



PDE PROVIDER

Professional Development for the Expert

Course Title

Liquidity, Trading, and Price
Determination in Equity Markets:
A Finance Course Application

Instructor

Alan Fata, DBA

Credit

3 PDU

Questions

15

Professional Development for the Expert, Inc.
(571) 619-5550, info@pdeprovider.com

Adaptation Statement

- *This course is chapter 2 titled “Liquidity, Trading, and Price Determination in Equity Markets: A Finance Course Application” adapted from the book titled “Liquidity, Markets and Trading in Action, An Interdisciplinary Perspective”, which can be downloaded for free from the following links: <https://open.umn.edu/opentextbooks/textbooks/liquidity-markets-and-trading-in-action>*
- *The book “Liquidity, Markets and Trading in Action, An Interdisciplinary Perspective” by DenizOzenbas, MichaelS.Pagano, RobertA.Schwartz, and Bruce W. Weber is used under a Creative Commons Attribution 4.0 International License, except where otherwise noted.*



- *Check additional references and sources in the original document.*
- *This adaptation has reformatted the original text, and have replaced some images and figures to make the resulting whole more shareable. This adaptation has not significantly altered or updated the original text.*
- *Few modifications have been made for the purpose of presenting this course on this website.*

Liquidity, Trading, and Price Determination in Equity Markets: A Finance Course Application

2

The relationship between fundamental information and the price of equity shares is critically important. Fundamental information encompasses a vast array of items that pertain to individual firms, to industries, and to the broad, macro economy. In investment courses, the relationship is considered with respect to portfolio formation. In corporate finance, the relationship is considered with respect to asset valuations and the determination of a firm's cost of capital.

The transformation of fundamental information into share prices starts with the information set and extends to investors (both individual and institutional) and then to the marketplace where equity shares are traded and share prices determined. In so doing, fundamental information is transformed into three factors: (1) expected future returns, (2) uncertainty concerning future returns (an investment's risk), and (3) the difficulty of buying and selling shares in the market (liquidity risk).

In broad brush, this is how it works. Assume that a stock's expected 1-year forward price is \$55 a share. If shares are currently priced at \$50, the expected return on the investment is 10%. If, concurrently, the risk-free rate of interest is 4%, the stock is priced to yield a 6% premium. What accounts for the premium? Two things: risk and illiquidity.

Risk exists because what a stock's actual price will be one year from now is unknown in the present. The stock's *expected* share price is \$55. One year later, the price could turn out to be nicely higher than \$55 or disappointingly lower. Thus, the investment is risky, and very importantly, investors are risk averse. Accordingly, the premium compensates them for accepting risk. But is that all it compensates investors for? No, investors are also averse to illiquidity.

Risk pertains to a future share value, while illiquidity matters when shares are bought or sold. Here is a simple, intuitive definition of what the term liquidity means: the ability to buy or to sell shares reasonably quickly, in reasonable amounts, and at reasonable prices. In a frictionless environment, the market would be perfectly liquid, trading would be costless, and shares could be bought or sold

instantly at an appropriate price. But equity markets are not frictionless,¹ trading is not at all costless (we explain more about this in Sect. 2.2), and transaction costs are higher the more illiquid a market is.

Let us back up for a moment. How are investors compensated for risk? By a risk premium. How are they compensated for buying shares that they know can be difficult to sell in the future? By an illiquidity premium. Accordingly, let us repeat: with the risk-free rate at 4%, buying shares at \$50 while expecting a 1-year forward price of \$55 yields a premium of 6%, and this premium compensates investors both for accepting risk and for bearing the cost of illiquidity.

So what is liquidity? As we have just said, a good intuitive definition of this slippery term is the ease with which shares can be traded. Can they be traded quickly? Can they be traded in reasonable quantity and at a reasonable price? If the answer is yes, yes, yes, then we can say that the market for a company's shares is liquid. But what benchmark might there be for assessing, for an order of a given number of shares, the time taken to fill it and the price at which the trade has been made? And can the assessments of time, price, and size be aggregated into a single quantitative measure of liquidity? They cannot, so where do we stand? Hang on, we return to a further discussion of liquidity in Sect. 2.7 of this chapter.

For most stocks, speed is not an issue in today's modern electronic markets. What about size? Size is not an issue for smaller, retail-sized orders, but it is a major challenge for institutional-sized orders (for instance, an order of 50,000 shares, 100,000 shares, or more). What about price? Have you observed how rapidly prices change in short, intraday intervals? They bounce around, often with such rapidity that you can look at a price one instant, blink, and then look again and the stock's share value has changed. Clearly, in this environment, trading at a "reasonable" price is difficult to accomplish, and it is not even easy to know what a reasonable price is.

One glance at a computer screen with "real-time" prices will convince you of this. At times, price rises (or falls) over a series of trades, turns direction, and then shoots back down (or up). What might explain this volatility? Finding prices that best reflect the broad market's desire to hold shares is complex and dynamic. The process is called *price discovery*. We pursue this thought further in Sect. 2.4 of this chapter.

Buying and selling shares is clearly not costless. Costs exist in the form of commissions and fees. They also exist for a participant who wants to consummate a trade quickly by buying or selling "at market" because there is a spread between the price at which one can buy shares (the lowest posted asking quote on the market) and the price at which one can sell shares (the highest posted bid quote on the market). The difference between the best buy and the best sell quotes is the *bid-ask*

¹Friction is the total implicit and explicit costs associated with the execution of a financial transaction.

spread. The spread can be thought of as the cost of buying or selling with immediacy. Then, there is the cost big traders incur when their large buy orders push price up, or their large sell orders push price down. This is referred to as a *market impact cost*. There is also an opportunity cost. We identify and discuss these costs in Sect. 2.2.

And so buying and selling shares to implement a portfolio decision is not costless. The costs referred to in the previous paragraph exist because equity markets are not frictionless. The distinction between frictionless and non-frictionless markets is of tremendous importance. The notion of “frictionless” is in the same spirit as the concept of frictionless physics.² This contrasts markedly from a non-frictionless equity market where: 1) commissions are not zero, 2) there are fees and taxes, 3) trading with immediacy requires paying the bid-ask spread, and 4) buying a large number of shares can push price up, and selling a large number can push price down.

One reality is a root cause of much of the complexity that surrounds trading, liquidity creation, and price determination: investors commonly differ in their interpretations of the fundamental information that applies to specific stocks, industries, and the broad economy, and their differing interpretations translate into their having different expectations about what a stock’s future price will be. We refer to this as *divergent expectations*. What are divergent expectations attributable to? Answer: information sets are of enormous size. Moreover, they are incomplete, replete with complexities, ambiguities, and inaccuracies (surprise, surprise). What is the effect of expectations being divergent? It accounts for:

- Discovering reasonable prices in a marketplace being a difficult, complex process
- Good trading being a challenging activity
- Prices being excessively volatile in brief intervals of time
- The design of equity market structure being of critical importance

The difference between homogeneous expectations and divergent expectation merits more attention. Here is how it works. As we have noted, market participants price their orders with regard to the future values that they expect their investments to deliver. Would a community of investors have identical expectations of future values (the means and variances of returns), or might their expectations differ? If their expectations are identical, we say that they have *homogeneous expectations*. If they differ, we say that investors have *divergent expectations*. The distinction is of major importance. Here is one reason why. If investors’ expectations are homogeneous, shares can be thought of as having fundamental (“intrinsic”) values that can be found by stock analysts. If investors’ expectations are divergent, shares do not have fundamental values, and share prices must be found in the marketplace where trades are made. And price discovery is a major function of a stock exchange. What are

² As discussed in Chapter 1, some topics may be discussed initially assuming frictionless environments, and then friction is added to measure the impact.

your thoughts about this? What do you believe best describes investor expectations: homogeneous or divergent?

In this introductory section of the chapter, we have touched on an array of thoughts regarding information, risk, liquidity, trading costs, market structure, and their effects on price determination. Aside from business school students (and other interested people), who should understand these concepts? Of course, the list includes portfolio managers, traders, and the exchanges themselves, but these are not the only stakeholders. As we have noted in the Preface, also important are corporate CEOs and other corporate officials of publicly traded companies, information technology professionals who build and maintain trading systems, regulators who oversee the markets, business journalists, and the investing public.

