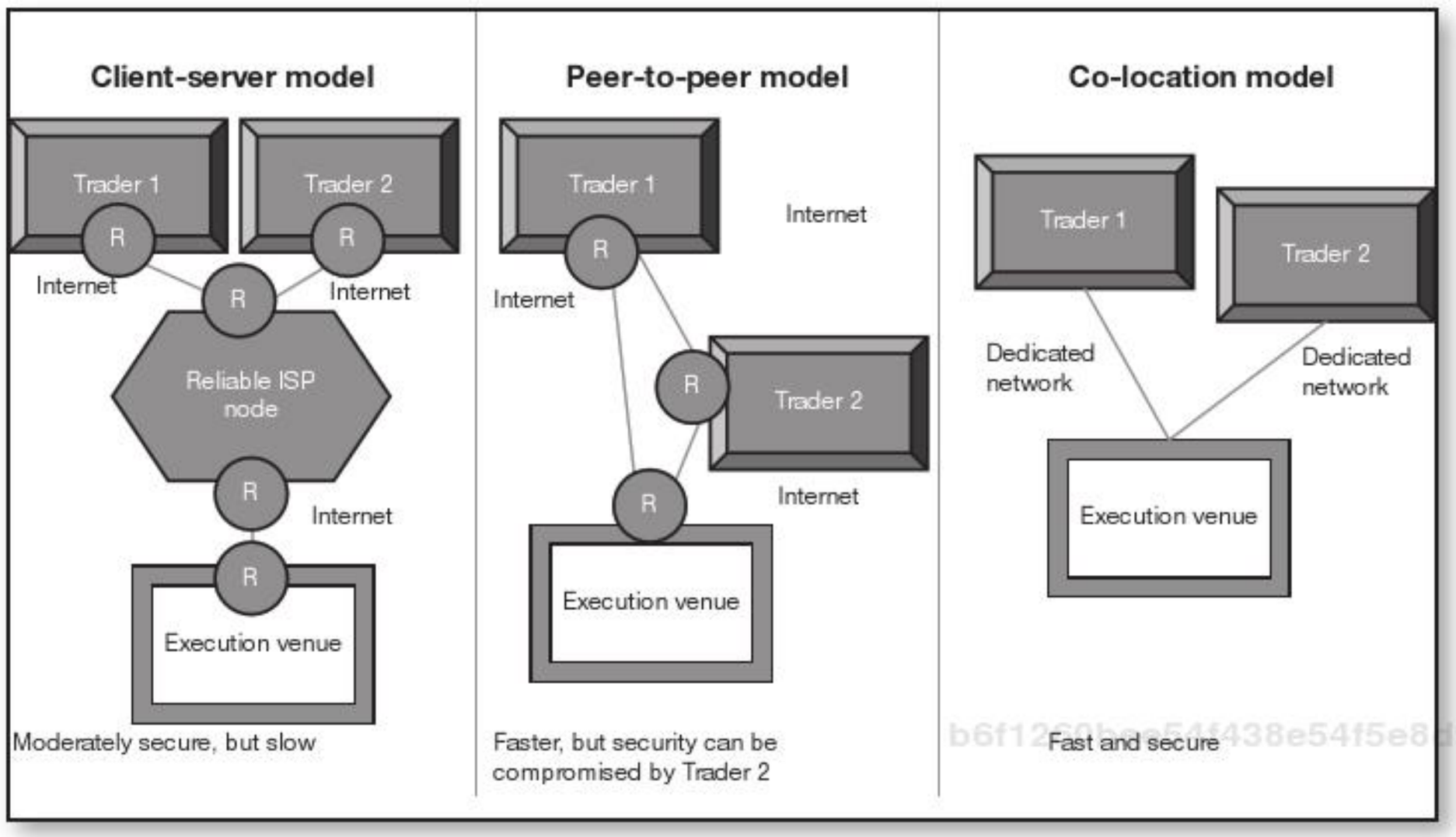


FIGURE 2.10 Messaging Architecture in Client-Server, Peer-to-Peer, and Co-location Models

Figure 2.10 describes three common Internet communication models prevalent in trading, including the so-called co-location model. In the co-location model, traders' servers are housed in the same secure facility as the exchange's matching servers. In the co-location scenario, trading servers have dedicated network access to the exchange servers. The dedicated network access comprises a private secure communication line from the trading servers directly to the exchange, minimizing the risk of malicious intervention and ensuring a safe environment for all market participants. Co-location also provides such benefits as speed advantage: the servers of a Chicago trader co-located with Nasdaq, for example, would allow the trader to shave anywhere from 17 to 22 milliseconds in a round-trip order latency due to the physical distance between New York and Chicago, all in addition to the immeasurable savings resulting from the security of the co-located trading connection. Figure 2.11 summarizes latency among selected co-location centers worldwide.

	New York, NY	Washington, DC	Toronto, Canada	Chicago, IL	London, U.K.	Frankfurt, Germany	Sao Paolo, Brazil	Tokyo, Japan
Newark, NJ	0.314	3.400	9.470	15.175	65.763	74.383	109.414	141.640
New York, NY		4.057	9.784	15.291	65.533	74.153	109.100	141.756
Washington, DC			13.270	14.175	69.083	78.164	111.960	140.640
Toronto, ON				10.795	75.233	83.853	118.884	136.910
Chicago, IL					80.740	89.360	123.485	127.595
London, U.K.						8.620	183.253	215.825
Frankfurt, Germany							183.253	215.825
Sao Paolo, Brazil								249.950

FIGURE 2.11 Latency Incurred by Electronic Signals Traveling over Optical Fiber Networks between Pairs of Locations

Aldridge, Irene. Wiley Trading Ser. : High-Frequency Trading : A Practical Guide to Algorithmic Strategies and Trading Systems (2). Somerset, US: John Wiley & Sons, Incorporated, 2013. ProQuest ebrary. Web. 8 March 2017.
Copyright © 2013. John Wiley & Sons, Incorporated. All rights reserved.

b6f1260bee54f438e54f5e8d966ba6d2
ebrary

Within most co-location data centers, servers are positioned at various distances from the exchange server itself, raising natural concerns about "fairness" of connections of all traders co-located in a given facility. A physical distance difference of as little as 100 feet may result in one microsecond (one millionth of a second) time delay on every message sent and received, giving traders co-located near the exchange servers a potential advantage. To address such issues, the Nasdaq co-location center guarantees an equidistant length of fiber-optic cable from the servers to the exchange to the servers of each trader co-located in the Nasdaq's facility. The cable lengths are identical down to the millimeter, and can be seen coiled near the servers of traders physically close to servers of the exchange.

Although some market participants believe co-location to be unaffordably expensive, the real numbers point to the opposite. At a data center in Secaucus, New Jersey, for example, a private company, Equinix, offers co-location-equivalent proximity services with the minimum monthly charges broken down as follows:

b6f1260bee54f438e54f5e8d966ba6d2
ebrary

- A cabinet for the trader's hardware equipped with biometric security scanners and air-conditioning runs $1,500 per month.
- A 20-amp 120-volt primary power source costs $350 per month.
- An additional 20-amp 120-volt power source designed for redundancy costs an additional $175 per month.
- Finally, a connection to the ultra-fast communication network linking various data centers around the world runs an additional $325 per month.

The grand total of the proximity setup adds up to just $2,350 per month, a negligible cost for any serious investor.

Network Throughput

The messaging architecture is a resource available to all market participants, yet it is not free to all. Exchanges, for example, have to continuously enhance their infrastructure to ensure that the bandwidth of their connections is broad enough to allow uninhibited message traffic among all interested traders. Perhaps the biggest challenge to exchanges and other order-matching venues is the sheer volume of order cancellations. According to Hautsch and Huang (2011), on Nasdaq, 95 percent of all limit orders are canceled within one minute from the time the orders are placed. Hasbrouck and Saar (2011) report similar activity grouped into brisk order placement and cancellation "runs." While this activity may seem malicious to an uninitiated observer, the explanation for such behavior is quite simple: as described in detail in Chapters 10, automated market makers need to quote close to the market price—"stay on top of the book"—in order to be successfully and promptly matched, thus ensuring a steady revenue stream. Once the market moves away from the quotes of the market maker, it is in the best interests of the market maker to cancel the orders and to resubmit them at the new best bid and best offer prices. In addition, as explained in Chapter 12, on exchanges with time-price priority of limit order books, market participants may place and then innocuously cancel excessive numbers limit orders to secure their execution priority, in a practice known as "layering."

b6f1260bee54f438e54f5e8d966ba6d2
ebrary

Aldridge, Irene. Wiley Trading Ser. : High-Frequency Trading : A Practical Guide to Algorithmic Strategies and Trading Systems (2). Somerset, US: John Wiley & Sons, Incorporated, 2013. ProQuest ebrary. Web. 8 March 2017.
Copyright © 2013. John Wiley & Sons, Incorporated. All rights reserved.

b6f1260bee54f438e54f5e8d966ba6d2
ebrary

In such dynamics, many trading venues are caught within a vicious circle: on the one hand, they are competing to attract the market makers, but on the other, many order-cancelling market makers are eroding network resources, resulting in missed quotes and other delays for all market participants. Even the co-location does not help fully navigate the bandwidth issue, as co-location space also faces capacity constraints: Nasdaq has seen so much demand in its co-location hangar in Mahwah, New Jersey, that it is reportedly running out of space to offer to the parties interested in co-locating there. As discussed in Chapters 3 and 12, a promising solution to the network bandwidth issue, known as *pro-rata execution*, has been developed and implemented at selected exchanges.

Software

b6f1260bee54f438e54f5e8d966ba6d2
ebrary

High-frequency trading systems are ultimately software applications deployed over hardware and messaging described above. As with any software system, an HFT system begins with an idea, known as an *algorithm* in computer-science lingo, that is subsequently coded in a chosen computer language into a full-blown software program. The term *algorithm* is properly defined as logic, or a sequence of high-level actions, developed to explain to a computer how to implement a given task at hand. The algorithm does not delve into the details of actual coding or programming the system, but may still take into account the idiosyncrasies of the hardware and messaging structure on which the algorithm will be ultimately implemented. Algorithms are often visualized in diagrams. The key elements of the algorithm diagrams are summarized in Figure 2.12.

The algorithm elements shown in Figure 2.12 will be used throughout the book to explain algorithm designs of common HFT strategies. Figure 2.13 illustrates the step-by-step process of the following simple market-making algorithm:

1. Begin program.
2. Check market conditions: Are market conditions suitable for market making?
3. If yes, start market making.
4. If no, wait one minute.
5. Repeat step 2.

b6f1260bee54f438e54f5e8d966ba6d2
ebrary

The algorithm presented in Figure 2.13 is "nested," or comprises two additional algorithms marked in Figure 2.13 only as "Check market conditions" and "Start market making." The nested tasks can be the critical "secret sauce" that distinguishes good HFT systems from bad ones. Usually, the tasks are designed on the basis of advanced research, where the task is selected among several competing ideas given

FIGURE 2.12 Common Elements of an Algorithm

b6f1260bee54f438e54f5e8d966ba6d2
ebrary

Aldridge, Irene. Wiley Trading Ser. : High-Frequency Trading : A Practical Guide to Algorithmic Strategies and Trading Systems (2). Somerset, US: John Wiley & Sons, Incorporated, 2013. ProQuest ebrary. Web. 8 March 2017.
Copyright © 2013. John Wiley & Sons, Incorporated. All rights reserved.

b6f1260bee54f438e54f5e8d966ba6d2
ebrary

positive results of rigorous testing. The two nested tasks shown in Figure 2.13 are explained in detail in Chapter 15.

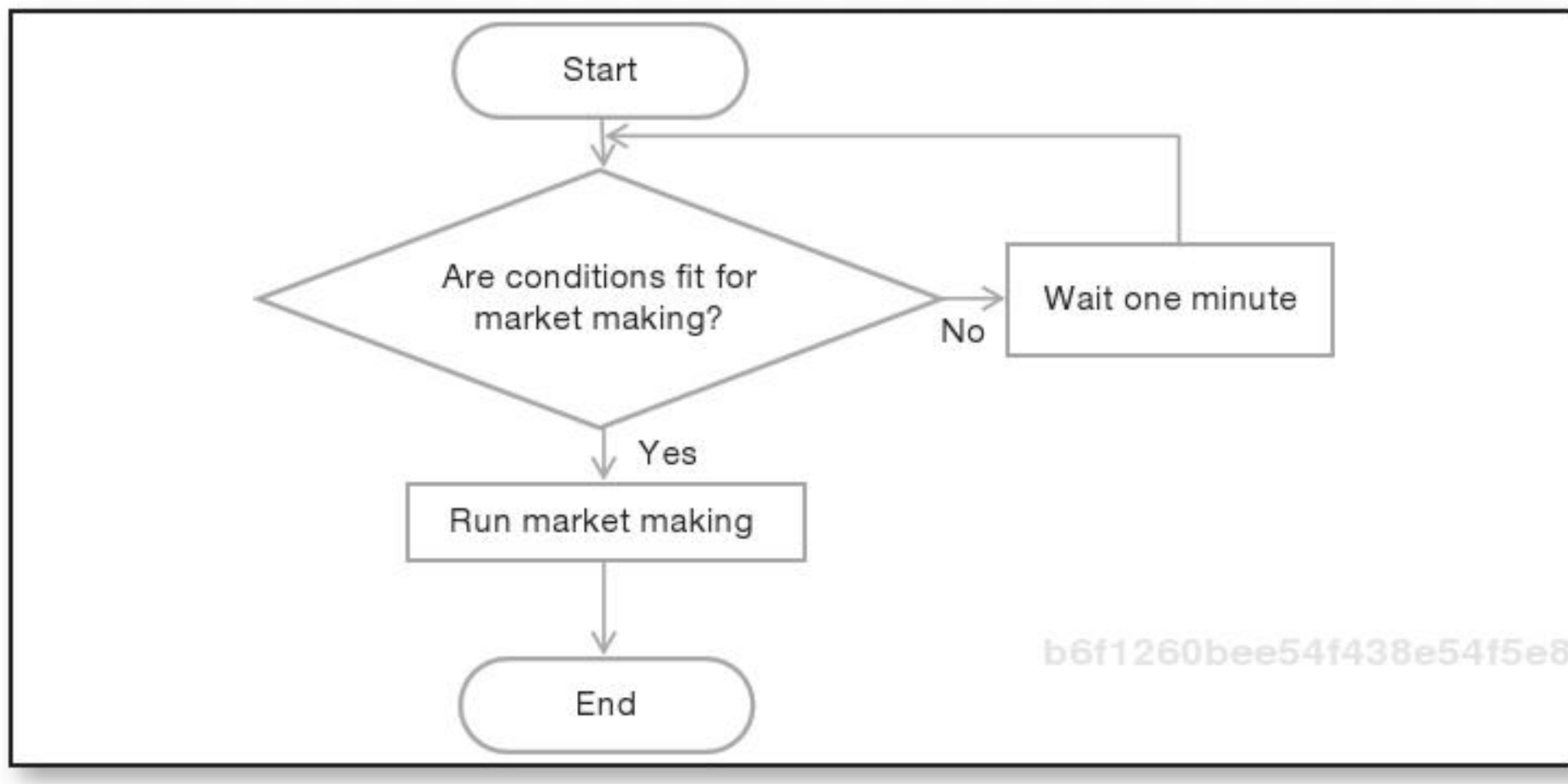


FIGURE 2.13 Sample Market-Making Algorithm

The term *algorithm* is often used synonymously with the terms *high-frequency trading, systematic trading, electronic trading,* and *low-latency trading*. The distinctions among the terms, however, are significant enough to warrant some explanation. A *system* usually refers to a highly methodical approach to a process. Systematic trading, therefore, is following some rigid frameworks, but does not have to be fully automated. A trader can be considered systematic if he manually places trades when certain indicators form a specific pattern. The term *systematic* was coined to distinguish traders practicing methodical allocations from traders using their intuition or discretion, and hence known as *discretionary* traders. All high-frequency and algorithmic traders are also systematic.

