1

The Ecology of Financial Markets

Buyer: How much is it?
Seller: £1.50.
Buyer: OK, I'll take it.
Seller: It's £1.60.
Buyer: What? You just said £1.50.
Seller: That was before I knew you wanted it.
Buyer: You cannot do that!
Seller: It's my stuff.
Buyer: But I need a hundred of those!
Seller: A hundred? It's £1.70 apiece.
Buyer: This is insane!
Seller: It's the law of supply and demand, buddy. You want it or not?

(Translated from "6", by Alexandre Laumonier)

A market is a place where buyers meet sellers to perform trades, and where prices adapt to supply and demand. This time-worn idea is certainly broadly correct, but reality is rather more intricate. At the heart of all markets lies a fundamental tension: buyers want to buy low and sellers want to sell high. Given these opposing objectives, how do market participants ever agree on a price at which to trade?

As the above dialogue illustrates, if a seller was allowed to increase the price whenever a buyer declared an interest to buy, then the price could reach a level so high that the buyer was no longer interested – and vice-versa. If traders always behaved in this way, then conducting even a single transaction would require a long and hard negotiation. Although this might be feasible if trades only occurred very infrequently, modern financial markets involve many thousands of transactions every single day. Therefore, finding a mechanism to conduct this process at scale, such that huge numbers of buyers and sellers can coordinate in real time, is an extremely complex problem.

Centuries of market activity have produced many possible solutions, each with their own benefits and drawbacks. Today, most markets implement an electronic, continuous-time double-auction mechanism based on the essential idea of a *limit order book (LOB)*, which we introduce in Chapter 3. However, as a brief glance at the financial press will confirm, ensuring market stability and “fair and orderly trading” is still elusive, and it remains unclear whether modern electronic markets are any less prone to serious problems than old-fashioned trading pits.

Given the tremendous impact of the digital revolution on society as a whole, why has the advent of computerised trading not solved these age-old problems once and for all? One possible answer is that trading intrinsically leads to instabilities. This viewpoint, which is increasingly supported by a growing body of empirical evidence, will lie at the very heart of our present journey into financial markets.

1.1 The Rules of Trading

We begin our discussion of financial markets by exploring the mechanisms that allow trading to take place, from old-style auctions to modern electronic markets.

1.1.1 The Walrasian Auction

As we emphasised above, markets are attempts to solve the seemingly impossible problem of allowing trades between buyers, who want to buy at an ever-lower price, and sellers, who want to sell at an ever-higher price. One possible way to solve this never-ending back-and-forth problem is to require that whenever a trader specifies a price, he or she makes a firm commitment to trade. This simple idea forms the heart of a classic market organisation called a **Walrasian auction**.[1]

In a Walrasian auction, traders communicate their buying or selling desires to an auctioneer, who collects and records this information. Buyers are invited to post bids that state the maximum price at which they are willing to buy, while sellers are invited to post offers that state the minimum price at which they are willing to sell. When posting bids or offers, each trader enters into a firm commitment to trade if he or she wins the auction. In the language of modern financial markets, this commitment is called **liquidity provision**.

A Walrasian auctioneer gathers these bids and offers into an order book. This order book describes the quantities that are available for purchase or sale at each specified price, as declared by the market participants. In a Walrasian auction, the auctioneer keeps the order book invisible, so that market participants cannot change their minds by observing what others are posting.

[1] The concept of the Walrasian auction first appeared as the design of French mathematical economist Léon Walras (1834–1910), as a gambit to understand how prices can reach their equilibrium such that supply matches demand.