2.1 Order Types

Recognizing that investors communicate with the market via the orders that they submit, we next turn our attention to two basic, plain vanilla types: market orders and limit orders.

A market order is an order to trade at the best available price, while a limit order specifies a price limit. For example, “buy 100 shares with a limit price of \$50” means do not buy at any price greater than \$50. Equivalently, “sell 100 shares at a limit price of \$51” means do not sell at any price lower than \$51. Limit orders provide liquidity to the market because once they are posted, they sit on the book waiting for a counterparty to submit a market order. Market orders are “liquidity-taking” because they execute quickly against whatever the current best posted price is on the book and, in so doing, they eliminate the liquidity-providing limit order from the book. Thus, limit orders are “makers” of liquidity, while market orders are “takers” of liquidity.

A limit order is passive because after it has been posted, the order simply sits on the book waiting for a willing counterparty to submit a market order. If a counterparty does not materialize, the order does not execute. Consequently, there is a risk that the limit order will not execute (which is referred to as *non-execution risk*). In contrast, this cost is not borne by the trader who uses a market order.

In addition to non-execution risk, while posted on a transparent limit order book, a limit order can be “picked off” following an unexpected advent of unfavorable information (if the limit order trader has not yet received the news and has not withdrawn the order in time). This (along with non-execution risk) is a cost that a limit order trader incurs that a market order trader does not incur. On the other hand, limit orders can execute at better prices than market orders because they save the bid-ask spread. Consequently, resolving these trade-offs and choosing between a limit order or a market order calls for some strategic decision-making.

Other special order types are also used. Some of the more popular ones are:

- A “stop order” is an order that becomes active only if a trade is made at or through the stop price. For example, “sell 100 shares stopped at \$30, limit \$29”

means that once a trade has been made at \$30 or below, the order is activated and displayed on the book with a limit price of \$29.

- “Discretionary, not-held orders” give the broker the freedom to make the execution at any time and at a price that is fit and reasonable, given the investor’s goals. “Not-held” means that the broker is “not held responsible” if an attempt to get a better price fails and the order is eventually executed at an inferior price.
- “Pegged orders” are orders where the limit price is pegged to a benchmark such as the NBB (national best bid), the NBO (national best offer), or the midpoint of the two, and thus, the limit price changes as the benchmark price changes.
- “Iceberg orders” reveal only a small portion of the full order. Once the small, revealed portion of an iceberg order is executed, another small portion is posted on the book and displayed to the market.

Orders can also have special conditions attached to them. Two common ones are:

- Fill or kill (if the order cannot be filled upon arrival, cancel it)
- All or nothing (if the order cannot be filled in its entirety, do not fill it at all)

In today’s markets, more traders are relying heavily on algorithmic trading (also called algo trading or computerized trading) to enter their orders. Algo trading is based on computer programs that follow defined decision rules (algorithms) to generate and submit orders. Algorithmic trading typically operates at speeds that are impossible for a human trader to match (e.g., milliseconds). Nevertheless, a human participant must still design an algo and decide when to activate it.

For three reasons, some orders are not revealed to the market, and some are not even transmitted to a trading facility.

1. An exchange’s limit order book can only accept orders conditioned on price alone. The reason for this is that orders on the book must be subject to aggregation so that the total number of shares at a price can be unequivocally stated. Orders with special conditions (such as fill or kill) cannot be aggregated, are therefore kept separate, and are not disclosed.
2. Large traders do not want to reveal the full size of their orders because displaying this information will cause an undue market impact.
3. Participants who handle their orders strategically find it effective to enter an order only when they feel that the conditions are favorable for doing so.

The unrevealed (hidden) orders and the liquidity they provide are called *latent*. The latent orders translate into latent liquidity.

2.2 Trading Costs

Trading is a complex process because it is not costless. Let us take a closer look at what the costs are.

Trading costs fall into two categories: *explicit* and *implicit*. Explicit costs are visible and easily measured. They include commissions, fees, and taxes. Implicit costs are trickier to measure. They include the bid-ask spread that was described earlier, opportunity costs, and market impact.

Bid-ask spreads are a natural property of a continuous market. Bid and ask prices are established by limit order traders and/or by market makers. Because matched and crossed orders trigger transactions that eliminate them from the book, market bid-ask spreads are always positive. Further, with discrete prices, the spread must be at least as large as the smallest allowable price variation (currently one cent for stocks in the United States). Market order traders buy at the ask and sell at the bid, and for them, the bid-ask spread is the cost of a round trip (buying and then selling, or selling short and then buying).³

Opportunity costs are incurred when a trader is not able to complete a trade or when there are execution delays. This cost is particularly relevant for limit order traders. A limit order buyer incurs an opportunity cost if a stock's price rises and his/her limit order, because it was priced too low, remains unexecuted on the book. A seller similarly incurs this cost if a stock's price declines and his/her limit order to sell, because it was priced too high, remains unexecuted on the book.

Market impact (also referred to as "price impact") is encountered by large traders. For one thing, a large order sent as a single block to the market can "walk the book." That is, it can execute in part at the best counterpart bid or offer, and after it has cleared out the shares at the best price, it moves to the next price rung on the ladder (higher if it is a buy order and further down if it is a sell order), and so on. Or if the large order is brought to market in smaller pieces over an extended period, the process can trigger short-term trending that adversely impacts the prices a trader obtains. And there is one more thing: selling (or buying) conveys a negative (or positive) signal for a stock that causes price movements that other participants will jump on, thereby creating a price trend that augments the market impact cost for a large order.

Along with reducing returns, trading costs also cause investors to rebalance their portfolios less frequently. Accordingly, this results in investors holding portfolios that, in a costless environment, they would not deem optimal.

Some trading desks employ a metric called *implementation shortfall* (IS) to measure trading costs. IS is the difference between the price of an actual trade (after taking into consideration all commissions, fees, and taxes) and the price of a hypothetical trade based on a benchmark value. A commonly used benchmark is the midpoint of the national best bid and offer (NBBO), either at the time the trade decision was made or when the process of getting the order executed was first initiated. IS can be seen as a proxy of how good a trader is at implementing his/her trading decisions.

³Short selling is defined as selling shares you do not own but have borrowed. You borrow them from your broker with the hope of buying them back later at a lower price to close out your short position.

2.3 What Drives Trading?

In this section, we address a fundamental question: what drives trading? The various players in the market include speculators, hedgers, arbitrageurs, and market makers. On any given trading day, players who are meeting in the marketplace have different investment goals, different time horizons, and different appetites for risk and liquidity. The players also have different amounts of information and different (divergent) expectations about the stock. As this diverse set of players meet, the mixture leads to trading. As noted in the previous section (and we will repeat here), divergent expectations are attributable to information sets being huge, complex, imprecise, incomplete, inaccurate, and ambiguous.

Not only do participants have different expectations, they can also change their individual valuations at any time, either because of their own independent reassessments or upon learning the thoughts of others. A divergence of expectations, along with the attending interdependencies between different people's valuations, profoundly affects the dynamic process of price formation.

Trading results from two different types of shocks: *information shocks* and *liquidity shocks*. Information shocks are due to the advent of news concerning a company, its industry, and/or the macroeconomy.⁴ For instance, at the microeconomic (firm/industry) level, a drug company receives FDA approval for a new drug, and its expected 1-year forward share price shoots up. Or at the macroeconomic level, the chairman of the Federal Reserve announces positive unanticipated information about the prospects for the US economy, and major stock indexes vault higher.

All other reasons for trading are referred to as liquidity shocks. The meaning of "liquidity" when used with respect to a "liquidity shock" differs from "liquidity" when used with respect to the "liquidity of a market." We turn to the liquidity of a market later in the chapter.

Regarding liquidity shocks, they occur, not because of the advent of new information, but because of a change in some investors' individual desires to hold shares of a stock. For instance, an individual comes into money and buys shares, or needs money and sells shares. Or he/she has reassessed his/her expectation of a stock's 1-year forward price, or his/her risk and/or illiquidity tolerances have changed. Liquidity shocks can also be attributed to technical trading and the use of participant-unique algorithmic trading strategies. In contrast to information shocks that investors generally respond to in similar (but not necessarily identical) fashion, liquidity shocks are independent (uncorrelated) from investor to investor. Following a liquidity shock that pushes a stock's price away from an equilibrium value, the price will revert back toward the equilibrium it was pushed away from. On the other hand, an information shock changes a stock's equilibrium value, and there will be no reversion back.