The term *electronic* describes the execution preferences of the trader: whether he chooses to place orders electronically or, perhaps, over the telephone. All high-frequency trading, algorithmic trading, and low-latency trading are necessarily electronic, but systematic trading may involve nonelectronic components. The reverse, however, does not have to be the case; many electronic trading systems route only orders that may or may not be placed algorithmically. As most markets and traders are moving on to electronic platforms, however, the term *electronic trading* is becoming implicit and obsolete.

Low-latency trading refers to trading that utilizes fast connectivity between traders and exchanges. As described in the previous section, latency measures the time distance between the trader and the exchange. Most latency measurements are currently recorded in microseconds. High-frequency trading systems often also happen to be low-latency, but the reverse does not have to hold: low-latency systems are often deployed by low-frequency traders to obtain better prices on trades.

Once an algorithm is designed, it is broken down into components and coded in a language understood by computers. The goal of coding is to accurately translate the logic of the algorithm into computer "speak" and, in the process, to create as little delay as possible during "run time," when the computer will read and interpret

Aldridge, Irene. Wiley Trading Ser. : High-Frequency Trading : A Practical Guide to Algorithmic Strategies and Trading Systems (2). Somerset, US: John Wiley & Sons, Incorporated, 2013. ProQuest ebrary. Web. 8 March 2017.
Copyright © 2013. John Wiley & Sons, Incorporated. All rights reserved.

b6f1260bee54f438e54f5e8d966ba6d2
ebrary

the code. The code written directly to FPGA chips is presently the fastest. Still, many high-frequency traders deploy standard non-FPGA architecture and rely on languages such as C++ and Java to code their systems. While C++ remains the fastest computer language easily understood by humans, many systems are coded in Java with workarounds of its slowest components. Thus, for example, the famed Nasdaq OMX system is reportedly coded in Java with Java garbage collection disabled and replaced with C++-like direct memory access for increased speed. Chapter 16 describes best practices of coding implementation.

The code outlining the actual trading logic of an algorithm tends to be quite short. In many successful cases, the trading logic comprises as few as 50 lines of code. In addition to the actual decisions to buy and sell, however, every HFT system incorporates supporting quote data retrieval functionality that may number in 10,000+ lines of code, as well as the trade send-off and acknowledgment receipt applications that can also take as much as 5,000 lines of code. Perhaps the most lengthy, yet mandatory, component of each HFT system is its risk management checks and balances, which can total 50,000+ lines of code. Risk management of HFT is discussed in detail in Chapter 14 of this book.

b6f1260bee54f438e54f5e8d966ba6d2
ebrary

Summary

Algorithmic execution is inseparable from today's markets. It is a necessary function that delivers considerable value to all investors, large and small. With plummeting technology costs, most investors today can afford to build and use advanced algos, including algos designed for high-frequency trading, previously available only to a select few market participants. Services such as co-location provide added benefits of security and speed.

End-of-Chapter Questions

b6f1260bee54f438e54f5e8d966ba6d2
ebrary

1. Would you encrypt your trading orders before transmitting them to the execution venue over the Internet? Explain.
2. Mr. Smith has read about the "arms race" of computer technology in the financial services industry and decides to invest into the latest super computer to increase the odds of fast order transmission. Is Mr. Smith's investment justified? Where does most message congestion occur in the cyber-universe today?
3. What is co-location?
4. On average, how much slower is the transmission of trade order messages in comparison with quote messages?
5. What is a heartbeat?
6. The best offer on exchange A contains 300 units of instrument X, the best offer on exchange B contains 500 units, and the best offer on exchange C contains just 100 units. Your customer wants you to buy 550 units on his behalf. How would you break up the customer's order and send them to exchanges under the minimal impact algorithm?

b6f1260bee54f438e54f5e8d966ba6d2
ebrary

Aldridge, Irene. Wiley Trading Ser. : High-Frequency Trading : A Practical Guide to Algorithmic Strategies and Trading Systems (2). Somerset, US: John Wiley & Sons, Incorporated, 2013. ProQuest ebrary. Web. 8 March 2017.
Copyright © 2013. John Wiley & Sons, Incorporated. All rights reserved.

b6f1260bee54f438e54f5e8d966ba6d2
ebrary

b6f1260bee54f438e54f5e8d966ba6d2
ebrary

b6f1260bee54f438e54f5e8d966ba6d2
ebrary

b6f1260bee54f438e54f5e8d966ba6d2
ebrary

Aldridge, Irene. Wiley Trading Ser. : High-Frequency Trading : A Practical Guide to Algorithmic Strategies and Trading Systems (2). Somerset, US: John Wiley & Sons, Incorporated, 2013. ProQuest ebrary. Web. 8 March 2017.
Copyright © 2013. John Wiley & Sons, Incorporated. All rights reserved.

b6f1260bee54f438e54f5e8d966ba6d2
ebrary

CHAPTER 3

Market Microstructure, Orders, and Limit Order Books

b6f1260bee54f438e54f5e8d966ba6d2
ebrary

The study of market microstructure originated over four decades ago, and the core principles remain true today: most market participants rely on limit orders and market orders. These core order types remain most used despite an explosion of various derivative order types. Despite the continuity of these key principles still, much has changed. In the 1980s At one time, the peculiarities differentiating tick data dynamics from daily and monthly data were observed, but there was no way to incorporate the differences into a trading strategy. Today, that tick data can be captured easily and trades initiated quickly, it is possible to build trading strategies taking into account market microstructures. This chapter delves into the modern microstructure of markets, describing orders, matching processes, and rebate structures, among other issues defining today's markets.

b6f1260bee54f438e54f5e8d966ba6d2
ebrary

Types of Markets

Financial markets are venues that allow investors and other market participants to buy and sell securities with a peace of mind that all transactions will be properly accounted and settled. In this respect, financial markets are not that different from markets for other nonfinancial products, such as a neighborhood grocer. When a customer enters a grocery store, he expects immediate execution of his transaction, an exchange of his money for merchandise—food. The grocer's cash register takes

b6f1260bee54f438e54f5e8d966ba6d2
ebrary

Aldridge, Irene. Wiley Trading Ser. : High-Frequency Trading : A Practical Guide to Algorithmic Strategies and Trading Systems (2). Somerset, US: John Wiley & Sons, Incorporated, 2013. ProQuest ebrary. Web. 8 March 2017.
Copyright © 2013. John Wiley & Sons, Incorporated. All rights reserved.

b6f1260bee54f438e54f5e8d966ba6d2
ebrary

on the settlement function: the receipt itemizes the produce the customer bought and the total amount of money the grocer collected, nowadays most often from the customer's account rather than in cash form.

Furthermore, the grocery customer expects that the food he acquires is in good condition and that the transaction is permanent: that the grocer has full rights to sell the customer the product. Most centralized financial markets incorporate similar quality control for financial products: the products offered for purchase and sale on the floors of each market tend to be standardized, and their quality can be ensured via a thorough prelaunch due-diligence process as well as via the real-time data distribution to all market participants. For example, futures contracts traded on exchanges have a specific well-defined structure. Initial public offerings for stocks tend to be well scrutinized. Yet, just like grocers do not evaluate suitability of particular foods to each person, financial markets are unaware of a client's risk profile or accreditation, leaving the risk of investing decisions to the traders themselves.

b6f1260bee54f438e54f5e8d966ba6d2
ebrary

Since the beginning of trading history, most financial and nonfinancial markets have been organized by product type. In nonfinancial settings, stores carrying lumber differ from stores selling clothes. Food markets used to be highly fragmented by type as well: fishmongers, bakers, butchers, and ice-cream makers all used to have their own shops, if not districts. Similarly, financial markets have been historically organized by the type of the traded instrument, each trading only equities, futures, options, or other securities. Recently, however, nonfinancial markets have been gearing toward megastores to increase their efficiency and take advantage of common distribution and procurement frameworks and skills. Similarly, many trading venues are now venturing in the cross-asset space, with the traditional foreign exchange player iCap launching a fixed-income trading offering, and equity exchange Best Alternative Trading Systems (BATS) considering entering the foreign exchange trading space.

The history of the financial exchanges is not always linear. The New York Stock Exchange (NYSE), for example, was first formed as a for-profit entity at the end of the eighteenth century, when two dozen stockbrokers signed an agreement to stop undercutting each other's commissions and instead retain at least 0.25 percent of each trade. In subsequent years, as the U.S. economy grew, more public listings became available, and the interest in investing grew from the U.S. public, stockbroking became increasingly lucrative and seats on the NYSE rose in value.

b6f1260bee54f438e54f5e8d966ba6d2
ebrary

During major recessions, however, lavish lifestyles of stockbrokers drew the scrutiny of money-losing investors and regulators. In 1934, in the midst of the Great Depression, the NYSE was required to register and accept oversight from the then newly formed U.S. Securities and Exchange Commission (SEC). In 1971, during the post–Vietnam War recession, the NYSE was converted into a not-for-profit entity in a bid to cap excessive compensation of brokers and transfer some of their intake to the investors. The nonprofit mandate failed to live up to its expectation, with exchange-registered brokers creating a secondary market for their seats on the exchange that by the 1990s had reached multiple-million-dollar price tags.

The past two decades have witnessed increasing competition in the exchange space, and these competitive market forces appear to have succeeded in helping investors retain a larger share their earnings away from brokers. New exchanges,

b6f1260bee54f438e54f5e8d966ba6d2
ebrary

Aldridge, Irene. Wiley Trading Ser. : High-Frequency Trading : A Practical Guide to Algorithmic Strategies and Trading Systems (2). Somerset, US: John Wiley & Sons, Incorporated, 2013. ProQuest ebrary. Web. 8 March 2017.
Copyright © 2013. John Wiley & Sons, Incorporated. All rights reserved.

b6f1260bee54f438e54f5e8d966ba6d2
ebrary

such as Nasdaq, deployed technology in lieu of bonus-collecting human traders, transferring some of the resulting savings into the pockets of investors and producing other benefits in the process. Technology has lowered error rates, increased execution times, and, perhaps most importantly, increased transparency of process to investors. Other early exchanges, such as the Chicago Mercantile Exchange, followed trajectories similar to that of NYSE, and today face competition from relatively new exchange entrants, such as the Intercontinental Commodity Exchange (ICE).

Today's equity markets comprise over a dozen various exchange venues, run by now-industry stalwarts such as the NYSE and Nasdaq, and relatively recent arrivals, such as BATS, DirectEdge, and others. Since all equity exchanges are subject to U.S. SEC oversight and trade standardized products, exchanges presently compete to differentiate their offerings mainly on liquidity and costs. In a bid to attract liquidity (more on this later in this chapter), most exchanges have done away with membership fees and now offer free registration. In addition, equity exchanges are differentiating themselves on price structures, not only rewarding large traders but also segmenting traders based on whether they place market orders or limit orders.

b6f1260bee54f438e54f5e8d966ba6d2
ebrary

In addition to the regular exchanges, a new breed of matching venues has emerged, known as *dark pools.* Unlike an exchange, where the entire limit order book is available for observation, dark pools do not disclose their limit order books, instead keeping them "in the dark."Trading in a dark limit order book has appealed to large investors who are concerned about the information they may reveal by placing their orders in a traditional exchange, a *lit* market. Liquidnet is an example of an equities dark pool.

Limit Order Books

The cumulative trade size of all limit orders available to meet incoming market orders at any given time on a specific trading venue is known as *liquidity.* The larger the number of limit order traders available on the exchange, and the larger the size of each trader's limit orders, the more liquid the given trading venue. Liquidity is also necessarily finite in today's markets: the number of limit orders is measurable, and each limit order has a finite size. Liquidity was first defined by Demsetz (1968).

b6f1260bee54f438e54f5e8d966ba6d2
ebrary

To account for limit orders, the majority of contemporary exchanges are organized as so-called centralized limit order books (CLOBs), also referred to as a double-sided auction. The CLOBs were pioneered in the United States in the early 1970s and adopted in Europe in the 1980s. In a CLOB model, all incoming limit orders are recorded in a "book": a table with columns corresponding to sequential price increments, and rows recording sizes of limit orders posted at each price increment. Figure 3.1 illustrates the idea. The limit order book information can be distributed to all other market participants as Level II data, discussed in detail in Chapter 4.

In theory, limit order books are often assumed to be symmetric about the market price, with the distribution of limit buy orders mirroring that of limit sell orders. Furthermore, in many risk management applications, order books are also assumed

b6f1260bee54f438e54f5e8d966ba6d2
ebrary

Aldridge, Irene. Wiley Trading Ser. : High-Frequency Trading : A Practical Guide to Algorithmic Strategies and Trading Systems (2). Somerset, US: John Wiley & Sons, Incorporated, 2013. ProQuest ebrary. Web. 8 March 2017.
Copyright © 2013. John Wiley & Sons, Incorporated. All rights reserved.

b6f1260bee54f438e54f5e8d966ba6d2
ebrary

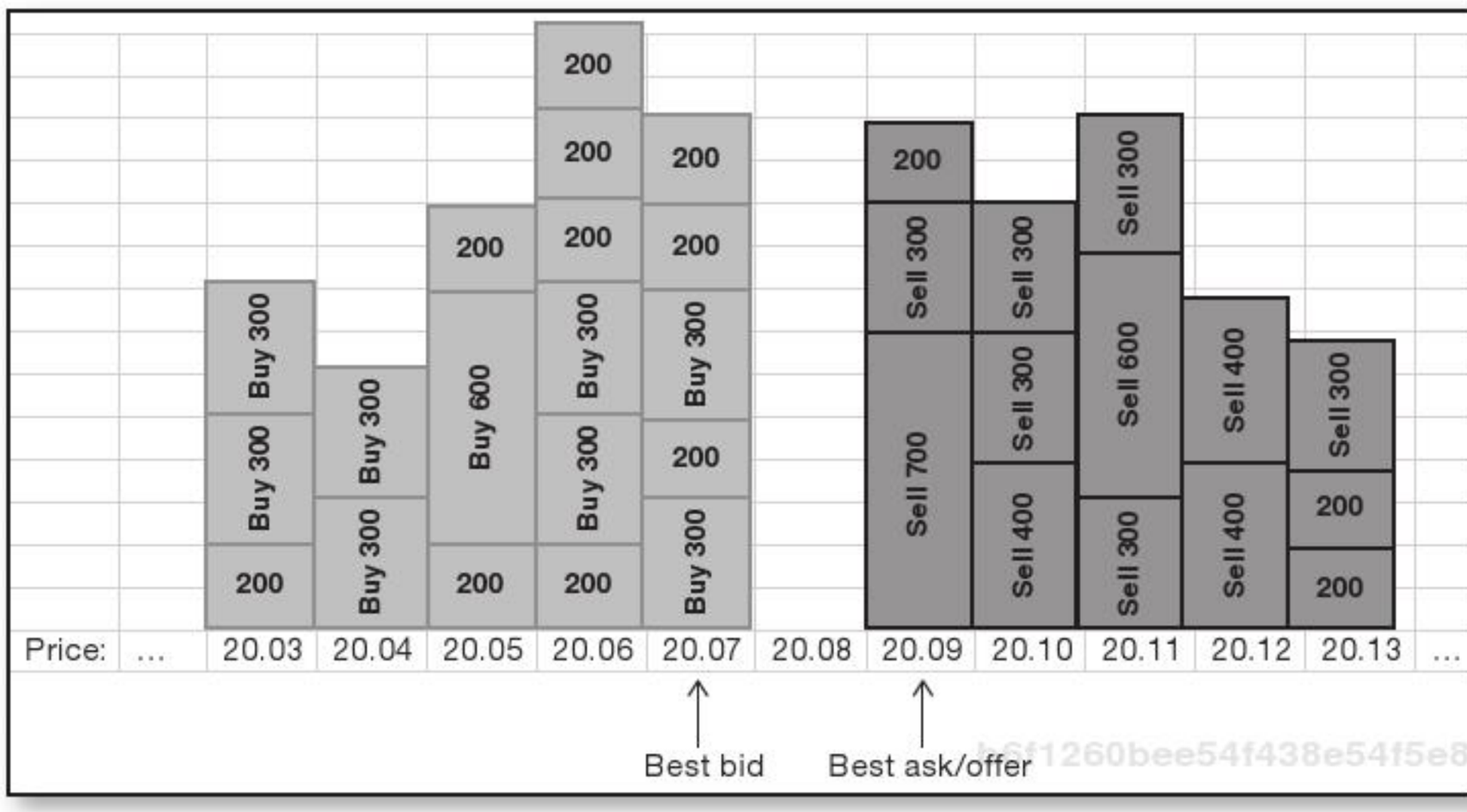


FIGURE 3.1 Sample Snapshot of a Limit Order Book. All limit buy orders are on the left-hand side of the book; all limit sell orders are on the right-hand side of the book.

to follow normal bell-curve distributions. Neither of the assumptions tends to hold: order books are seldom normal and are often asymmetric.

As Figure 3.2 shows, when a new limit order arrives, it is placed into a *limit order queue* corresponding to its price. Since all prices in today's markets are subject to a minimum increment, or *tick*, price-based bins are clearly delineated. The limit buy orders at the highest price form the best bid, with the price of these orders reflected in the best bid price, and the aggregate size reported as the best bid size. Similarly, the limit sell orders posted at the lowest price form the best ask, with respective price and size information. Best ask is sometimes referred to as *best offer*. At any given moment of time, there exists a finite aggregate size of all limit orders posted at each price.

When a market buy order arrives, it is matched with the limit sell orders, beginning with those placed at the best ask price. If the size of the incoming market buy order is greater than the size of the best ask queue, the market order "sweeps" through other offer queues in the direction of increasing price, "eating up" liquidity available at those price ticks. Sweeping leaves a significant gap in limit orders on the ask side, instantaneously increasing the bid-ask spread, and potentially inducing slippage in subsequent market buy orders. The order-matching process is similar for market sell orders that end up matched with the available limit buy orders aggregated on the bid size of the book. Limit buy orders with the prices equal or higher than the prevailing best bid are executed like market buy orders. Similarly, low-priced limit sell order are usually treated as market sell orders.

If the size of the incoming buy order is smaller than the size of the best ask, and the aggregate best ask queue is composed of several limit sell orders placed at the best ask price, the decision of which of the limit sell orders is matched against the market buy order may differ from one exchange to another. While most exchanges at present practice price-time priority, also known as the first-in-first-out (FIFO)

b6f1260bee54f438e54f5e8d966ba6d2
ebrary

Aldridge, Irene. Wiley Trading Ser. : High-Frequency Trading : A Practical Guide to Algorithmic Strategies and Trading Systems (2). Somerset, US: John Wiley & Sons, Incorporated, 2013. ProQuest ebrary. Web. 8 March 2017.
Copyright © 2013. John Wiley & Sons, Incorporated. All rights reserved.

b6f1260bee54f438e54f5e8d966ba6d2
ebrary

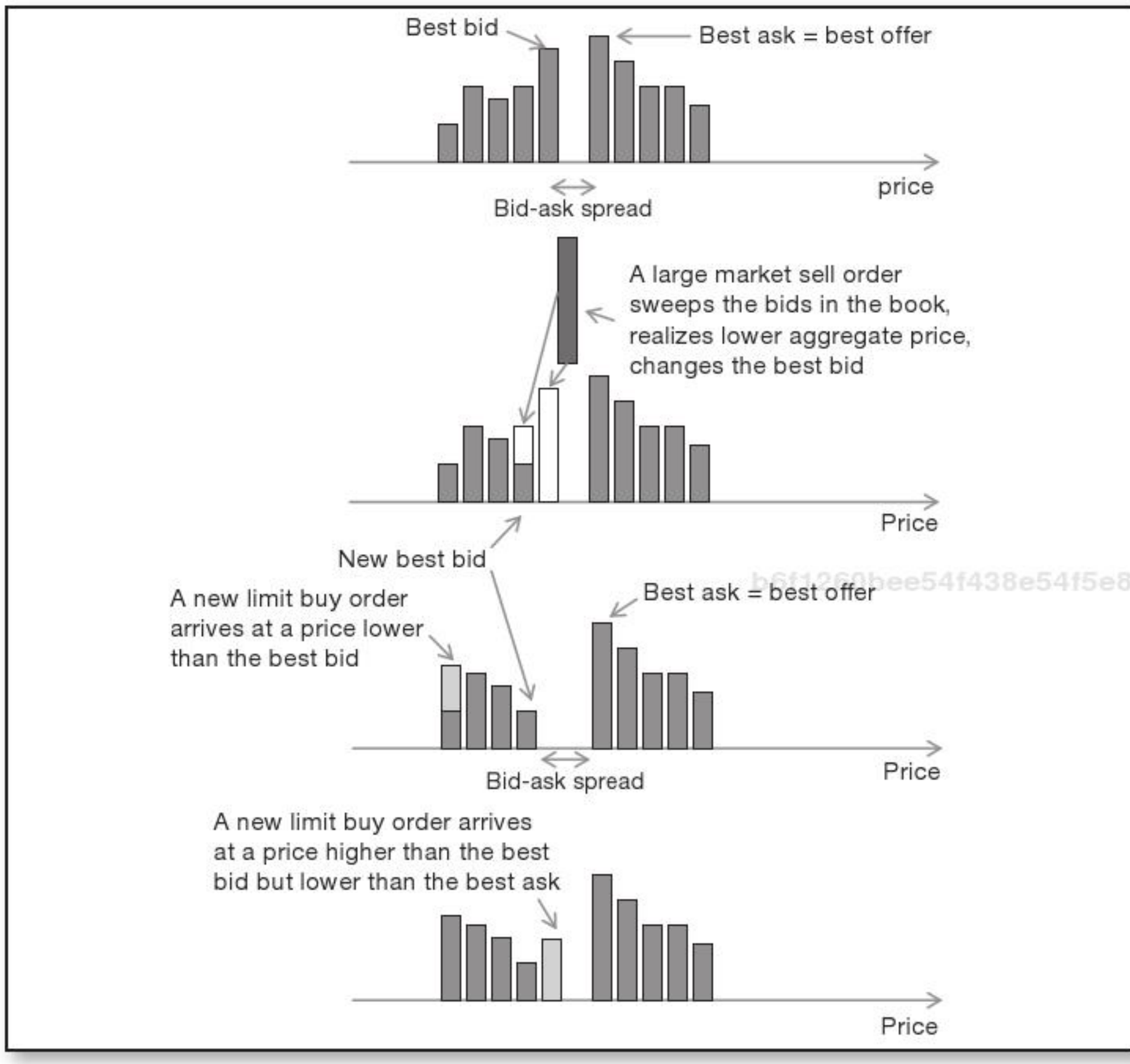


FIGURE 3.2 Sample Dynamics in a Limit Order Book

execution schedule for limit orders, several other exchanges now match a fixed proportion of each limit order at a given price in a process known as pro-rata matching.

In time-price priority, or FIFO, execution, the limit order that arrived first is also the first of that price bin to be matched with the incoming market order. Figure 3.3 illustrates the FIFO matching process. FIFO, known as the continuous auction, has been shown to enhance transparency of trading via the following measures (see Pagano and Roell, 1996; Jain, 2005; and Harris, 2003):

- Reducing information asymmetry—all traders have access to the limit order book information.
- Enhancing liquidity—a CLOB structure incentivizes traders to add limit orders, thereby increasing market liquidity.
- CLOB's organization supports efficient price determination by providing a fast and objective order-matching mechanism.
- Uniform rules for all market participants ensure operational fairness and equal access.

Aldridge, Irene. Wiley Trading Ser. : High-Frequency Trading : A Practical Guide to Algorithmic Strategies and Trading Systems (2). Somerset, US: John Wiley & Sons, Incorporated, 2013. ProQuest ebrary. Web. 8 March 2017.
Copyright © 2013. John Wiley & Sons, Incorporated. All rights reserved.

b6f1260bee54f438e54f5e8d966ba6d2
ebrary

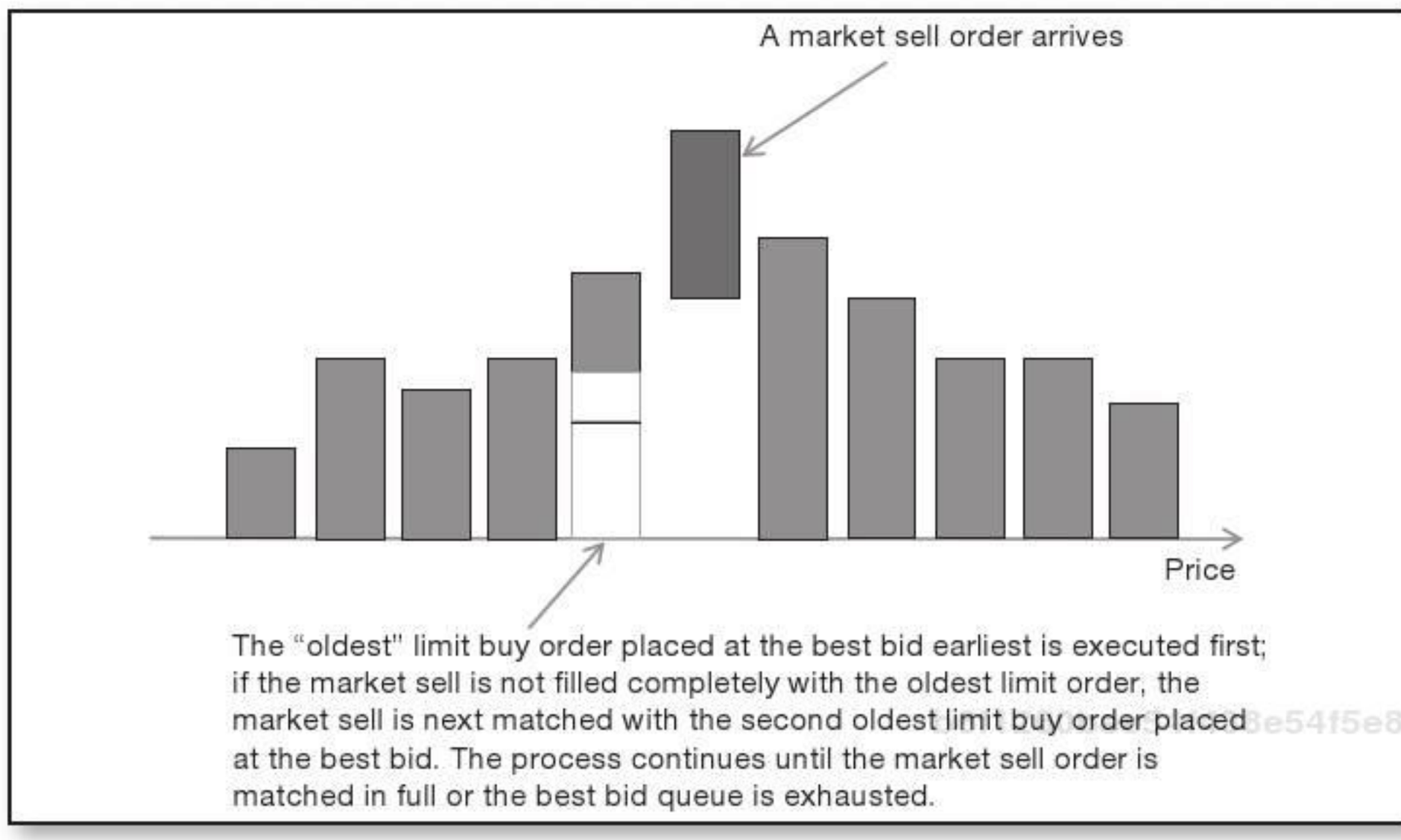


FIGURE 3.3 Price-Time Priority Execution

While most execution venues are based on FIFO, some exchanges, like Chicago Mercantile Exchange (CME), Chicago Board Options Exchange (CBOE), Philadelphia Stock Exchange (PHLX), and Intercontinental Commodity Exchange (ICE) have switched to pro-rata execution schedules. CME's pro-rata schedule matches an incoming market buy order with a fixed proportion of each limit order posted at the best ask. Similarly, an incoming sell order is matched with equal fractions of all of the limit orders posted at the best bid. The larger the limit order at the best bid and the best ask, therefore, the larger the fill of that order. Figure 3.4 shows the pro-rata process.