⁴In Chap. 3 on macroeconomics, we offer a detailed exploration of how information shocks can affect trading and the liquidity of financial markets.

2.4 Price Discovery: A Major Function of a Marketplace

Price discovery means finding a stock's value that best reflects the broad market's desire to hold its shares. For a simple reason, the process is protracted and complex: investors have divergent expectations. Let us consider a highly simplified example. Assume that some investors value a stock at \$25 a share while others value it at \$24. What then should the stock's price be? This can be determined only by participants submitting their orders to the market and by their orders being translated into trades and transaction prices.

Good price discovery is, of course, important for those individuals who are participating in a trade.⁵ It is also important for a broad array of other uses: derivatives trading, estate valuations, mutual fund valuations and redemptions, marking positions to market, and dark pool pricing.⁶ Market-produced prices are also important to firms for assessing their costs of capital, for making share and stock options/warrants issuance and repurchase decisions, and for computing various price-related ratios (for instance, price-to-earnings and market-to-book ratios).

Price discovery, however, is not the only challenge. Quantity discovery is another big one. Institutional participants, because they want to minimize the market impact of their large orders, approach the market wrapped in a veil of secrecy. How do they find each other and trade if they are all trying to stay hidden? Answer: not easily and not always successfully. Large traders commonly "hold their orders in their pockets" or send them to an off-exchange, non-transparent trading facility (commonly referred to as a dark pool). Large traders also "slice and dice" their orders for submission to a "lit" (public) market, entering them in small pieces over an extended period of time.

2.5 Trading: The Implementation of an Investment Decision

We turn next to differentiating trading from investing. Before doing so, we call your attention to two important distinctions.

- *Brokers versus dealers*: A broker is an intermediary who, as an agent, brings an investor's orders to the market and, for this service, is paid a commission. A dealer (also called a market maker) trades with the investor, not as an agent, but as a principal who buys shares from customers who want to sell and who sells shares to customers who want to buy. In further contrast to a broker, the dealer does not charge a commission but instead posts bid quotes

⁵We discuss price discovery due to information arrival in Chap. 3.

⁶Marking to market is an accounting practice that involves restating the value of an asset to reflect its current market levels. A dark pool is a private financial forum for trading securities. We discuss dark pools in Sect. 2.8.

at which he/she will buy that are lower than the ask quotes at which he/she will sell.

- *Sell-side traders versus buy-side traders*: Investment banks and brokerage houses, either as brokers or dealers, are on the sell side of the market, selling trading services to buy-side customers. On the buy side of the market are retail customers and institutional investors. In the past, order handling and trading were predominantly done by the sell side but, in the advanced markets of today, many large institutional investors have their own well-developed trading desks.

With an eye on institutional investors, let us consider the difference between portfolio managers and buy-side traders. Investment decisions are made by portfolio managers while, as part of the same company, a separate entity referred to as a *buy-side trading desk* handles the implementation of the investment decisions. On both the buy side and the sell side, much experience is needed to be a good trader. On the buy side, a good trading desk can add to the overall performance of a portfolio. The opposite is also true – a poor trading desk can impair overall fund performance.

Trading, like investing, is a “professionalized” activity, but traders have a different career path than portfolio managers, and their skill sets are different. Like portfolio managers, traders require solid educational training, and continuing education is also called for in the ever-changing, complex world in which they operate. Traders need to keep up to date with technology, market structure innovations, and, of course, regulatory requirements.

Time has a different meaning for traders than for portfolio managers. A portfolio manager can take days, weeks, and, at times, a good deal longer to investigate a stock’s risk, return, and liquidity characteristics. But once an investment decision has been made and is passed to the trading desk, the clock accelerates and time acquires a different meaning. For a trader who has been given an order to work at the start of a trading day, 12:00 noon can seem like the long run. So we ask: what occurs in the short run that makes time so important? The answer: accentuated intraday price volatility.

2.6 Intraday Price Volatility

The challenge and excitement that attend trading are attributable to one thing: the turbulent price movements that occur over brief intervals of time (even within sub-seconds). In an environment of accentuated intraday volatility, market prices will move sharply in very brief intervals of time, and as they do, trading opportunities suddenly pop up and then quickly vanish. A trading desk’s own order handling decisions can cause adverse price moves, and poor order placement and imperfect market timing are costly. When costs are incurred, the gains that an asset manager might otherwise have realized from a good investment decision can be seriously eroded.

When focusing on volatility, think not of prices themselves but of percentage price changes that we refer to as “returns.” While we commonly say “price

Exhibit 2.1 Variance ratios

	Open/Day	Close/Day	Day/Month
ALL S&P500	4.36	1.82	1.31
TOP 50 SIZE	3.50	1.40	1.35

volatility,” we actually measure the standard deviation or variance of returns. Short-run (for instance, daily) stock return variances are substantially greater than long-run (for instance, monthly) stock return variances. However, there would not have been a difference between short-term and long-term variance if the intraday prices were efficiently set. The accentuation can be seen with reference to Exhibit 2.1.

Exhibit 2.1 shows ratios for opening half-hour to daily, closing half-hour to daily, and daily to monthly variances for SP 500 stocks during 2019.⁷ We calculate these ratios for each stock separately and present the average ratio. We adjust for measurement interval length by multiplying the numerator by 13 (the number of half-hour periods per day) for the open/day and the close/day ratios and by 21 (the number of trading days per month) for the day/month ratios. In row 1, we present the average ratio for all SP 500 stocks, and in row 2, we present the average for the top 50 stocks in terms of market capitalization. The accentuation of shorter period volatility is evident from the ratios shown: all are considerably greater than one, where one is the value that represents the benchmark for a perfect, frictionless market. The ratio is especially high for the open/day measure, as the opening half-hour is a period of price discovery (and hence accentuated volatility) following the overnight market close. Later in this chapter, we consider various marketplace realities that can account for the accentuated turbulence of intraday prices. At this point, we turn to a key finance variable: liquidity.

2.7 Liquidity

Liquidity is of major importance for equity markets. However, as we have previously said, it is not subject to simple definition and quantification, and for large investors in particular, finding it is a never-ending challenge. Liquidity is usually thought of in terms of both transaction time (the time it takes to complete a trade) and transaction cost (including the bid-ask spread and market impact). But without a good workable definition, an assessment of liquidity is generally based more on people’s perceptions than on generally accepted quantitative analysis. One thing, however, is quite apparent: when a market lacks liquidity, participants know it. For small cap stocks in particular, the illiquidity problem is especially acute, and as such, it has attracted intensified regulatory attention.

Across the broad spectrum of companies and for the economic growth of the entire macroeconomy, liquid, well-functioning markets are of major importance. A stock market, by offering liquidity, gives people the ability to easily buy and sell

⁷We included only the stocks that remained in the index the whole year.

shares when desired. With regard to selling, the need for an exit door is particularly critical. Investors will not buy a stock in the first place if they have insufficient assurance that they will be able to sell it sometime in the future with reasonable facility and at a reasonable price. As we just said, this is particularly important for small companies that are looking to raise money in the primary markets where new shares are issued.

So how might the liquidity of a market be quantitatively assessed? Liquidity has been viewed as the depth, breadth, and resiliency of a market. Depth refers to the size of posted orders that are at or close in price to the best bid and offer in the market. Breadth refers to having orders at numerous price points up and down the order book in relatively close proximity to the best bid and offer. Together, depth and breadth represent the number and size of revealed orders on the book. If the book has sufficient depth and breadth, orders can be executed quickly and in reasonable size.

Resiliency is a stock price's ability to retain an equilibrium value and to speedily regain an equilibrium value if pushed away by a liquidity shock. To repeat, resiliency encompasses two dimensions: (1) the initial impact of a liquidity shock and (2) the speed with which a dislocation is repaired. The initial magnitude depends on the breadth and depth of the market, and the bounce-back depends on the existence of unrevealed, latent liquidity.

Liquidity is supplied in two different ways: revealed (posted) liquidity and latent (not posted) liquidity. Revealed liquidity includes the quotes posted by dealers and by limit orders traders. But some traders' orders are not displayed, and these are the latent ones. Latent orders include orders with special conditions (such as "fill or kill" or "all or nothing" instructions identified in Sect. 2.1). Latent liquidity also includes orders that are "held in traders' pockets" while being worked strategically by a broker. Orders kept on investors' trading desks are also latent (their revelation to a market is commonly triggered by machine-driven algos). Because price moves can trigger the entry of latent liquidity into the market, latent liquidity is dynamic (as opposed to revealed orders on the book which are static). Static (posted) liquidity can be measured, while dynamic (non-revealed) liquidity is not observable. Accordingly, it is not possible to obtain a sufficiently broad measure of this vitally important characteristic of a market. However, illiquidity's *footprint* in a market can be assessed. In Sect. 2.6, we have called attention to illiquidity's big footprint: accentuated intraday price volatility.

2.8 Equity Market Structures

How orders are turned into trades depends on the rules and regulations that define a market's structure. Much development has occurred on the market structure front in recent years with striking technological advances, the emergence of new trading facilities, and an intensification of intermarket competition. Nevertheless, despite some positive developments, two issues have remained challenging: (1) providing

	Bid Size	Bid	Ask	Ask Size	
			24.55	10	
			24.54	50	
			24.53	12	
			24.52	10	
Bid-Ask Spread			24.51	4	
(24.51 – 24.49)	0	24.50			
	5	24.49			
Air pocket	0	24.48			
	100	24.47			
	20	24.46			
	15	24.45			

Exhibit 2.2 Limit order book for ABC stock

reasonable liquidity for smaller stocks and, for all stocks, (2) amassing sufficient liquidity for large, institutional-sized orders.

Trading systems can be classified according to three generic structures:

1. *Continuous order-driven markets.* A market is continuous if a trade can be made whenever, in continuous time, a buy and a sell order meet in price. In an order-driven market, prices are established by posted limit orders. Limit orders to sell set the prices at which market order traders can buy, and limit orders to buy set the prices at which market order traders can sell. The limit orders are posted on a “limit order book,” and in continuous trading, a trade is made whenever a buy order matches (or crosses) a sell order during normal trading hours (US markets open at 9:30 am and close at 4:00 pm). A representative limit order book for a continuous order-driven market is shown in Exhibit 2.2.
2. *Periodic call auctions.* In contrast to a continuous order-driven market, a call auction is a periodic order-driven market. With a call auction, participant orders are batched together for simultaneous execution at a single point in time, such as at the opening or closing of the trading day, and all executed orders transact at the same price – the *clearing price*. When the market is called, all buy orders equal to and greater than the clearing price are executable, as are all sell orders equal to or less than the clearing price. Clearing prices are set at values that maximize the number of shares that execute. By batching multiple orders and transactions together, a call auction concentrates liquidity, and in so doing, it can decrease intraday price volatility and reduce transaction costs for participants.⁸ A representative limit order book for a call auction is shown in Exhibits 2.3 and 2.4.

⁸The integration of revealed and latent liquidity could be better harmonized in a call auction as the latent liquidity provider could be more comfortable with revealing the orders in a batched trading, periodic environment. Why? For one reason: he/she can get price improvement.

Indicative Price	Bids	Cumulative	Price	Cumulative	Asks	Shares that would Trade
	0	0	24.55	402	97	0
	0	0	24.54	305	42	0
	25	25	24.53	263	55	25
	23	48	24.52	208	69	48
	15	63	24.51	139	32	63
	19	82	24.50	107	27	82
24.49	26	108	24.49	90	38	90
	24	132	24.48	52	32	52
	76	208	24.47	20	20	20
	84	292	24.46	0	0	0
	67	359	24.45	0	0	0

Exhibit 2.3 Call auction book

Indicative Price	Bids	Cumulative	Price	Cumulative	Asks	Shares that would Trade
	0	0	24.55	402	97	0
	0	0	24.54	305	42	0
	25	25	24.53	263	55	25
	23	48	24.52	208	69	48
	45	93	24.51	139	32	93
\$24.50	19	112	24.50	107	27	107
	26	138	24.49	90	38	90
	24	162	24.48	52	32	52
	76	238	24.47	20	20	20
	84	322	24.46	0	0	0
	67	389	24.45	0	0	0

Exhibit 2.4 Call auction book following the submission of another buy order (of 30 lots at \$24.51)

3. *Continuous dealer markets.* In a dealer market, multiple dealers (also called market makers) post the prices at which public customers can buy or sell shares. A dealer posts two-sided quotes: a bid quote at which the market maker will buy shares from a customer looking to sell and an ask quote at which the market maker will sell shares to a customer looking to buy. A dealer market is commonly referred to as *quote driven* (in contrast with the *order driven* market that we have just discussed). Dealers do not speculate in long-term price movements

Dealer	Bid	Ask	Dealer
Lion	20.40	20.41	Tiger
Fox	20.39	20.42	Bull
Tiger	20.38	20.42	Fox
Bear	20.38	20.43	Lion
Bull	20.37	20.44	Bear

Exhibit 2.5 Dealer market

of the instruments they trade. Instead, they seek to profit from small differences between their buying and selling prices, and they generally hold positions for short periods of time. The adage “stock sold to a dealer is still for sale” captures this reality of a quote-driven environment. A representative screen for a quote-driven market is shown in Exhibit 2.5.

2.8.1 Hybrid Markets

Hybrid markets combine call auctions and continuous trading. Stock exchanges internationally open and close their continuous markets with a call auction. Moreover, when under stress, price discovery can break down in the continuous market, and when this occurs, trading is halted and a call auction is used to reopen the market. The call auction reopening procedure is used because batching orders in multilateral, call auction trading facilitates order handling, sharpens price discovery, and enhances transparency. Call auctions, however, do not provide immediacy, but continuous markets do. Immediacy is appealing to many market participants, and thus, a hybrid market structure that combines these two market structures offers significant advantages.

2.8.2 Handling Large Orders

There are additional procedures available to meet the specialized trading needs of large participants:

1. Block trading is used for the sizable orders of large investors (typically institutional) who are trying to minimize the adverse impacts their large orders have.
2. Large orders can be negotiated by telephone or via an electronic interface.
3. To achieve better executions, a trader will commonly “slice and dice” a large “parent order” into smaller “child orders” that are sequentially submitted to the market over a longer period of time.
4. Also on the scene are “dark pools,” trading facilities that allow orders to be entered without disclosing participants’ trading interests.

Dark pools, so named to emphasize their lack of transparency, are private trading forums as opposed to public exchanges like the New York Stock Exchange and Nasdaq that are referred to as “lit” because of their higher levels of transparency. While dark pools came about primarily to facilitate block trading by institutional investors, they are also being used for trading small, internalized retail-sized orders received by brokerage houses.⁹ That is, the orders are being executed without having been sent to a public exchange.

So here is the situation:

1. Price discovery takes place through trading in public exchanges.
2. These prices are used for trading in the dark pools.
3. Much quantity discovery takes place in the dark pools, particularly for large investors.

The partial separation of price discovery and quantity discovery is not without drawbacks. Equilibrium values for price and quantity are better attained when the price and quantity variables are solved for simultaneously.

2.9 Financial Markets and the Process of Turning Orders into Trades

In this section, we drill down in more detail on how orders are handled and turned into trades in continuous order-driven markets, in call auctions, and in dealer (quote-driven) markets, the three primary market structures.

2.9.1 Trades in Continuous Order-Driven Markets

As we discussed earlier, in the continuous order-driven market, prices are determined by limit orders that are placed in what is referred to as a “limit order book.” How does this work? Let us take a closer look.

As we have noted, a market order does not specify a price, whereas a limit order does. A limit price is a maximum price for buy orders and a minimum price for sell orders (which is why they are referred to as “limit orders”). In continuous trading, limit orders generally execute at the price at which they have been entered, while limit orders in a call auction, unless their price is the same as the clearing price, are price improved.