b6f1260bee54f438e54f5e8d966ba6d2
ebrary

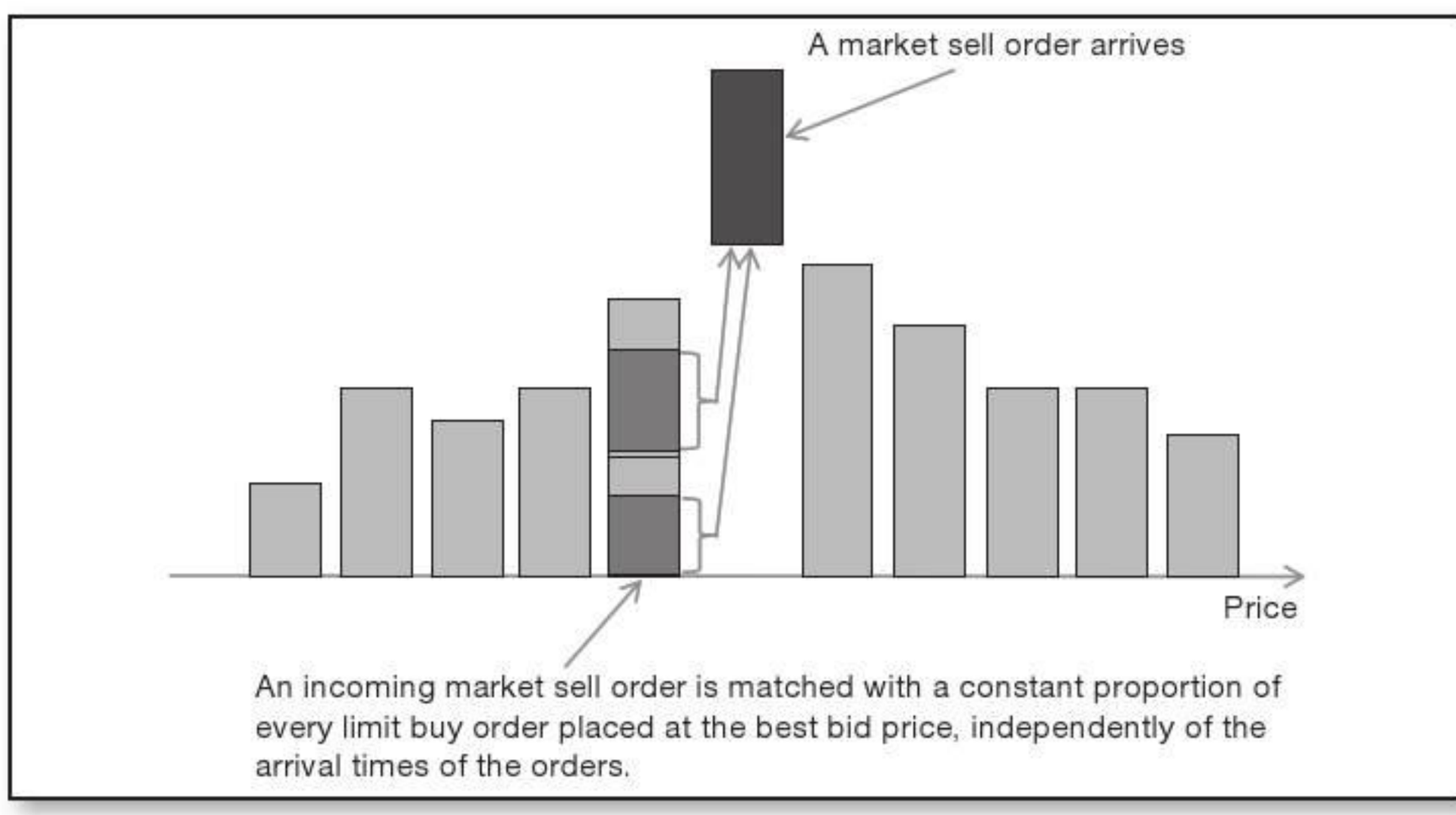


FIGURE 3.4 Pro-Rata Execution

b6f1260bee54f438e54f5e8d966ba6d2
ebrary

Aldridge, Irene. Wiley Trading Ser. : High-Frequency Trading : A Practical Guide to Algorithmic Strategies and Trading Systems (2). Somerset, US: John Wiley & Sons, Incorporated, 2013. ProQuest ebrary. Web. 8 March 2017.
Copyright © 2013. John Wiley & Sons, Incorporated. All rights reserved.

b6f1260bee54f438e54f5e8d966ba6d2
ebrary

The main advantage of the pro-rata matching from the exchange point of view is the built-in incentives for traders to place large limit orders, and, therefore, to bring liquidity to the exchange. The pro-rata matching encourages traders to post large limit orders without special compensation like rebates discussed below, thereby increasing exchange profitability. In addition, pro-rata matching eliminates incentives to place and then cancel limit orders with the intent to secure time priority of execution, reducing message traffic to and from the exchange.

In a nutshell, the pro-rata incentive works as follows: a trader desiring to execute a limit order knows that only a fraction of his order will be filled under the pro-rata schedule. The exact size of the filled portion of the order will depend on the cumulative size of limit orders placed at the same price by other limit order traders. The higher the aggregate size of limit orders in a specific price bin, the lower the percentage of all orders that will be filled in response to an incoming market order of the opposite sign. To increase his chances of filling the entire order, therefore, the trader is likely to place a limit order with a larger size than his intended order, with the explicit hope that the fraction of the order that will get filled is of the exact size as his intended order.

b6f1260bee54f438e54f5e8d966ba6d2
ebrary

Aggressive versus Passive Execution

Orders can be described as passive or aggressive. Aggressive orders do not mean malicious orders, and passive orders do not indicate orders that are destined to be taken advantage of. Instead, the aggressiveness or passivity of orders refers to the proximity of the order price to the prevailing market price. Figure 3.5 illustrates the concept.

FIGURE 3.5 Aggressive and Passive Orders: An Illustration

b6f1260bee54f438e54f5e8d966ba6d2
ebrary

A limit order far away from the market price (a low-priced limit buy order, or a high-priced limit sell order) is considered passive. The closer the limit order is to the market price, the more aggressive is the order. A market order is the most aggressive order, "crossing the spread" to be matched with the best-priced limit order on the opposite side of the limit order book. Limit orders crossing the spread are treated like market orders in the execution queue, and are also considered aggressive.

While market orders enjoy immediate and nearly guaranteed fills, market orders cross and pay the spread, incur transaction fees of the exchange and broker-dealers, and face price uncertainty. In today's markets, price uncertainty can be the costliest component associated with the market order execution. From the time the market

b6f1260bee54f438e54f5e8d966ba6d2
ebrary

Aldridge, Irene. Wiley Trading Ser. : High-Frequency Trading : A Practical Guide to Algorithmic Strategies and Trading Systems (2). Somerset, US: John Wiley & Sons, Incorporated, 2013. ProQuest ebrary. Web. 8 March 2017.
Copyright © 2013. John Wiley & Sons, Incorporated. All rights reserved.

b6f1260bee54f438e54f5e8d966ba6d2
ebrary

order is placed to the time the execution is recorded, the market price may "slip," resulting in worse execution than the prevailing price at the time the market order was placed. The slippage may be due to several factors:

- Several market orders may arrive at the exchange and be executed between the time a given market order is placed and the time it is executed. Each of the arriving market orders may deplete the matching liquidity in the order book, adversely moving the market price. Such a situation is particularly common at times of news releases, when many traders and their algorithms simultaneously process information and place orders in the same direction.
- A market order that is large relative to the available depth of the order book may sweep through the book, executing fractional pieces of the order against limit orders at different price levels.
- Additional market conditions, such as market disruptions, may also result in significant slippage.

b6f1260bee54f438e54f5e8d966ba6d2
ebrary

By contrast, the price of a limit order is fixed when the order is placed. A limit order is added to the limit order book, where it "sits" until the prevailing market price reaches it and a market order is executed against it. Limit orders also generally avoid "crossing the spread," a cost of paying the market spread incurred by market orders. Highly aggressive limit orders executed as market orders cross the spread, but obtain as good or better execution price than their specified limit price. Limit orders are also subject to positive or negative transaction costs, which vary from one trading venue to another.

For all their price advantages, limit orders are subject to an important risk—the risk of nonexecution. A limit order is executed only when it is matched with a market order of the opposite direction. A market price may quickly move away from a limit order, leaving it unexecuted. An unexecuted limit order may present a particular problem when placed to close a position, and misses the opportunity to eliminate market risk of the trade. And yet, unexecuted limit orders placed to open a position also incur a cost, that of the opportunity to engage in the trading strategy.

b6f1260bee54f438e54f5e8d966ba6d2
ebrary

Complex Orders

In response to competition from new entrants in the matching business, trading venues have diversified their order offerings. For example, in response to competition from dark pools, selected exchanges expanded the number of available orders, creating so-called *iceberg orders*. Iceberg orders allow limit-order traders to display only a portion of their order in the limit order book, and keep the rest of their liquidity in the dark. In FIFO limit order books, iceberg orders are executed on a time priority basis: when matched against a smaller order, the nonexecuted part of the iceberg is being placed back at the end of their limit-order book queue. Unlike the orders in a dark pool, the size information of an iceberg is revealed after the iceberg is matched in part or in full: the matched size is disseminated to other traders as a trade tick. As a rule, iceberg orders cost more than do limit and market orders.

b6f1260bee54f438e54f5e8d966ba6d2
ebrary

Aldridge, Irene. Wiley Trading Ser. : High-Frequency Trading : A Practical Guide to Algorithmic Strategies and Trading Systems (2). Somerset, US: John Wiley & Sons, Incorporated, 2013. ProQuest ebrary. Web. 8 March 2017.
Copyright © 2013. John Wiley & Sons, Incorporated. All rights reserved.

b6f1260bee54f438e54f5e8d966ba6d2
ebrary

Other specialized orders have sprung up as well, to generate additional revenues from higher transaction costs and to serve the following potential needs of customers:

- *Limit risk.* Most trading venues and broker-dealers now offer a range of orders for containing market risk. The order examples include hard and trailing stop orders, where the position is liquidated when a price move exceeds the predetermined threshold in the adverse direction (see Chapter 14 for more information on stops).
- *Speed of execution.* Orders in this category try to enable the fastest execution possible and include, in addition to the vanilla market order, a market-on-close order that often guarantees to catch the closing price, a midpoint match order that bests the best bid limit order by attempting to negotiate crossing only half of the prevailing spread, and the sweep-to-fill order simultaneously clears the order book of the size requested in the order. The sweep-to-fill order may be executed faster than the market order, since a large market order is often executed by sweeping the limit order book gradually over time.
- *Price improvement.* Such orders include a block order in options that improves the price by obtaining large-volume discounts on transaction costs.
- *Privacy.* Privacy-providing orders deliver dark liquidity, and include iceberg orders and hidden orders, among others. The hidden order, as its name suggests, is not displayed in the limit order books. The iceberg order displays a limited portion of the order in the limit order book, as discussed in the beginning of this section.
- *Time to market.* Orders in the time-to-market group include fill-or-kill orders for market orders desiring the most immediate liquidity. A fill-or-kill order is canceled if the matching liquidity is not immediately available. Conversely, a good-till-canceled limit order falling in the same order category is kept in the limit order book until it is canceled or another maximum period of time set by the trading venue (e.g., a quarter).
- *Advanced trading.* These orders include additional quantitative triggers, such as implied volatility in options.
- *Algorithmic trading.* Orders in this category offer execution via order-slicing algorithms, such as percentage of volume (POV), described in detail in Chapter 17.

Trading Hours

Traditionally, many trading venues operated from 9:30 a.m. to 4:00 p.m. Eastern time. In today's globalized markets, more effort is placed on expanding accessibility of trading. As a result, many exchanges today offer premarket trading and afterhours trading, cumulatively known as *extended-hours trading*. Extended hours in equities, for example, allow market access from 4:00 a.m. to 8:00 p.m. Eastern time. The extended-hours volume is considerably thinner than that observed during the normal trading hours. Still, selected brokers use the after-hours trading to fill their customers' market-on-close orders.

Aldridge, Irene. Wiley Trading Ser. : High-Frequency Trading : A Practical Guide to Algorithmic Strategies and Trading Systems (2). Somerset, US: John Wiley & Sons, Incorporated, 2013. ProQuest ebrary. Web. 8 March 2017.
Copyright © 2013. John Wiley & Sons, Incorporated. All rights reserved.

b6f1260bee54f438e54f5e8d966ba6d2
ebrary

Modern Microstructure: Market Convergence and Divergence

The electronization of markets has left an indelible footprint on all modern markets, streamlining some aspects of trading and fragmenting others. Among the trends in market convergence are the following developments:

- Most markets today can be accessed via Financial Information eXchange (FIX) messaging protocol. The FIX protocol is an XML-like specification that allows market participants to send and receive quotes, orders, and order cancellation and execution acknowledgments, among other messages necessary for fast and efficient trading. The FIX protocol is administered by an independent not-for-profit body, further facilitating proliferation of the protocol.
- Most markets worldwide are now designed as limit order books (LOBs). The Singaporean Stock Exchange was one of the last entities to use a different market structure, but converted to LOB in the past decade.

Among the key trends in market divergence is the progressing fragmentation of markets among the asset classes:

- Equities are subject to the National Best Bid and Offer (NBBO) rule, whereas all equities are to be executed at the aggregated and disseminated NBBO or better price. If the exchange is unable to execute an incoming market order at the NBBO, the exchange is obligated to route the order to an exchange with NBBO quotes.
- Futures exchanges do not have centralized pricing, but are subject to unique margining and daily mark-to-market requirements.
- Foreign exchange markets do not have centralized quotes or exchanges at all. All trades are continued to be negotiated over the counter (OTC), even though many OTC platforms are now fully electronic. Yet selected large market participants may be granted access to an interdealer network, an exchange-like entity.
- Option markets are numerous, with little activity on average.
- Following the Dodd-Frank regulation, new asset classes such as fixed income and swaps will be coming online or expanding in electronic forms. Each of the asset classes has distinct peculiarities, resulting in further fragmentation of the overall securities frontier.

Fragmentation exists within each asset class as well. The following sections discuss peculiarities within selected asset classes.

Fragmentation in Equities

U.S. equities can be traded in dark pools and on lit exchanges. Dark pools are exchange-like entities where the order book is "dark"—not displayed to any participant of that pool. According to Pragma Securities (2011), about 22 percent of

Aldridge, Irene. Wiley Trading Ser. : High-Frequency Trading : A Practical Guide to Algorithmic Strategies and Trading Systems (2). Somerset, US: John Wiley & Sons, Incorporated, 2013. ProQuest ebrary. Web. 8 March 2017.
Copyright © 2013. John Wiley & Sons, Incorporated. All rights reserved.

b6f1260bee54f438e54f5e8d966ba6d2
ebrary

the U.S. aggregate equity volume is presently traded in the dark pools. The singular advantage of dark pools lies in their ability to match large orders without revealing information associated with the order size, as the orders are not observable. The frequently cited disadvantages of dark pools include the lack of transparency and related issues. Unlike "lit" exchanges, dark pools do not offer differentiated pricing for limit and market orders—the dark pool limit order book is not disclosed to market participants.

The remaining 78 percent of the U.S. equity volume is executed on "lit" exchanges—venues where the order book is fully transparent and can be disseminated in full to interested market participants. But even within the lit markets category, the landscape is quite fragmented as the exchanges compete to set optimal fees.

In the lit exchanges, the U.S. equities are required to be executed at the NBBO or better quotes, compiled from the best bid and ask quotes available on all the member exchanges and disseminated by the Securities Information Processor (SIP). The aggregated best quotes are then disseminated back to exchanges as NBBO references. The NBBO execution rule was put forth by the Securities and Exchange Commission (SEC) in 2005 under the Regulation NMS, with the explicit purpose of leveling the playing field: under the NBBO rule, every best limit order, whether placed by a large institution or an individual investor, has to be displayed to all market participants. (Prior to the NBBO rule, individual investors were at the mercy of broker-dealers, who often failed to route investors' limit orders to exchanges, even when said limit orders were at the top of the market—better than the best quote available at the time.) Under the NBBO rule, exchanges that cannot execute at the NBBO due to the dearth of liquidity are required to route incoming market orders to other exchanges where the NBBO is available. As a result, traders placing market orders on lit exchanges are guaranteed that their orders are executed at the best possible prices available nationwide. Dark pools are exempt from the NBBO requirement.

b6f1260bee54f438e54f5e8d966ba6d2
ebrary

Under the NBBO rule, exchanges can match trades only when they can execute at the NBBO, that is, when the NBBO-priced limit orders are recorded within their limit order books. Such NBBO limit orders can be achieved using two distinct approaches:

1. The exchange can compete to attract the top-of-the-book liquidity—limit orders priced at NBBO or better.
2. The exchange can compete to attract market orders, while simultaneously serving as a proprietary market maker posting NBBO limit orders.