We have said this before, but it merits repeating. Traders who place limit orders avoid paying the bid-ask spread. Those who place market orders incur the cost represented by the spread. But limit order placers incur the risk that their orders might never execute. Also, if a limit order does execute, it could be at a disadvantageous

⁹Internalization occurs when a brokerage firm that has received both buy and sell orders matches and executes them internally rather than sending them to an exchange.

price (buying at a share price as it starts to fall because of the arrival of unanticipated negative news, or selling at a share price as it starts to rise because of the arrival of unanticipated positive news). Because limit order traders face these risks, they require an incentive for being a liquidity provider, and saving the spread is their compensation.

Exhibit 2.2 shows how a limit order book might look for a hypothetical stock, ABC. The exhibit shows limit orders to buy (“bids”) at prices ranging from \$24.45 to \$24.49 and limit orders to sell (“asks,” which are also referred to as “offers”) at prices ranging from \$24.51 to \$24.55. The columns labeled “bid size” and “ask size” show the number of round lots entered at each limit order price (a round lot is 100 shares). For example, in the bid size column, the number “5” at \$24.49 indicates that there are five “round lots” (500 shares) to buy ABC stock at a price of \$24.49. These 500 shares might be one trader’s order or multiple traders’ orders because the limit order book aggregates all orders that are entered at a given price.

Exhibit 2.2 displays two prices where there are no orders on the book (\$24.50 and \$24.48). The zero bid size at 24.48 is a gap in the order book that can occur when the market for ABC stock is not very liquid. Gaps such as this one are *air pockets*, and they are present when an order book is “thin” (not “deep”). Ideally, we would like the order book to be deep and to have no air pockets.

The absence of orders at \$24.50 is not an air pocket. It is a price point within the bid-ask spread. As we have noted, the spread is the lowest (“best”) offer (\$24.51) minus the highest (“best”) bid (\$24.49). In Exhibit 2.2, the spread is shown to be \$0.02 per share (two cents). The two-cent spread can exist when the minimum price change is one cent because, for a strategic reason, no participant has chosen to post a limit order within one cent of a counterpart order. To understand this strategic reason, consider a buyer who is entertaining the possibility of placing a limit order at \$24.50 when, for one cent more, he/she could buy with certainty at \$24.51. With a one cent spread, there is not much to save by placing a limit order and more to lose because of the chance that it will not execute. Thus, the certainty of executing at \$24.51 exerts a “pull” on the incoming order, and the pull keeps it from being entered at \$24.50. So instead of placing a limit order at \$24.50, the buyer enters a market order that executes at \$24.51. The pull of certainty accounts for spreads that are wider than the minimum price change of one penny. It is this “pull of certainty” that explains why the absence of an order within the spread is not just another air pocket.

Exhibit 2.2 shows a snapshot of the order book at a single point in time. In the order book shown in Exhibit 2.2, a market order to buy will execute at the best offer (\$24.51), and a market order to sell will execute at the best bid (\$24.49). These posted prices will change over time as new orders arrive and as existing orders are executed or cancelled.

Suppose that a new market order to buy 5000 shares (50 round lots) is submitted to ABC’s limit order book. As shown in Exhibit 2.2, the number of shares available at the best offer of \$24.51 is only four round lots (400 shares). This means that not all of the 5000 shares to buy by market order can be purchased at the best asking price of \$24.51. In this case, the order will *walk the book*. Walking the book means

that any market order that exceeds the size shown at the best quote will trade at ever higher prices (if it is a buy order) or at ever lower prices (if it is a sell order) until it is filled. Accordingly, with walk-the-book pricing, the first 400 shares of the 5000 share order are bought at \$24.51, the next 1000 shares are bought at \$24.52, the next 1200 shares are bought at \$24.53, and the last 2400 shares are bought at \$24.54.

The overall purchase price of \$24.5312 for the 5000 shares is a weighted average of the four different execution prices, where the weights are the numbers of shares traded at each price point. We thus have $((4 \times 24.51) + (10 \times 24.52) + (12 \times 24.53) + (24 \times 24.54))/50 = \24.5312 . Because this buy order is much larger than the number of shares available for sale at the best offering price (\$24.51), the overall weighted average price of \$24.5312 is 2.12 cents per share higher. For the 5000 share order, the additional cost is \$106 ($\0.0212×5000). The higher purchase price will, of course, reduce the buyer's net return on ABC's stock. These additional costs add up and compound over time because they can be incurred each time the investor trades.

2.9.2 Trades in Call Auction Markets

As discussed earlier, a call auction is a periodic (as opposed to a continuous) order-driven market. With a call auction, participants' orders are batched together for simultaneous execution at a single clearing price at a pre-announced point in time. When the market is called, all buy orders equal to and greater than the clearing price are executable, as are all sell orders equal to or less than the clearing price. Trading prices are set at values that maximize the number of shares that execute.

By batching many transactions together, a call auction concentrates liquidity. In so doing, it can significantly decrease transaction costs for participants. Calls also facilitate better "quantity discovery" because they allow larger orders to be executed with reduced information leakage and thus lower market impact costs.¹⁰ Large buyers and sellers who might be reluctant to send their orders to a continuous market may submit them to the call because of market impact costs being lower. The integration of revealed and latent liquidity could also be better achieved in a call auction because latent liquidity providers can get price improvement (with rare exceptions, this does not happen with continuous order-driven trading). Two important operational details of any call auction are the pricing mechanism used and how buy-sell imbalances at the call auction price are handled. The usual pricing mechanism in a call auction is to choose the price that maximizes the number of shares that trade. The usual rationing mechanism to deal with an order imbalance is time priority (first in, first out).

Exhibit 2.3 shows the accumulated orders on the book of a call auction. Buy orders accumulate from the high price to lower prices on the left. Sell orders accumulate from the low price to higher prices on the right.

¹⁰The concern for large buyers / sellers that other traders might find out about their orders which would lead to trades at unfavorable prices is referred to as "information leakage".

For instance, there is a bid to buy 15 round lots priced at \$24.51. Anyone who places a bid at \$24.51 is surely willing to buy at that price or at a price that is lower. The cumulative quantity to buy at \$24.51 is 63 (the number of round lots entered at bid prices of \$24.51 and higher). Likewise, the call auction sellers who place an ask at 24.51 per share would certainly be willing to sell at 24.51 and higher. Since there are cumulative numbers of 63 bids and 139 asks at \$24.51, the minimum of those numbers, 63 round lots, would trade at that price. The indicative clearing price is the price at which, as we have noted, the maximum number of shares would be traded. In this example, the indicative price is \$24.49 with 90 round lots executed (the minimum of 108 cumulative buys and 90 cumulative sells).

Continuing to refer to Exhibit 2.3, you can see that at \$24.49, the number of round lots that trade is maximized at 90 by noting that at one tick up, 82 round lots (less than 90) would trade (the minimum of 82 and 107) and that at one tick down, 52 round lots (again less than 90) would trade (the minimum of 52 and 132). Thus, if the market gets called at this instant, \$24.49 would be the realized clearing price. The cumulative number of round lots to buy is 108 at \$24.49; the 90 out of the 108 that execute is determined by applying the price priority rule. Notice that with an execution price of 24.49, the bids above 24.49 and the asks below 24.49 receive price improvement (i.e., the realized transaction prices are improved vis-à-vis the limit prices on the orders that the participants had submitted to the call).

Now let a new buy order comes in at \$24.51 for 30 round lots before the market is called. This will increase the total number of bids by 30 at 24.51 and every price point below it. As a result, the new maximum number of shares that would be traded is now 107 at a new indicative price of \$24.50. And so call auctions progress as orders arrive and the book builds. As new orders keep coming in, the indicative price keeps fluctuating until the market is “called,” at which point the auction price and the number of shares that trade are established.

We have presented a basic example of call auction trading, but the calls can differ in any number of ways from market to market. For instance, each market would have further rules about how to prioritize the trades when there is an imbalance between the bids and the asks at the clearing price because only by rare chance will the two sides of the market match exactly. When the number of shares on the two sides differs, the lesser of the two sides determines the total number of shares that executes, and shares on the bigger side of the market have to be rationed. Generally, as we have noted, the shares on the larger side that execute are determined by applying a time priority rule (the first orders placed get executed first). If application of the time priority rule does not produce an exact match, a further rationing rule is required, and various alternatives are possible. This is part of the complexity of designing a real-world market. In the example above, these additional rules would govern which 107 bids out of the 112 would end up being executed. Another way in which calls can differ from one another is in the amount of information about the call book that is disseminated while the book is still building prior to when the market is called. The full book may not be revealed, and only the indicated clearing price is shown to the public.

2.9.3 Trades in Continuous Dealer Markets

As described earlier, in a dealer market, dealers state the prices at which public customers can buy or sell shares. Thus, a dealer market is commonly referred to as quote driven. A dealer posts two-sided quotes: a bid quote at which the market maker will buy shares if a customer is looking to sell and an ask quote at which the market maker will sell shares if a customer is looking to buy.

In Exhibit 2.5, we see a market consisting of five competing dealers. In this market, the current best bid is quoted by Lion (\$20.40), and the current best ask is quoted by Tiger (\$20.41). In other words, Lion is eager to buy, and Tiger is eager to sell, and therefore, they are offering the most aggressive quotes in the market. Their quotes set the NBBO, the national best bid and offer.

Customers may choose to send their orders to any dealer that they prefer regardless of the prices that the dealer is currently quoting. This practice is called *preferencing*. The receiving dealers have the right to reject the orders, but they rarely do. Instead they will accept the order and execute it at the best bid or offer quoted in the market.

Dealers trade against public customers and other dealers. They adjust their quotes as the day progresses, as market conditions evolve, and as their own inventory levels fluctuate. As dealers see substantial portions of the aggregate order flow, they gain insight into the balance between public buying and selling pressures. Collectively, the competing dealers play the central role in discovering prices in the quote-driven environment.

2.10 Regulation, Technology, and the Quality of Market Structure¹¹

For many years, three powerful forces have been reshaping the operations and quality of the equity markets: regulation, competition, and technology. In this section, we focus primarily on the first, regulation, and on its interrelation with the other two. In so doing, we concentrate on US markets, but the picture in many ways is the same in other major equity markets around the world. Chapter 4 deals, in further detail, with technology.

Regulators have focused a great deal on competition, primarily on how it exists between the exchanges and other trading venues. In so doing, they have relied, in good part, on competition to drive exchange fees down to competitive levels and to encourage technological innovation. The approach has been effective regarding fees, but it has come at a cost. Strengthening competition between different trading facilities fragments the order flow, and fragmentation can have harmful effects on the quality of price discovery.

¹¹This section is a modified reprint with permission of “Perspectives: The Interplay Between Regulation, Competition, and Technology and the Transformation of Our Equity Markets” by Ozenbas and Schwartz, *The Journal of Portfolio Management*, November 2020.

In one way, regulation can impede, not spur, competition and market structure development. New innovations must receive regulatory approval, and because regulators like to proceed with great caution, getting approval is generally a lengthy process. A good example involves the pioneering introduction of the first electronic call auction trading facility, the *Wunsch Auction System*. Founded in 1990, this new market became the Arizona Stock Exchange (AZX) in 1991. Not surprisingly, the New York Stock Exchange (NYSE) did not welcome a novel competitor, and the Securities and Exchange Commission (SEC) was concerned about the effect an electronic call could have on the Big Board's operations.¹² After a great deal of lengthy deliberation, the SEC did give the Wunsch Auction System a green light, but in so doing, the commission imposed a critical restriction: the call auction could not be held until 30 minutes after the market's 4:00 pm close. Not many traders hang around their desks after the main market has closed, and with this constraint, the new trading facility could not succeed. After years of trying, in October 2001, the AZX closed down because of insufficient order flow. But it had been on the right track. In 2020, the dollar volume of trading at the opening and closing calls which were by then being run by the NYSE accounted for about 9.7% of the total daily trading volume of the same stocks.¹³

Technology development has massively transformed the equity markets in the decades that followed the 1975 Amendments. Consider the speed with which orders are handled and turned into trades: in the pre-electronic era, the trade clock ticked at a slow enough pace for humans to follow price formation on a trade-to-trade basis; today, markets can change from microsecond to microsecond, and the trade-to-trade evolution of price formation cannot be followed by eye, only by computer. Consider the fragmentation of the order flow: in the past, the New York Stock Exchange enjoyed an 80% market share for its listed stocks; today, the NYSE's market share is hovering around 20%, and trading is dispersed over roughly 40 dark pool trading facilities and 16 exchanges. Consider the intermarket linkages: they used to be weak and slow; today's electronic markets are fast and interconnected, not only across exchanges and the off-exchange facilities but also between the stock and derivative markets. Consider how trades are accomplished: in the past, they were made by human-to-human interaction, either face-to-face or by phone; in today's super rapid, super interconnected markets, trades are also being made by computer orders meeting computer orders without direct human intervention (i.e., computer-driven algorithmic – algo – trading).

There is another big one we can thank technology for: the availability of data. In 1975, end of day, closing prices were reported in the papers, and that was about it; today, we have electronically delivered, intraday data with a microsecond time stamp for quotes, prices, trading volumes, and market indices, and the sheer amount of this data is enormous. Additionally, detailed audit trails are now available for regulators to peruse.

¹²“Big Board” is a common nickname for the NYSE.

¹³Calculated by the authors using 2020 data from the TAQ and CRSP databases for 2593 stocks listed on the NYSE (some small stocks were eliminated due to missing data).

Put this together and what do we have? We have speed, we have competition, we have an extraordinarily complex environment, and the regulators are faced with a huge challenge with regard to their obligation to police the markets for abuses of power and position.¹⁴ Moreover, the electronic markets can be fragile. With electronic order submission and executions, every condition must be planned for because preprogrammed computers cannot implement adjustments that human participants are capable of making when they are given reasonable freedom to do so. That is, a computer code has to be rewritten, while a human can make adjustments and corrections on the fly. Does this reality call for further regulatory intervention, or should the markets be left to sort it out by themselves?

Two things are clear: (1) the markets are the innovators, and (2) keeping regulations properly aligned with ever-evolving market structures is necessary but not easily accomplished. Regulatory oversight is necessary because of the enormous importance equity markets have for the macroeconomy, because a technological breakdown can have disastrous impacts on the financial markets, and because a poorly functioning secondary market (where already listed shares are traded) can make it more difficult for listed companies to raise funds in the primary markets (where new shares are issued). However, good regulatory policy is extraordinarily difficult to formulate. The issues are complex and regulatory intervention can give rise to *unintended consequences* (in medicine, “unintended consequences” are called “side effects”). In part, the complexity is attributable to various issues being very thorny. Here are three examples:

- There is a trade-off between promoting intermarket competition (which calls for fragmenting the order flow across trading venues) and promoting competition within the order flow (which calls for consolidating the order flow). What is the right balance between these two kinds of competition?
- All participants value transparency (the rapid public display of transaction prices, quotes, and trading volume). However, many participants, large institutional investors in particular, do not want their own trading intentions disclosed. So how should the conflict between the collective desire for transparency and the individual needs for opacity be resolved?
- Free competition between firms has been widely relied on in the United States to advance economic development, and for much of its history, market structure has evolved naturally in a free environment. Government (with exceptions of course) does not tell a firm such as GAP how exactly to produce clothes. Should it be

¹⁴Manipulating a market by “spoofing,” for instance, presents a particularly interesting challenge to regulators. Spoofing is the act of placing a bid or offer quote with the *intent* of cancelling the quote before it executes. Note the word “intent.” How might regulators infer a trader’s intent from his or her order placement decisions? There can be valid reasons for withdrawing an order soon after it has been placed. To name two: (1) the broad market or the market for the specific stock has changed, and (2) the trader has placed orders for two or more stocks and one or more of them has executed. Also, detecting and controlling spoofing is far from a simple matter because orders can be submitted to a sizable number of alternative facilities as well as to the futures and options markets. So this is what the regulators face: intent is hard to prove, and in the complexity of the markets, manipulators can find cover.

telling the equity markets how to produce trades and transaction prices? Yes, some government intervention in equity market structure development is no doubt called for, but what should the balance be between free market development and regulatory intervention?

Government regulation in the US equity markets operates on three levels: the US Congress, the Securities and Exchange Commission (SEC, established in 1934), and on the state level. In addition, equity markets have their own *self-regulatory organizations* (SROs). The New York Stock Exchange and the National Association of Securities Dealers had their own SROs until July 2007 when the SEC approved the merger of the two to form the Financial Industry Regulatory Authority (FINRA).