The two business models of exchanges readily translate into the new fee structures of exchanges. In addition to clearing fees, exchanges now offer divergent pricing for suppliers of liquidity and takers of liquidity. Depending on whether the exchange follows the business model 1 or 2, the exchange may pay liquidity providers for posting limit orders, or pay takers of liquidity for bringing in market orders. Such payments, amounting to negative transaction costs, are known as *rebates*.

b6f1260bee54f438e54f5e8d966ba6d2
ebrary

Aldridge, Irene. Wiley Trading Ser. : High-Frequency Trading : A Practical Guide to Algorithmic Strategies and Trading Systems (2). Somerset, US: John Wiley & Sons, Incorporated, 2013. ProQuest ebrary. Web. 8 March 2017.
Copyright © 2013. John Wiley & Sons, Incorporated. All rights reserved.

b6f1260bee54f438e54f5e8d966ba6d2
ebrary

The two different business models have driven the exchanges into two distinct camps, "normal" and "inverted" exchanges, based on their fee structure. The normal exchanges offer the following fees: normal exchanges charge traders for placing market orders, taking away liquidity, and offer rebates for placing limit orders, bringing in liquidity. The NYSE, for example, charges $0.21 for a 100-share market order, and pays anywhere from $0.13 to $0.20 for a 100-share limit order. The exact value of the NYSE rebate is determined by the aggregate monthly volume of the trader—the higher the volume, the higher the rebate. The NYSE fee structure displayed online on May 14, 2012, is shown in Figure 3.6.

In contrast, the so-called "inverted" exchanges pay traders small rebates to remove liquidity (place market orders), and charge fees to place limit orders. Boston Exchange (Nasdaq OMX BX) is an example of an inverted exchange. There, traders with market orders for fewer than 3.5 million shares per day are paid $0.0005 per share, while traders placing market orders for 3.5 million or more shares per day are paid $0.0014 per share. Traders adding displayed limit orders under 25,000 per day are charged $0.0015 to $0.0018 per share. Yet, large limit order traders—traders placing limit orders for 25,000 shares per day or more—are paid rebates at the rate of $0.0014 per share. The snapshot of distribution of trading costs on the BX as of October 10, 2012, is shown in Figure 3.7.

As a result of the regulatory framework and pricing divergence among equity exchanges in the United States, various exchanges show different rates of availability

b6f1260bee54f438e54f5e8d966ba6d2
ebrary

b6f1260bee54f438e54f5e8d966ba6d2
ebrary

NYSE Arca Rates per 100 Shares

		TAPE A (NYSE-LISTED)			
Tier	**Tier Requirement(s)**	**Rebate for Adding[1]**	**Fee for Removing**	**Routing to NYSE[2]**	**Routing to Other Venues**
Tier 1	NYSE Arca Daily Adding as of % of US CADV in excess of 0.70%	$ (0.30)	$ 0.30	$0.21/$0.23	$ 0.30
Tier 2	NYSE Arca Daily Adding as of % of US CADV in excess of 0.30%	$ (0.29)	$ 0.30	$0.21/$0.23	$ 0.30
Tier 3	NYSE Arca Daily Adding as of % of US CADV in excess of 0.20%	$ (0.25)	$ 0.30	$0.21/$0.23	$ 0.30
Step-Up Tier 1	NYSE Arca Daily Adding as % of US CADV in excess of 0.15% over the	$ (0.295)	$ 0.30	$0.21/$0.23	$ 0.30

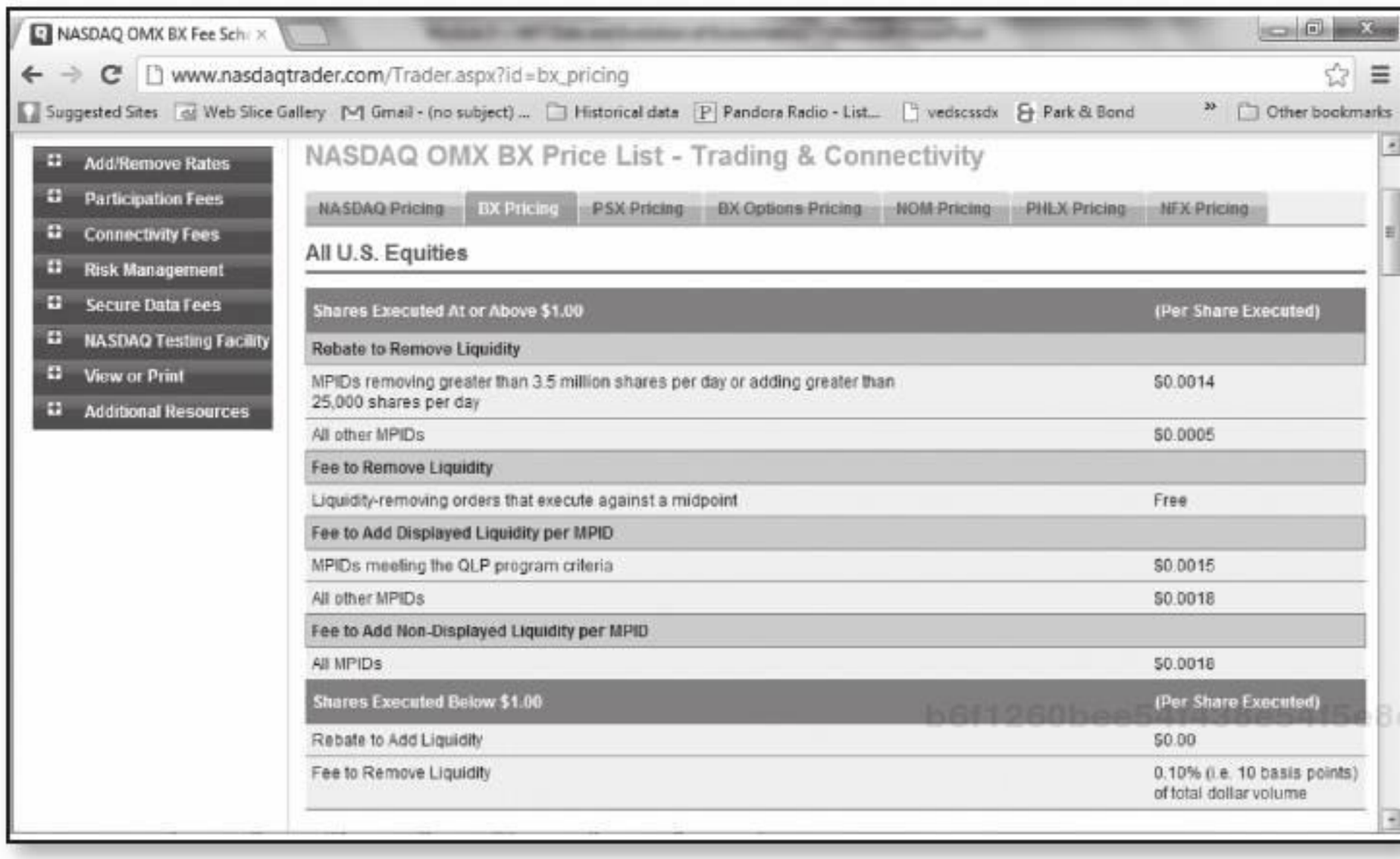
FIGURE 3.6 Fee Structure of Orders Placed and Routed to and from NYSE

Source: NYSE web site

b6f1260bee54f438e54f5e8d966ba6d2
ebrary

Aldridge, Irene. Wiley Trading Ser. : High-Frequency Trading : A Practical Guide to Algorithmic Strategies and Trading Systems (2). Somerset, US: John Wiley & Sons, Incorporated, 2013. ProQuest ebrary. Web. 8 March 2017.
Copyright © 2013. John Wiley & Sons, Incorporated. All rights reserved.

b6f1260bee54f438e54f5e8d966ba6d2
ebrary



Shares Executed At or Above $1.00	(Per Share Executed)
Rebate to Remove Liquidity	
MPIDs removing greater than 3.5 million shares per day or adding greater than 25,000 shares per day	$0.0014
All other MPIDs	$0.0005
Fee to Remove Liquidity	
Liquidity-removing orders that execute against a midpoint	Free
Fee to Add Displayed Liquidity per MPID	
MPIDs meeting the QLP program criteria	$0.0015
All other MPIDs	$0.0018
Fee to Add Non-Displayed Liquidity per MPID	
All MPIDs	$0.0018
Shares Executed Below $1.00	**(Per Share Executed)**
Rebate to Add Liquidity	$0.00
Fee to Remove Liquidity	0.10% (i.e. 10 basis points) of total dollar volume

FIGURE 3.7 Fees on Nasdaq OMX BX, an Inverted Exchange
Source: Nasdaq web site

of NBBO on their order books as well as different market shares. As the intuition would suggest, the exchanges able to provide the highest occurrence of NBBO quotes are the ones able to secure the highest market share. Figure 3.8 displays the relationship between the NBBO availability rates and the market share rates. As shown in Figure 3.8, Nasdaq and NYSE on average have the highest availability of NBBO and obtain the highest share of trades.

b6f1260bee54f438e54f5e8d966ba6d2
ebrary

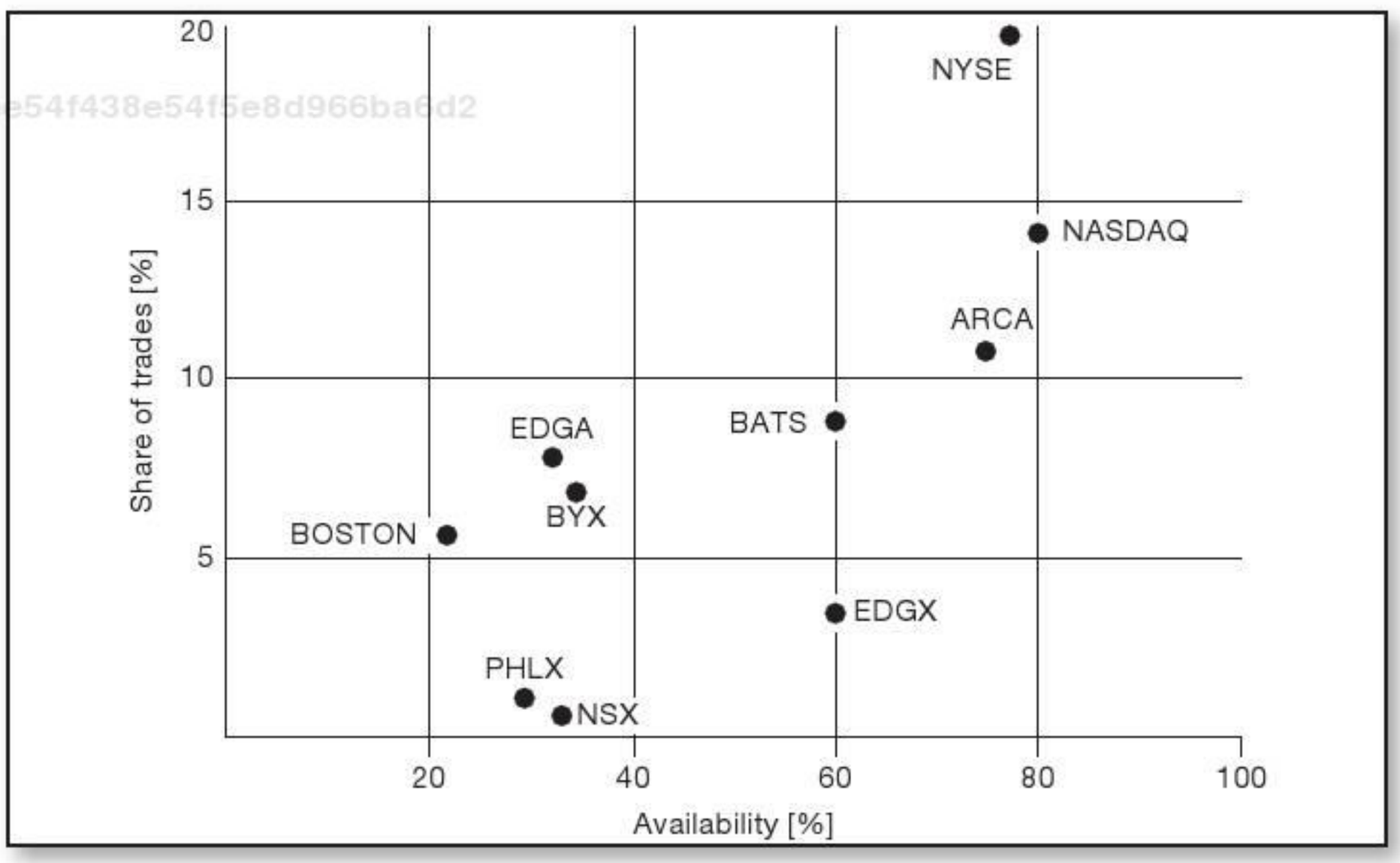


FIGURE 3.8 Availability of NBBO versus Share Volume
Source: Pragma Securities (2011)

b6f1260bee54f438e54f5e8d966ba6d2
ebrary

Aldridge, Irene. Wiley Trading Ser. : High-Frequency Trading : A Practical Guide to Algorithmic Strategies and Trading Systems (2). Somerset, US: John Wiley & Sons, Incorporated, 2013. ProQuest ebrary. Web. 8 March 2017.
Copyright © 2013. John Wiley & Sons, Incorporated. All rights reserved.

b6f1260bee54f438e54f5e8d966ba6d2
ebrary

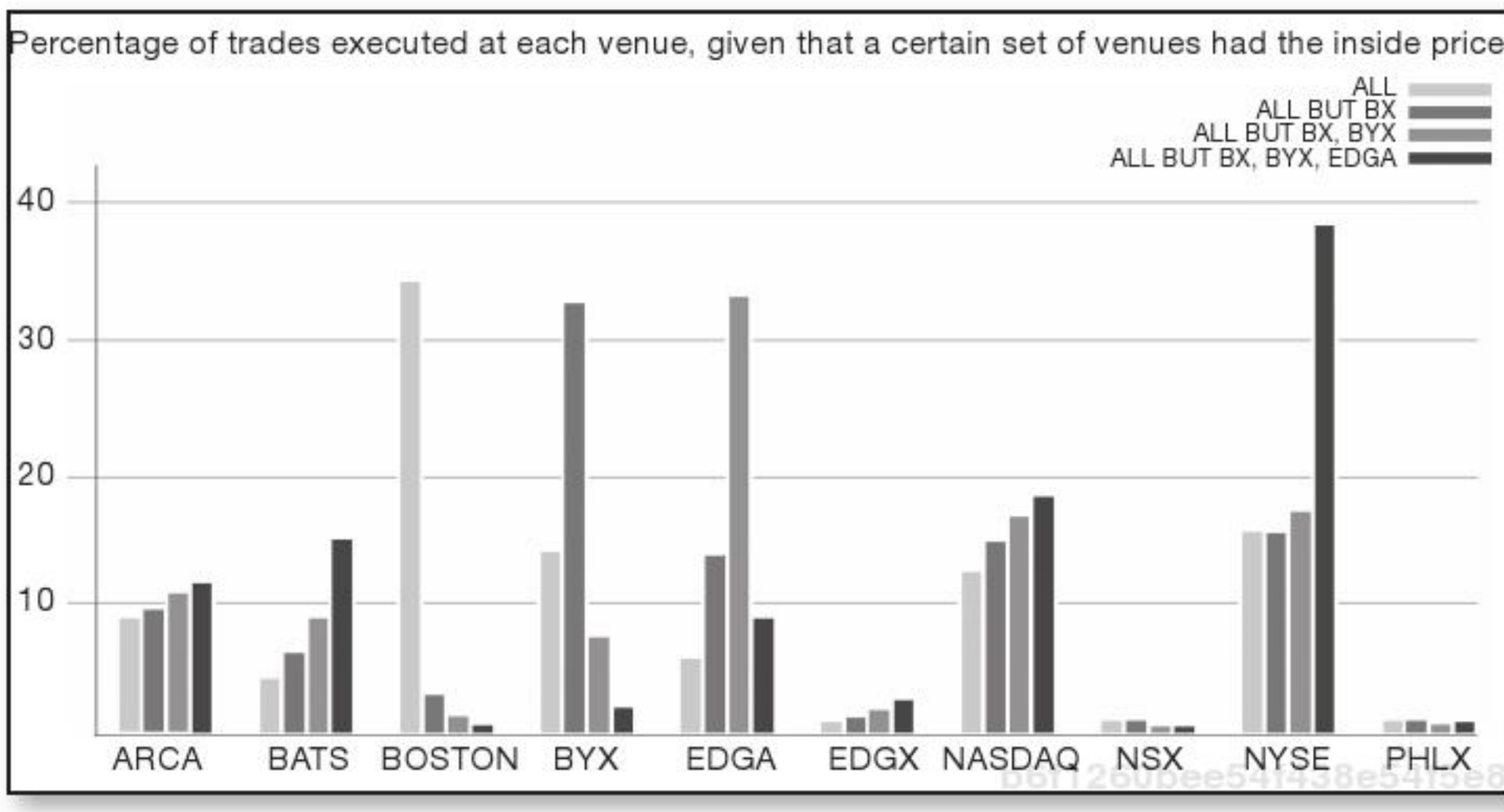


FIGURE 3.9 Percentage of Trades Executed at Each Trading Venue When NBBO Is Present at Certain Venues
Source: Pragma Securities (2011)

Figure 3.9 illustrates what happens when the NBBO is available on all or only some exchanges. When the best bid and the best ask at NBBO are available on all exchanges, over 30 percent of trade volume is executed at the BX, where the inverted fee structure attracts market order traders seeking NBBO liquidity. When every exchange except for BX has the NBBO limit orders, a similarly high number trades is routed to the BATS Y-Exchange (BYX). BYX, like BX, offers inverted fee structure, but with much smaller fees and rebates for removing liquidity. When the NBBO is not available solely on BX and BYX, the 30+ percent market share moves to Direct Edge's EDGA, an exchange with normal but low fee structure. When the NBBO is available everywhere except for BX, BYX, and EDGA, the trading volume moves on largely to the NYSE, although Nasdaq and BATS also benefit from the flow. The economic incentives associated with inverted price structures clearly work: the asymmetric order flow enjoyed by BX creates the first-taker advantage for Nasdaq, the parent company of BX.

High market share of trades does not alone contribute to the profitability of the exchange, as other fees can provide significant contributions to the bottom line of the exchanges. For example, both normal and inverted exchanges charge traders' fees for routing orders to other trading venues when the NBBO is not available on the current exchange. Thus, even an exchange with lowest proportion of NBBO can remain profitable by attracting enough orders and charging fees for routing said orders elsewhere.

Fragmentation in Futures

Futures exchanges operated in a manner similar to equity exchanhes, but without rebate pricing implemented or foreseeable at the time this book was written. In a bid to attract liquidity, and unlike the rebate structure prevalent in equities, futures

Aldridge, Irene. Wiley Trading Ser. : High-Frequency Trading : A Practical Guide to Algorithmic Strategies and Trading Systems (2). Somerset, US: John Wiley & Sons, Incorporated, 2013. ProQuest ebrary. Web. 8 March 2017.
Copyright © 2013. John Wiley & Sons, Incorporated. All rights reserved.

b6f1260bee54f438e54f5e8d966ba6d2
ebrary

exchanges have been deploying pro-rata matching, discussed earlier. Other differences among futures exchanges are predominantly based on their operational and risk management decisions, such as when to draw the circuit breakers and so on, described in detail in Chapter 13.

Fragmentation in Options

Various options exchanges have sprung up over the past decade. A majority of the exchanges, however, operate on similar principles and are driven by the same group of market makers.

Due to the sheer number of options with different expiration dates and strike price combinations, most options are barely traded. The same 10 or so broker-dealers tend to provide liquidity on most equity options markets, with large spreads being common.

b6f1260bee54f438e54f5e8d966ba6d2
ebrary

The relative lack of activity in options trading enables market surveillance for insider information. A large trade in the long-dated option or an option with a strike price far away from the current market price of the underlying usually represents a bet on someone's special knowledge, not yet revealed to the rest of the markets.

Fragmentation in Forex

In Forex, interdealer brokers begin competing with broker-dealers for direct access to end traders. The traditionally interdealer brokers Currenex and iCAP, for example, now accept selected institutional customers. In breaking with other exchanges, iCAP offers a 250-microsecond validity on all of its top-of-the-book foreign exchange (forex) quotes.

b6f1260bee54f438e54f5e8d966ba6d2
ebrary

Fragmentation in Fixed Income

While most fixed income traditionally traded OTC, some signs point to a potential near-term exchange-ization of the fixed-income market. Thus, iCAP has planned to launch a fixed-income matching engine using the Nasdaq OMX technology. The iCAP offering would further target institutional investors by providing a one-second validity of all top-of-the-book quotes.

Fragmentation in Swaps

Swaps have also traditionally traded OTC, with privately negotiated contracts. Under the Dodd-Frank Act, swaps are required to be standardized and traded electronically on exchanges. A new class of trading venues, collectively known as *swaps execution facilities* (SEFs), was jointly shaped by the industry and regulators at the time this book was written.

b6f1260bee54f438e54f5e8d966ba6d2
ebrary

Aldridge, Irene. Wiley Trading Ser. : High-Frequency Trading : A Practical Guide to Algorithmic Strategies and Trading Systems (2). Somerset, US: John Wiley & Sons, Incorporated, 2013. ProQuest ebrary. Web. 8 March 2017.
Copyright © 2013. John Wiley & Sons, Incorporated. All rights reserved.

b6f1260bee54f438e54f5e8d966ba6d2
ebrary

Summary

Modern markets are complex businesses ever concerned with streamlining their operations with the goal of delivering the most immediate, cost-effective service to their customers. The competition of trading venues has led to the innovation and evolution in methods, pricing, and service models of exchanges and alternative trading systems. While technology remains the key driver in development of faster and leaner offerings, the solutions are becoming more customer-centric, producing trading products customized to clients' unique execution needs.

End-of-Chapter Questions

1. What is the difference between the normal and inverted exchanges?
2. What type of orders would you use to buy a large quantity? Why?
3. You are given the task of developing an algorithm optimizing liquidation (selling) of a large equity position. Your aim is to develop an algorithm that maximizes execution costs while minimizing speed of liquidation process. How would you develop such an algorithm giving normal and inverted price structures of modern equity exchanges?
4. You are receiving Level I and Level II data on a certain futures contract from a well-known exchange. At 13:45:00:01:060 GMT, the aggregate liquidity reported at the best bid of 12.7962 comprises 325 contracts. The next tick of data is a trade print with a timestamp of 13:45:00:01:075 GMT, recording a trade of 900 contracts at 12.7962. The following tick is a 13:45:00:01:095 GMT quote citing the best bid of 12.7962 with the size of 325 contracts. What has happened in the market from 13:45:00:01:060 GMT to 13:45:00:01:095 GMT?
5. The current right-hand side of the limit order book for a particular stock on a given exchange shows the following information: Best Ask: 100 shares at 35.67, 200 shares at 35.68, 100 shares at 35.69. What is the average price per share you are likely to receive on your market buy order of 250 shares if the National Best Bid is advertised as 35.67? What if the National Best Bid is 35.65 instead?
6. Is an iceberg order passive or aggressive?
7. The last recorded trade price in a given options market is 2.83. The prevailing best bid is 2.65 and the prevailing best ask is 2.90. Is a limit order to buy 10 contracts at 2.85 passive or aggressive? Why?
8. A quantitative researcher ("quant") develops his investing models using daily closing prices. What order types should an execution trader use to execute the quant's buy-and-sell decisions?

b6f1260bee54f438e54f5e8d966ba6d2
ebrary

Aldridge, Irene. Wiley Trading Ser. : High-Frequency Trading : A Practical Guide to Algorithmic Strategies and Trading Systems (2). Somerset, US: John Wiley & Sons, Incorporated, 2013. ProQuest ebrary. Web. 8 March 2017.
Copyright © 2013. John Wiley & Sons, Incorporated. All rights reserved.

b6f1260bee54f438e54f5e8d966ba6d2
ebrary

CHAPTER 4

High-Frequency Data[1]

b6f1260bee54f438e54f5e8d966ba6d2
ebrary

Trade and quote information is often distributed in Level I or Level II formats. Level I quotes include the best bid price, best ask price, best bid size, best ask size, and last trade price and size, where available. Level II quotes include all changes to the order book, including new limit order arrivals and cancellations at prices away from the market price. This chapter describes the details of data quotations and sampling methodologies and contrasts the data with their low-frequency counterparts.

What Is High-Frequency Data?

High-frequency data, also known as *tick data,* are records of live market activity. Every time a customer, a dealer, or another entity posts a so-called limit order to buy s units of a specific security with ticker X at price q, a bid quote $q^b_{t_b}$ is logged at time t_b to buy $S^b_{t_b}$ units of X. Market orders are incorporated into tick data in a different way, as discussed in this chapter.

b6f1260bee54f438e54f5e8d966ba6d2
ebrary

When the newly arrived bid quote $q^b_{t_b}$ has the highest price relative to all other previously arrived bid quotes in force, $q^b_{t_b}$ becomes known as "the best bid" available at time t_b. Similarly, when a trading entity posts a limit order to sell s units of X at price q, an ask quote $q^a_{t_a}$ is logged at time t_a to sell $S^a_{t_a}$ units of X. If the latest $q^a_{t_a}$ is lower than all other available ask quotes for security X, $q^a_{t_a}$ becomes known as "the best ask" at time t_a.

What happens to quotes from the moment they arrive largely depends on the venue where the orders are posted. Best bids and asks posted directly on an exchange will be broadcast to all exchange participants and other parties tracking quote data. In situations when the new best bid exceeds the best ask already in force on the exchange, $q^b_{t_b} \geq q^a_{t_a}$, most exchanges will immediately "match" such quotes, executing a

[1] A version of this chapter appears in F. Fabozzi, ed., *Encyclopedia of Financial Models* (3 volume set) (Hoboken, NJ: John Wiley & Sons, 2012).

b6f1260bee54f438e54f5e8d966ba6d2
ebrary

Aldridge, Irene. Wiley Trading Ser. : High-Frequency Trading : A Practical Guide to Algorithmic Strategies and Trading Systems (2). Somerset, US: John Wiley & Sons, Incorporated, 2013. ProQuest ebrary. Web. 8 March 2017.
Copyright © 2013. John Wiley & Sons, Incorporated. All rights reserved.

b6f1260bee54f438e54f5e8d966ba6d2
ebrary

trade at the preexisting best ask, $q_{t_a}^{a}$ at time t_b. Conversely, should the newly arrived best ask fall below the current best bid, $q_{t_a}^{a} \leq q_{t_b}^{b}$, the trade is executed at the preexisting best bid, $q_{t_b}^{b}$ at time t_a.

Most dark pools match bids and asks by "crossing the spread," but may not broadcast the newly arrived quotes (hence the mysterious moniker, the "dark pools"). Similarly, quotes destined for the interdealer networks may or may not be disseminated to other market participants, depending on the venue.

Market orders contribute to high-frequency data in the form of "last trade" information. Unlike a limit order that is an order to buy a specified quantity of a security at a certain price, a market order is an order to buy a specified quantity of a security at the best price available at the moment the order is "posted" on the trading venue. As such, market orders are executed immediately at the best available bid or best ask prices, with each market buy order executed at the best ask and each market sell matched with the best bid, and the transaction is recorded in the quote data as the "last trade price" and the "last trade size."

b6f1260bee54f438e54f5e8d966ba6d2
ebrary

A large market order may need to be matched with one or several best quotes, generating several "last trade" data points. For example, if the newly arrived market buy order is smaller in size than that of the best ask, the best ask quote may still remain in force on most trading venues, but the best ask size will be reduced to reflect that the portion of the best ask quote has been matched with the market order. When the size of the incoming market buy order is bigger than the size of the corresponding best ask, the market order consumes the best ask in its entirety, and then proceeds to be matched sequentially with the next available best ask until the size of the market order is fulfilled. The remaining lowest-priced ask quote becomes the best ask available on the trading venue.

Most limit and market orders are placed in so-called "lot sizes": increments of certain number of units, known as a lot. In foreign exchange, a standard trading lot today is US$5 million, a considerable reduction from a minimum of $25 million entertained by high-profile brokers just a few years ago. On equity exchanges, a lot can be as low as one share, but dark pools may still enforce a 100 share minimum requirement for orders. An order for the amount other than an integer increment of a lot size, is called "an odd lot."

b6f1260bee54f438e54f5e8d966ba6d2
ebrary

Small limit and market "odd lot" orders posted through a broker-dealer may be aggregated, or "packaged," by the broker-dealer into larger-size orders in order to obtain volume discounts at the orders' execution venue. In the process, the brokers may "sit" on quotes without transmitting them to an executing venue, delaying execution of customers' orders.

How Is High-Frequency Data Recorded?

The highest-frequency data are a collection of sequential "ticks," arrivals of the latest quote, trade, price, order size, and volume information. Tick data usually has the following properties:

- A timestamp
- A financial security identification code

b6f1260bee54f438e54f5e8d966ba6d2
ebrary

Aldridge, Irene. Wiley Trading Ser. : High-Frequency Trading : A Practical Guide to Algorithmic Strategies and Trading Systems (2). Somerset, US: John Wiley & Sons, Incorporated, 2013. ProQuest ebrary. Web. 8 March 2017.
Copyright © 2013. John Wiley & Sons, Incorporated. All rights reserved.

b6f1260bee54f438e54f5e8d966ba6d2
ebrary

- An indicator of what information it carries:
  - Bid price
  - Ask price
  - Available bid size
  - Available ask size
  - Last trade price
  - Last trade size
- Security-specific data, such as implied volatility for options
- The market value information, such as the actual numerical value of the price, available volume, or size

b6f1260bee54f438e54f5e8d966ba6d2
ebrary

A timestamp records the date and time at which the quote originated. It may be the time at which the exchange or the broker-dealer released the quote, or the time when the trading system has received the quote. At the time this article is written, the standard "round-trip" travel time of an order quote from the ordering customer to the exchange and back to the customer with the acknowledgment of order receipt is 15 milliseconds or less in New York. Brokers have been known to be fired by their customers if they are unable to process orders at this now standard speed. Sophisticated quotation systems, therefore, include milliseconds and even microseconds as part of their timestamps.

Another part of the quote is an identifier of the financial security. In equities, the identification code can be a ticker, or, for tickers simultaneously traded on multiple exchanges, a ticker followed by the exchange symbol. For futures, the identification code can consist of the underlying security, futures expiration date, and exchange code.