For the most part, regulation exists to police the markets for abuses of power and position, including fraud, manipulation, and trading on insider information. However, for almost half a century, government regulation by the Congress and the SEC has extended into overseeing the very structure of the securities markets.

The first major government regulatory foray into market structure occurred when the US Congress enacted the 1975 Amendments to the Securities Exchange Act of 1934. At the time, the Congress was concerned about the extent to which equity market dealers were profiting from unduly wide bid-ask spreads, and they were unhappy that commissions had been fixed at unjustifiably high levels. In addition, the amendments were enacted following the creation of the Securities Investor Protection Corporation (SIPC) in 1970 by an act of the Congress. SIPC is a nonprofit corporation that financially protects the clients of its member brokerage firms if those brokerage firms are forced into bankruptcy. Consequently, government, having become directly involved in the financial stability of the broker-dealers, now had a vested interest in these firms staying healthy and avoiding bankruptcy.

The Amendments included two items of particular importance: (1) it precluded trading commissions from ever being fixed (as they previously had been), and (2) it mandated the development of a *National Market System (NMS)*. Regarding the NMS, four objectives were stipulated:

- Enhance the economic efficiency of transactions (i.e., reasonable transaction costs).
- Ensure fair competition among brokers, dealers, and markets.
- Ensure the broad availability of information on quotations and transactions.
- Provide the opportunity, consistent with efficiency and best execution, for investors' orders to be executed without the participation of a dealer.

The SEC was charged with implementing the 1975 Amendments. This was not an easy task. For starters, in mandating the development of a National Market System, the Congress provided no definition of what an NMS was. Much

discussion, many meetings, and a number of conferences followed, and further regulations were enacted.

Today, the markets are far more connected by computers, various trading costs have been reduced, and considerably more information on quotes and transactions is available (thanks again to computers). And with regard to the fourth bullet above, dealers are far less prominent in today's equity markets. However, further regulations were enacted with respect to achieving the goals of the 1975 legislation. Below are four regulatory developments in particular that affected the role played by competition, that were initiated by the SEC over the years that followed the Securities Acts Amendments of 1975.

1997: Order Handling Rules. Comprehensive changes were mandated in 1997 by the SEC in the rules governing share trading in the United States. Known as *Order Handling Rules*, the regulation effectively ended market makers' dominance in price setting. These rules primarily affected Nasdaq, the world's second-largest stock market at the time (following the New York Stock Exchange), as Nasdaq was primarily designed as a dealer market. More transparency was required. Moreover, the previously private electronic systems used to trade big orders were opened up to the public. As a result, prices on these systems, which were often better than those offered by the market makers, became visible and available to the public. Following this regulatory initiative, several new execution venues including alternative trading systems (ATSs) and electronic communication networks (ECNs) opened.¹⁵ Their arrival led the way to a steady decrease in the market share of all traditional stock markets (including Nasdaq and the NYSE).

2001: Decimalization. The SEC ordered all US stock markets to convert to trading in decimals by April 2001. Prior to this date, prices had been quoted in 1/8^{ths} and, more recently, in 1/16^{ths} of a dollar, as opposed to decimals, the norm in other international equity markets.¹⁶ The main intent of this rule was to decrease the bid-ask spread and, hence, one of the costs of trading. With decimal pricing, the minimum tick size shrunk to a penny and bid-ask spreads thightened. This, however, had another consequence: decreased spreads made dealer operations significantly less profitable and the structure of the dealer market changed.

NYSE Rule 390. On May 17, 1792, twenty four stockbrokers signed an agreement at 68 Wall Street in New York City. According to legend, they met under a buttonwood tree, and the document they signed became known as the Buttonwood Agreement. It was historic. The Buttonwood Agreement marked the founding of the New York Stock Exchange.

¹⁵ Electronic communication networks (ECNs) are a type of alternative trading system (ATS) that trades listed stocks and other exchange-traded products. Unlike dark pools, another type of ATS, ECNs, display orders in the consolidated quote stream. Like ATSs, ECNs are required to register with the commission as broker-dealers and are also members of FINRA.

¹⁶ The 1/8th quotation dated from the colonial period, when the most common unit of currency used was the Spanish dollar, also known as "piece of eight," where a Spanish dollar was worth 8 Spanish silver reales.

The 24 brokers agreed on two points: their commissions were fixed at 0.25%, and they were to deal only with each other. Eventually, their agreement to deal only with each other became codified as NYSE Rule 390. Specifically, Rule 390 stated that an exchange member receiving an order for an exchange listed stock must bring that order to an exchange floor to be executed, that off-floor trading by exchange members was prohibited.

With regard to off-floor trading, in carrying out its congressional mandate, the SEC announced in June 1977 that NYSE Rule 390 was to be removed by the end of the year. A loud outcry against this was successfully raised by the industry, and at the last minute, the SEC postponed the rule's removal. But the threat of removal remained.

What motivated the regulators to seek Rule 390's elimination? In 1975, the NYSE was far and away the dominant exchange. Central to the exchange's market model was the specialist, a market maker who had the affirmative obligation to make a fair and orderly market for the stocks traded at his or her post. The NYSE's market share of the order flow was enormous, and all the orders for a specific stock went to the trading post of the specific specialist firm to which the stock was assigned. So think about it. Where in this model was competition? Where were the substantial competitive pressures that would keep trading fees low and innovation robust? These are the questions the regulators asked, and they did not like the answers.

However, in one way, the NYSE did face competition. Historically, there has been intense competition between the three largest exchanges in the United States – the NYSE, Nasdaq, and the American Stock Exchange (AMEX). Each has fought hard to attract the new listings of corporations that are going public and to get companies that are already public to switch their listings. This competition has certainly given the exchanges an incentive to improve the quality of their markets. But is this form of competition sufficient?

In any event, the battle over Rule 390 continued. The next regulatory action was in 1980 when a new SEC rule, Rule 19c-3, became effective. The new rule chipped away at Rule 390 by permitting the off-board trading of stocks that were listed after April 26, 1979. However, the 19c-3 stocks were relatively few in number, the large proportion of NYSE stocks remained subject to 390, and 19c-3 did not deliver the results the SEC was looking for. But the commission did not give up. Finally, on May 8, 2000, NYSE Rule 390 was repealed.

What effect did the repeal have? In the years and months leading up to May 2000, the NYSE's market share hovered around 80% to 90%. It took some time for participants to adapt and for the order flow to respond, but then the exchange's market share dropped precipitously. In 2003, it had fallen to 50%, and by 2019, it was hovering around 20%. What the regulators were looking for they got. The markets did become a great deal more competitive.

Regulation National Market System (Reg NMS) and the Order Protection Rule. Reg NMS is a set of rules established by the SEC in 2007 that are designed to

further strengthen the competitive structure of the US equity market. The major provision in Reg NMS that we focus on here is the Order Protection Rule. To ensure that investors buy and sell at the best available prices, the Order Protection Rule requires that the most aggressive quotes (highest bids and lowest offers) across the different trading venues be protected. Alternatively stated, the rule disallows executing a trade at a price inferior to the best bid and offer quotes (the protected quotes) in the market. For instance, if the best bid on Venue A is \$20.12 and the best bid on Venue B is \$20.15, an incoming sell order must be sent to and executed against B's quote of \$20.15. If the seller's order were to execute against A's \$20.12 quote, B's higher bid would have been "traded through." For this reason, the Order Protection Rule is also referred to as the Trade Through Rule.

The Order Protection Rule can certainly come across as being fair and reasonable. Think of how the buyer who entered the higher bid in Venue B would feel if his or her \$20.15 posting was traded through by a trade made in Venue A at the lower price of \$20.12. But do not forget that each quote has not only a price but also a size component to it. What if the incoming sell order is for 10,000 shares while, concurrently, the \$20.15 bid in Venue B is for 200 shares and the \$20.12 bid in Venue A is for 12,000 shares? And what if realizing a fast execution is of critical importance for the seller? Let us ask the question this way: should investors seeking large volume trades be required to first access small-sized quotations? Should venues be denied the ability to compete by offering good trading possibilities at prices above or below the national best bid and offer (NBBO)? What is a fair, equitable, and efficient way to direct order flow to different venues in a fragmented market? And speaking of fragmentation, a result of Reg NMS was more trading venues opening up, which further dispersed the order flow. Clearly, with regard to this complex issue, it is easier to ask questions than to provide simple answers to them.