The last trade price shows the price at which the last trade in the security cleared. Last trade price can differ from the bid and ask. The differences can arise when a customer posts a favorable limit order that is immediately matched by the broker without broadcasting the customer's quote. Last trade size shows the actual size of the last executed trade.

b6f1260bee54f438e54f5e8d966ba6d2
ebrary

The best bid is the highest price available for sale of the security in the market. The best ask is the lowest price entered to buy the security at any particular time. In addition to the best bid and best ask, quotation systems may disseminate "market depth" information: the bid and ask quotes entered posted on the trading venue at prices worse than the best bid and ask, as well as aggregate order sizes corresponding to each bid and ask recorded on the trading venue's "books." Market depth information is sometimes referred to as the Level II data and may be disseminated as the premium subscription service only. In contrast, the best bid, best ask, last trade price, and size information ("Level I data") is often available for a small nominal fee.

Panels A and B of Figure 4.1 illustrate a 30-second log of Level I high-frequency data recorded by New York Stock Exchange (NYSE) Arca for Standard & Poor's Depositary Receipts (SPDR) S&P 500 exchange-traded fund (ETF; ticker SPY) from 14:00:16:400 to 14:02:00:000 GMT on November 9, 2009. Panel A shows quote

b6f1260bee54f438e54f5e8d966ba6d2
ebrary

Aldridge, Irene. Wiley Trading Ser. : High-Frequency Trading : A Practical Guide to Algorithmic Strategies and Trading Systems (2). Somerset, US: John Wiley & Sons, Incorporated, 2013. ProQuest ebrary. Web. 8 March 2017.
Copyright © 2013. John Wiley & Sons, Incorporated. All rights reserved.

b6f1260bee54f438e54f5e8d966ba6d2
ebrary

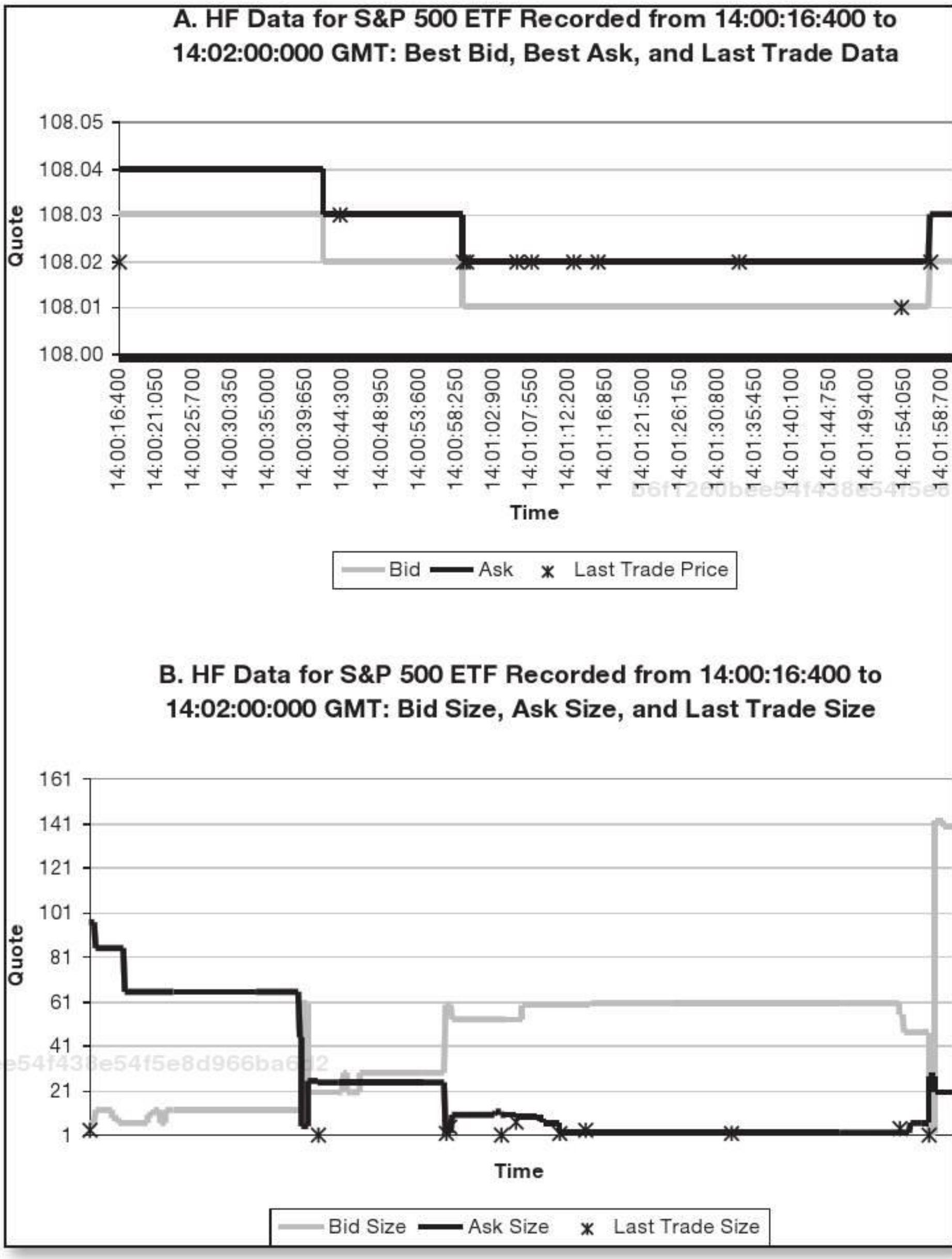


FIGURE 4.1 Level I High-Frequency Data Recorded by NYSE Arca for SPY from 14:00:16:400 to 14:02:00:000 GMT on November 9, 2009. Data source: Bloomberg

data: best bid, best ask and last trade information, while panel B displays corresponding position sizes (best bid size, best ask size, and last trade size).

Properties of High-Frequency Data

High-frequency securities data have been studied for many years. Yet they are still something of a novelty to many academics and practitioners. Unlike daily or monthly data sets commonly used in much of financial research and related applications, high-frequency

Aldridge, Irene. Wiley Trading Ser. : High-Frequency Trading : A Practical Guide to Algorithmic Strategies and Trading Systems (2). Somerset, US: John Wiley & Sons, Incorporated, 2013. ProQuest ebrary. Web. 8 March 2017.
Copyright © 2013. John Wiley & Sons, Incorporated. All rights reserved.

b6f1260bee54f438e54f5e8d966ba6d2
ebrary

data have distinct properties, which simultaneously can be advantageous and intimidating to researchers. Table 4.1 summarizes the properties of high-frequency data. Each property and its advantages and disadvantages are discussed in detail later in the article.

TABLE 4.1 Summary of Properties of High-Frequency Data

Property of HF Data	Description	Pros	Cons
Voluminous	Each day of high-frequency data contains the number of observations equivalent to 30 years of daily data.	Large numbers of observations carry lots of information.	High-frequency data are difficult to handle manually.
Subject to bid-ask bounce	Unlike traditional data based on just closing prices, tick data carry additional supply-and-demand information in the form of bid and ask prices and offering sizes.	Bid and ask quotes can carry valuable information about impending market moves, which can be harnessed to the researcher's advantage.	Bid and ask quotes are separated by a spread. Continuous movement from bid to ask and back introduces a jump process, difficult to deal with through many conventional models.
Not normally or lognormally distributed	Returns computed from tick are not normal or lognormal.	Many tradable models are still to be discovered.	Traditional asset pricing models assuming lognormality of prices do not apply.
Irregularly spaced in time	Arrivals of tick data are asynchronous.	Durations between data arrival carry information.	Most traditional models require regularly spaced data; need to convert high-frequency data to some regular intervals, or "bars" of data. Converted data are often sparse (populated with zero returns), once again making traditional econometric inferences difficult.
Do not include buy or sell trade direction information	Level I and Level II data do not include information on whether the trade was a result of a market buy or a market sell order	Data are leaner without trade direction information; trade information is more difficult for bystanders to extract.	The information on whether a trade is buyer initiated or seller initiated is a desired input in many models.

b6f1260bee54f438e54f5e8d966ba6d2
ebrary

High-Frequency Data Are Voluminous

The nearly two-minute sample of tick data for SPY shown in Figure 4.1 contained over 2,000 observations of Level I data: best bid quotes and sizes, best ask quotes and sizes, and last trade prices and sizes. Table 4.2 summarizes the breakdown of the data points provided by NYSE Arca for SPY from 14:00:16:400 to 14:02:00:000 GMT on November 9, 2009, and SPY, Japanese yen futures, and a euro call option throughout the day on November 9, 2009. Other Level I data omitted from Table 4.2 include cumulative daily trade volume for SPY and Japanese yen futures, and "greeks" for the euro call option. The number of quotes observed on November 9, 2009, for SPY alone would comprise over 160 years of daily open, high, low, close, and volume data points, assuming an average of 252 trading days per year.

b6f1260bee54f438e54f5e8d966ba6d2
ebrary

Aldridge, Irene. Wiley Trading Ser. : High-Frequency Trading : A Practical Guide to Algorithmic Strategies and Trading Systems (2). Somerset, US: John Wiley & Sons, Incorporated, 2013. ProQuest ebrary. Web. 8 March 2017.
Copyright © 2013. John Wiley & Sons, Incorporated. All rights reserved.

b6f1260bee54f438e54f5e8d966ba6d2
ebrary

TABLE 4.2 Summary Statistics for Level I Quotes for Selected Securities on November 9, 2009

Quote Type	SPY, 14:00:16:400 to 14:02:00:000 GMT	SPY, All Day	USD/JPY Dec. 2009 Futures, All Day	EUR/USD Call Expiring Dec. 2009 with Strike Price of 1.5100, All Day
Best bid quote	4 (3%)	5,467 (3%)	6,320 (5%)	1,521 (3%)
Best bid size	36 (29%)	38,948 (19%)	39,070 (32%)	5,722 (11%)
Best ask quote	4 (3%)	4,998 (2%)	6,344 (5%)	1,515 (3%)
Best ask size	35 (28%)	38,721 (19%)	38,855 (32%)	5,615 (11%)
Last trade price	6 (5%)	9,803 (5%)	3,353 (3%)	14 (0%)
Last trade size	20 (16%)	27,750 (14%)	10,178 (8%)	25 (0%)
Total	125	203,792	123,216	49,982

The quality of data does not always match its quantity. Centralized exchanges generally provide accurate data on bids, asks, and size. U.S. equity exchanges are required by law to archive and maintain reliable records of every tick of data, as well as to submit best quotes within one minute of their occurrence to the U.S. centralized ticker tape known as the *Securities Information Processor* (SIP). The information on the limit order book beyond the best bid and best offer is known as Level II data and can be obtained via special subscription.

In decentralized markets, such as foreign exchange and the interbank money market, no market-wide quotes are available at any given time. In such markets, participants are aware of the current price levels, but each institution quotes its own prices adjusted for its order book. In decentralized markets, each dealer provides his own tick data to his clients. As a result, a specific quote on a given financial instrument at any given time may vary from dealer to dealer. Reuters, Telerate, and Knight Ridder, among others, collect quotes from different dealers and disseminate them back, improving the efficiency of the decentralized markets.

There are generally thought to be three anomalies in interdealer quote discrepancies. First, each dealer's quotes reflect that dealer's own inventory. For example, a dealer that has just sold a customer $100 million of USD/CAD would be eager to diversify the risk of his position and avoid selling any more of USD/CAD. Most dealers, however, are obligated to transact with their clients on tradable quotes. To incite his clients to place sell orders on USD/CAD, the dealer temporarily raises the bid quote on USD/CAD. At the same time, to encourage his clients to withhold placing buy orders, the dealer raises the ask quote on USD/CAD. Thus, dealers tend to raise both bid and ask prices whenever they are short in a particular financial instrument and lower both bid and ask prices whenever they are disproportionately long in a financial instrument.

Second, in an anonymous marketplace, such as a dark pool, dealers as well as other market makers may "fish" for market information by sending indicative quotes that are much off the previously quoted price to assess the available demand or supply.

Third, Dacorogna et al. (2001) note that some dealers' quotes may lag real market prices. The lag is thought to vary from milliseconds to a minute. Some dealers quote moving averages of quotes of other dealers. The dealers who provide delayed quotes usually do so to advertise their market presence in the data feed. This was particularly true when most order prices were negotiated over the telephone, allowing a considerable delay between quotes and orders. Fast-paced electronic markets discourage lagged quotes, improving the quality of markets.

Aldridge, Irene. Wiley Trading Ser. : High-Frequency Trading : A Practical Guide to Algorithmic Strategies and Trading Systems (2). Somerset, US: John Wiley & Sons, Incorporated, 2013. ProQuest ebrary. Web. 8 March 2017.
Copyright © 2013. John Wiley & Sons, Incorporated. All rights reserved.

b6f1260bee54f438e54f5e8d966ba6d2
ebrary

High-Frequency Data Are Subject to the Bid-Ask Bounce

In addition to trade price and volume data long available in low-frequency formats, high-frequency data comprise bid and ask quotes and their associated order sizes. Bid and ask data arrive asynchronously and introduce noise in the quote process.

The difference between the bid quote and the ask quote at any given time is known as the bid-ask spread. The bid-ask spread is the cost of instantaneously buying and selling the security. The higher the bid-ask spread, the higher the gain the security must produce in order to cover the spread along with other transaction costs. Most low-frequency price changes are large enough to make the bid-ask spread negligible in comparison. In tick data, however, incremental price changes can be comparable or smaller than the bid-ask spread.

Bid-ask spreads usually vary throughout the day. Figure 4.2 illustrates the average bid-ask spread cycles observed in the institutional EUR/USD market for the last two weeks of October 2008. As Figure 4.2 shows, the average spread increases significantly during Tokyo trading hours, when the market is quiet. The spread then reaches its lowest levels during the overlap of the London and New York trading sessions, when the market has many active buyers and sellers. The spike in the spread over the weekend of October 18–19, 2008, reflects the market concern over the subpoenas issued on October 17, 2009, to senior Lehman executives in a case relating to potential securities fraud at Lehman Brothers.