Reg NMS had another effect, one that is attributable to an exception within the bill. Returning to our example, for Venue B's \$20.15 quote to be protected, it had to be immediately and automatically accessible. This exception had a consequence of major importance. The New York Stock Exchange, which for many years had resisted introducing electronic trading, now had no other choice. For its own quotes to be protected, the exchange had to offer fast and automatically accessible quotes. In other words, it had to institute an electronic trading platform. And it did.

The regulations we have discussed were intended to increase competition. They have and, along with technology development, they have had a powerful impact on market structure. Today, the speed with which traders can check quotes in multiple markets has become of major importance. This in turn has given rise to an enormous technology investment and to high frequency trading. Today, there are exchanges. Additionally, private trading platforms are competing with public markets, offering cheaper but, as we discuss earlier in this chapter, largely nontransparent alternatives

to public exchanges.¹⁷ Put it all together and what do we have? The landscape today is far more complicated.¹⁸ And there is another development of major importance. Today, dealers play a strikingly diminished role in bringing liquidity and price discovery to the markets.

The combined impact of regulation, competition, and technology has indeed transformed equity markets in the United States, and similar developments have been experienced in other major markets around the world. The difference between what markets are today and what they were when the 1975 Securities Acts Amendments were enacted is breathtaking. Few people would want to go back to where we were in the days of old. Nevertheless, problems concerning market quality persist. Intraday volatility remains significantly accentuated. The difficulty of executing large institutional orders remains formidable. Price discovery is still a challenge. The market for small capitalization stocks needs to be improved. And one problem that underlies all of the above persists: markets remain too illiquid, even for the largest capitalization stocks. Accordingly, we end this section with two questions. (1) How can more liquidity be attracted to the market? 2) How can latent (hidden) liquidity be more effectively integrated with revealed liquidity and, better yet, transformed into revealed liquidity?

2.11 Wrapping It Up: Market Efficiency in a Non-frictionless World

We started this chapter by calling your attention to something of critical importance: the relationship between fundamental information and the prices of equity shares. Understanding and appreciating this relationship requires basic institutional knowledge and, importantly, a solid comprehension of the following concepts:

1. Frictionless versus non-frictionless markets
2. Trading decisions versus investment decisions
3. Risk versus return versus liquidity
4. The drivers of trading: information shocks versus liquidity shocks
5. Explicit trading costs versus implicit trading costs
6. Posted liquidity versus latent liquidity
7. Measuring liquidity directly versus assessing illiquidity's footprints in the transactions tape
8. Short-period (e.g., intraday) price volatility versus longer-period (e.g., one month) price volatility

¹⁷According to the TABB Group, trading in dark pools comprised approximately 39% of all trading volume as of 2019.

¹⁸One example of complexity is that to attract more order flow, the competing exchanges now have complicated fee structures that include rebates and discounts which are generally referred to as maker-taker fees. The rebates are being paid to traders who add (make) liquidity, while the fees are being charged to traders who take liquidity from the market.

9. Homogeneous expectations versus divergent expectations
10. Price discovery, a major function of a stock exchange
11. Order types, most importantly, limit orders and market orders
12. The structure of an equity market

Let us circle back to the first entry on the list: frictionless versus non-frictionless markets. Think about it. The eleven other items on the list are important only because markets are not frictionless. With this in mind, let us consider the eleven other items on the list with the knowledge that the market is not frictionless.

In a frictionless environment, trading would be irrelevant. In a frictionless environment, only risk versus return would matter because all markets would be perfectly liquid. Liquidity shocks would not impact share prices. Trading costs, either explicit or implicit, would not exist. There would be no differentiation between posted liquidity and latent liquidity. Liquidity would not have to be measured, and there would be no illiquidity footprints in the transaction record. In a perfectly liquid world, short-period price volatility would not be accentuated and thus, if short-period and long-period volatility are both expressed in terms of the same measurement interval (e.g., per month or per year), the two measures would be identical. Because it would be costless to obtain and instantly evaluate all new information, every participant would be completely and identically informed, and under this condition, expectations could be homogenous. Totally accurate price discovery would be instantaneously achieved, and equity shares in the perfectly liquid, frictionless world would have fundamental values. Order submission would be instantaneous and costless, and differentiated order types would not be needed. And what could there possibly be to say about market structure if markets were perfectly liquid?

What do you think of the frictionless world? It is a difficult environment to imagine literally existing, and of course, it does not exist. To achieve a more complete comprehension of the workings of a real-world financial market, the realities of a non-frictionless market must be comprehended. Yet for some purposes, the simplifying assumption that markets are frictionless enables a rigorous, insightful model to be achieved. This certainly is the case for the Capital Asset Pricing Model (CAPM) that we dealt with in Chapter 1. Similarly, the Black-Scholes option pricing model assumes frictionless markets as well.

Analysis of a frictionless market yields another insight. In a frictionless environment, fundamental information is immediately and perfectly incorporated into share prices, and thus, share prices are instantly set with total efficiency. So what could cause a share price to change? Only new, unanticipated information. Current expectations based on the existing information set are not new information. By “new information,” we mean totally new and totally unpredictable.

Now, if informational change cannot be predicted, might one still be able to predict what the next price change will be? The answer is a resounding “no.” In statistical terms, in the perfectly efficient environment, stock price changes are, from one change to the next, uncorrelated. “Random walk” is the term used to describe this

property of an informationally efficient market. Let us be clear: it is returns (price changes) that are uncorrelated, and it is prices that take a random walk. Now reverse the logic. Because random walk is a property of a perfectly efficient market, random walk tests have been used to assess the informational efficiency of a market and, in so doing, to test the *efficient market hypothesis (EMH)*.

Have the random walk model and the EMH been validated? From our perspective on trading, the most relevant tests of the EMH are based on very short period, intraday data, and the tests have shown that intraday correlation patterns do indeed exist, thereby rejecting the hypothesis. This is consistent with what we said earlier in this chapter about intraday price volatility being accentuated. Both positive and negative intraday correlation patterns exist in complex combinations. However, the patterns keep shifting, and exploiting them with a profitable trading strategy is far from an easy task. Wise traders should recognize and deal with this reality. The intraday correlations say a lot about trading not being simple, and they underscore the need for having a well-designed market structure, along with good market structure regulation.

Equity markets have evolved tremendously in recent decades, and change is continuing apace. This is attributable to the three big drivers: competition, technology, and regulation, all three of which impact market structure. Regarding market structure, aside from issues concerning the abuse of power and position (such as insider trading and price manipulation) and technology's reliability, the quality of a market is equivalent to the quality of its liquidity provision. Illiquidity is a manifestation of impediments in trading that we have referred to as friction. We do not operate in a frictionless environment and never will. No matter how efficient our trading systems, there are limits to how liquid participants can expect a market to be.

Friction distorts the relationship between realized prices and underlying equilibrium values. Accordingly, prices bounce around equilibrium values, short-term volatility is accentuated, and correlation patterns are introduced in the return data. The bottom line is that, in a non-frictionless market, liquidity is not fully available when investment decisions are being implemented. Thus, substantial implementation costs can appreciably lower investment returns, and to the extent that they put downward pressure on share prices, they increase the cost of capital for listed companies.

A major economic *raison d'être* for a secondary market (where already listed shares are traded) is to enable shareholders to liquidate their positions with reasonable facility when so desired. But in stock markets today, liquidity is not consistently available for all stocks (both large and small), for all time periods (both throughout the trading day and inter-day), and for all market conditions (both normal and particularly stressful), and the efficient market hypothesis (EMH) should not be totally accepted. But neither should the EMH be totally rejected. It stands as a warning to anyone who thinks they have found a magic bullet, a "sure fire" formula that would enable trading based on mispricing to be consistently profitable.

We conclude this chapter with two thoughts: (1) markets are efficient enough to be a strict disciplinarian for any unduly rambunctious trader, and (2) trading frictions certainly have their dark side, but there is a bright side as well. Jobs exist because of them, and they make work a good deal more challenging, interesting, and exciting. This certainly is the case for equity trading.

Open Access This chapter is licensed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.