Bid-ask spreads typically increase during periods of market uncertainty or instability. Figure 4.3, for example, compares average bid-ask spreads on EUR/USD in the stable market conditions of July–August 2008 and the crisis conditions of September–October 2008. As the figure shows, the intraday spread pattern is persistent in both crisis and normal market conditions, but the spreads are significantly higher during crisis months than during normal conditions at all hours of the day. As Figure 4.3 also shows, the spread increase is not uniform at all hours of the day. The average hourly EUR/USD spreads increased by 0.0048 percent (0.48 basis points or pips) between the hours of 12 GMT and

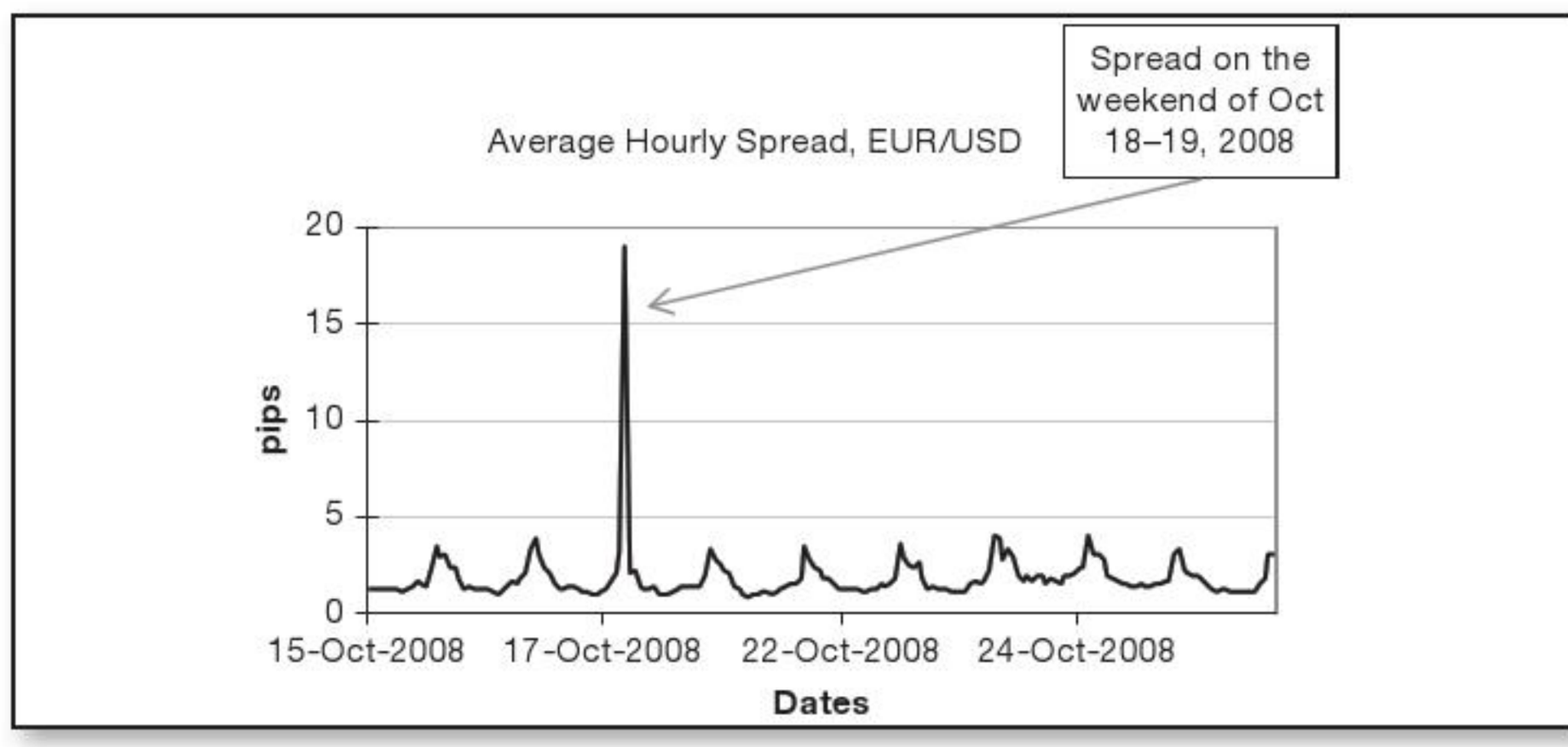


FIGURE 4.2 Average Hourly Bid-Ask Spread on EUR/USD Spot for the Last Two Weeks of October 2008 on a Median Transaction Size of US$5 Million

b6f1260bee54f438e54f5e8d966ba6d2
ebrary

Aldridge, Irene. Wiley Trading Ser. : High-Frequency Trading : A Practical Guide to Algorithmic Strategies and Trading Systems (2). Somerset, US: John Wiley & Sons, Incorporated, 2013. ProQuest ebrary. Web. 8 March 2017.
Copyright © 2013. John Wiley & Sons, Incorporated. All rights reserved.

b6f1260bee54f438e54f5e8d966ba6d2
ebrary

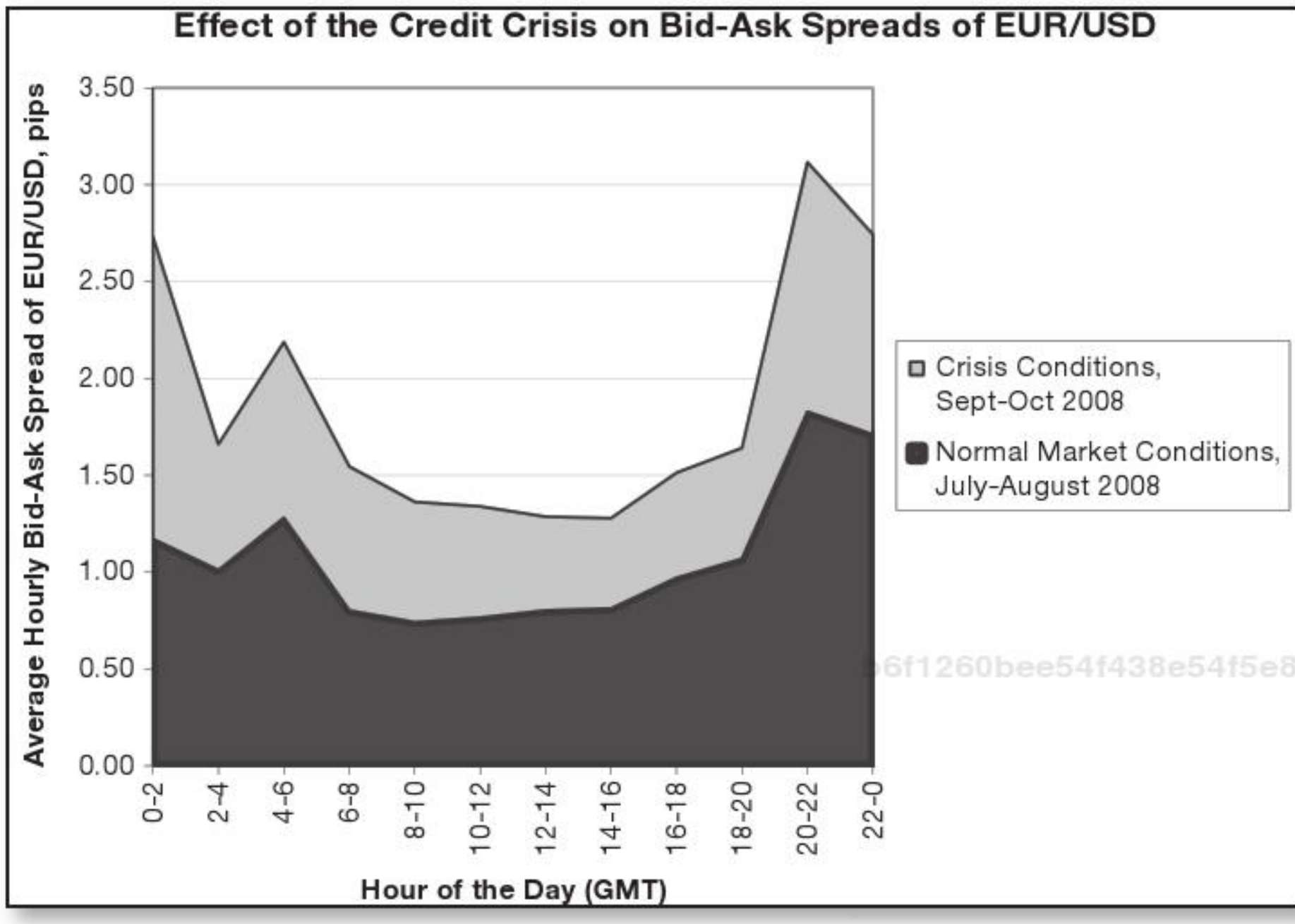


FIGURE 4.3 Comparison of Average Bid-Ask Spreads for Different Hours of the Day during Normal Market Conditions and Crisis Conditions

16 GMT, when the London and New York trading sessions overlap. From 0 to 2 GMT, during the Tokyo trading hours, the spread increased by 0.0156 percent, over three times the average increase during the New York/London hours.

As a result of increasing bid-ask spreads during periods of uncertainty and crises, the profitability of high-frequency strategies decreases during those times. For example, high-frequency EUR/USD strategies running over Asian hours incurred significantly higher costs during September and October 2008 as compared with normal market conditions. A strategy that executed 100 trades during Asian hours alone resulted in 1.56 percent evaporating from daily profits due to the increased spreads, while the same strategy running during London and New York hours resulted in a smaller but still significant daily profit decrease of 0.48 percent. The situation can be even more severe for high-frequency strategies built for less liquid instruments. For example, bid-ask spreads for NZD/USD (not shown) on average increased three times during September–October in comparison with market conditions of July–August 2008.

While tick data carry information about market dynamics, it is also distorted by the same processes that make the data so valuable in the first place. Dacorogna et al. (2001) report that sequential trade price bounces between the bid and ask quotes during market execution of orders introduce significant distortions into estimation of high-frequency parameters. Corsi, Zumbach, Muller, and Dacorogna (2001), for example, show that the bid-ask bounce introduces a considerable bias into volatility estimates. The authors calculate that the bid-ask bounce on average results in –40 percent first-order autocorrelation of tick data. Corsi et al. (2001) as well as Voev and Lunde (2007) propose to remedy the bias by filtering the data from the bid-ask noise prior to estimation.

Aldridge, Irene. Wiley Trading Ser. : High-Frequency Trading : A Practical Guide to Algorithmic Strategies and Trading Systems (2). Somerset, US: John Wiley & Sons, Incorporated, 2013. ProQuest ebrary. Web. 8 March 2017.
Copyright © 2013. John Wiley & Sons, Incorporated. All rights reserved.

b6f1260bee54f438e54f5e8d966ba6d2
ebrary

In addition to real-time adjustments to bid-ask data, researchers deploy forecasting techniques to estimate the impending bid-ask spread and adjust for it in models ahead of time. Future realizations of the bid-ask spread can be estimated using the model suggested by Roll (1984), where the price of an asset at time t, p_t, is assumed to equal an unobservable fundamental value, m_t, offset by a value equal to half of the bid-ask spread, s. The price offset is positive when the next market order is a buy, and negative when the trade is a sell, as shown in equation (3):

$$p_t = m_t + \frac{s}{2} I_t \tag{3}$$

$$\text{where } I_t = \begin{cases} 1, & \textit{market buy at ask} \\ -1, & \textit{market sell at bid} \end{cases}$$

If either a buy or a sell order can arrive next with equal probability, then $E[I_t] = 0$, and $E[\Delta p_t] = 0$, absent changes in the fundamental asset value, m_t. The covariance of subsequent price changes, however, is different from 0:

b6f1260bee54f438e54f5e8d966ba6d2
ebrary

$$\text{cov}\left[\Delta p_t, \Delta p_{t+1}\right] = E\left[\Delta p_t \Delta p_{t+1}\right] = -\frac{s^2}{4} \tag{4}$$

As a result, the future expected spread can be estimated as follows:

$$E[s] = 2\sqrt{-\text{cov}\left[\Delta p_t, \Delta p_{t+1}\right]} \text{ whenever } \text{cov}\left[\Delta p_t, \Delta p_{t+1}\right] < 0$$

Numerous extensions of Roll's model have been developed to account for contemporary market conditions along with numerous other variables. Hasbrouck (2007) provides a good summary of the models.

To use standard econometric techniques in the presence of the bid-ask bounce, many practitioners convert the tick data to "midquote" format: the simple average of the latest bid and ask quotes. The midquote is used to approximate the price level at which the market is theoretically willing to trade if buyers and sellers agreed to meet each other halfway on the price spectrum. Mathematically, the midquote can be expressed as follows:

b6f1260bee54f438e54f5e8d966ba6d2
ebrary

$$\hat{q}^m_{t_m} = \frac{1}{2}\left(q^a_{t_a} + q^b_{t_b}\right) \text{ where } t_m = \begin{cases} t_a, \textit{if } t_a \geq t_b \\ t_b, \textit{otherwise} \end{cases} \tag{5}$$

The latter condition for t_m reflects the continuous updating of the mid-quote estimate: $\hat{q}^m_{t_m}$ is updated whenever the latest best bid, $q^b_{t_b}$, or best ask quote, $q^a_{t_a}$, arrives, at t_b or t_a respectively.

Another way to sample tick quotes into a cohesive data series is by weighing the latest best bid and best ask quotes by their accompanying order sizes:

$$\tilde{q}^s_t = \frac{q^b_{t_b} s^a_{t_a} + q^a_{t_a} s^b_{t_b}}{s^a_{t_a} + s^b_{t_b}} \tag{6}$$

where $q^b_{t_b}$ and $S^b_{t_b}$ is the best bid quote and the best bid available size recorded at time t_b (when $q^b_{t_b}$ became the best bid), and $q^a_{t_a}$ and $S^a_{t_a}$ is the best bid quote and the best bid available size recorded at time t_a.

b6f1260bee54f438e54f5e8d966ba6d2
ebrary

Aldridge, Irene. Wiley Trading Ser. : High-Frequency Trading : A Practical Guide to Algorithmic Strategies and Trading Systems (2). Somerset, US: John Wiley & Sons, Incorporated, 2013. ProQuest ebrary. Web. 8 March 2017.
Copyright © 2013. John Wiley & Sons, Incorporated. All rights reserved.

b6f1260bee54f438e54f5e8d966ba6d2
ebrary

High-Frequency Data Are Not Normal or Lognormal

Many classical models assume lognormality of prices, allowing price diffusion models to work seamlessly, and resulting in several pricing models, such as Black-Scholes, to be considered fair approximations of market evolutions of related financial instruments. A necessary condition for lognormality of prices is the normal distribution of sequential price changes. As this section shows, however, sequential changes in most of the tick data, like midquotes and size-weighted quotes and trades, do not follow normal distribution, yet distribution of sequential trade ticks is close to that of normal. Trade tick data are, therefore, the best choice for modelers assuming lognormal prices.

Figure 4.4 compares the histograms of simple returns computed from midquote (panel A), size-weighted midquote (panel B) and trade-price (panel C) processes for SPDR S&P 500 ETF data recorded as they arrive throughout November 9, 2009. The data neglect the time difference between the adjacent quotes, treating each sequential quote as an independent observation. Figure 4.5 contrasts the quantile distribution plots of the same data sets with the quantiles of a standard normal distribution.

As Figures 4.4 and 4.5 show, the basic midquote distribution is constrained by the minimum "step size": the minimum changes in the midquote can occur at half-tick increments (at present, the minimum tick size is $0.01 in equities). The size-weighted midquote forms the most continuous distribution among the three distributions discussed. Figure 4.5 confirms this notion further and also illustrates the fat tails present in all three types of data distributions.

As clearly shown in Figure 4.5, of the three methodologies, tick-by-tick trade returns most closely fit the normal distribution, when heavy tails past four standard deviations are ignored. Midquote values and size-weighted midquotes alike begin to deviate from normality at just two standard deviations, while trade returns follow the normal up to four standard deviations.

b6f1260bee54f438e54f5e8d966ba6d2
ebrary

High-Frequency Data Are Irregularly Spaced in Time

Most modern computational techniques have been developed to work with regularly spaced data, presented in monthly, weekly, daily, hourly, or other consistent intervals. The traditional reliance of researchers on fixed time intervals is due to:

- Relative availability of daily data (newspapers have published daily quotes since the 1920s).
- Relative ease of processing regularly spaced data.
- An outdated view that "whatever drove security prices and returns, it probably did not vary significantly over short time intervals." (Goodhart and O'Hara 1997, pp. 80–81)

By contrast, high-frequency observations are separated by varying time intervals. One way to overcome the irregularities in the data are to sample it at certain

Aldridge, Irene. Wiley Trading Ser. : High-Frequency Trading : A Practical Guide to Algorithmic Strategies and Trading Systems (2). Somerset, US: John Wiley & Sons, Incorporated, 2013. ProQuest ebrary. Web. 8 March 2017.
Copyright © 2013. John Wiley & Sons, Incorporated. All rights reserved.