At some instant of time, which can be set by the will of the auctioneer or decided at random, the auctioneer sets a transaction price p^* such that the total volume exchanged at that price is maximised. The transaction price p^* is the only price such that no buyers and no sellers remain unsatisfied after the transaction, in the sense that all remaining buyers have a limit price below p^* and all remaining sellers have a limit price above p^*. For a given price p, let $V_+(p)$ denote the total volume of buy orders at price p and let $V_-(p)$ denote the total volume of sell orders at price p. Formally, the **supply curve** $\mathscr{S}(p)$ is the total volume of sell orders with a price less than or equal to p, and the **demand curve** $\mathscr{D}(p)$ is the total volume of buy orders with a price greater than or equal to p. In a discrete setting, the supply and demand at a given price p can be written as

$$\mathscr{S}(p) = \sum_{p' \leq p} V_-(p'); \tag{1.1}$$

$$\mathscr{D}(p) = \sum_{p' \geq p} V_+(p'). \tag{1.2}$$

In words, $\mathscr{S}(p)$ represents the total volume that would be available for purchase by a buyer willing to buy at a price no greater than p. Similarly, $\mathscr{D}(p)$ represents the total volume that would be available for sale by a seller willing to sell at a price no less than p.

For a given price p, the total volume of shares exchanged is given by

$$Q(p) = \min\left[\mathscr{D}(p), \mathscr{S}(p)\right], \tag{1.3}$$

because the volume cannot exceed either the volume for purchase or the volume for sale. As is clear from Equations (1.1) and (1.2), $\mathscr{S}(p)$ is an increasing function of p and $\mathscr{D}(p)$ is a decreasing function of p (see Figure 1.1). Intuitively, for a transaction to occur, the price must be a compromise between the buyers' and sellers' wishes. If the price is too high, buyers will be disinterested; if the price is too low, sellers will be disinterested. Therefore, the auctioneer must find a compromise price p^* such that

$$Q(p^*) = \max_p \min\left[\mathscr{D}(p), \mathscr{S}(p)\right]. \tag{1.4}$$

Note that $Q(p^*) > 0$ if and only if at least one pair of buy and sell orders overlap, in the sense that the highest price offered among all buyers exceeds the lowest price offered among all sellers. Otherwise, $Q(p^*) = 0$, so p^* is ill-defined and no transactions take place.

If $Q(p^*) > 0$, and in the theoretical case where $\mathscr{D}(p)$ and $\mathscr{S}(p)$ are continuous in p, the maximum of $Q(p)$ must occur when supply and demand are equal. The price p^* must therefore satisfy the equality

$$\mathscr{D}(p^*) = \mathscr{S}(p^*). \tag{1.5}$$

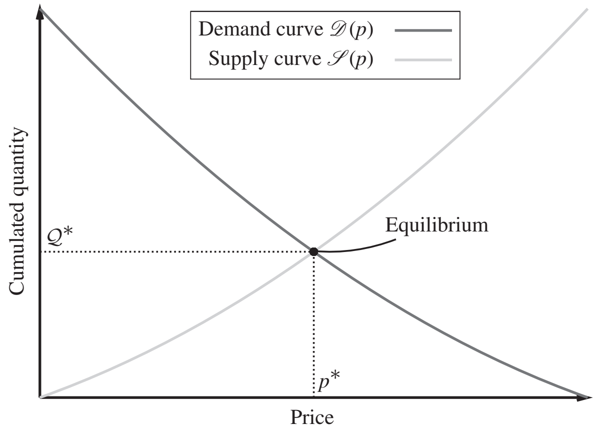

Figure 1.1. Illustration of the (increasing) supply curve $\mathscr{S}(p)$ and (decreasing) demand curve $\mathscr{D}(p)$ as a function of price p. According to Walras' law, the intersection of these curves defines the clearing price p^* and the total exchanged volume Q^*.

When the set of possible prices is restricted to a discrete price grid (as is the case in most markets), it is possible, and indeed common, that Equation (1.5) is not satisfied by the choice of p^* that maximises the total volume of trade. When this happens, some buyers or sellers remain unsatisfied. In this case, determining which of the possible buyers or sellers at a given price are able to trade requires a set of priority rules, which we discuss in detail in Section 3.2.1.

1.1.2 Market-Makers

In practice, Walrasian auctions are unsatisfactory for one major reason: they do not allow for any coordination between buyers and sellers. Therefore, a Walrasian auction might end in one of the following two scenarios:

(i) All buyers and sellers are unreasonably greedy, such that no buy and sell orders overlap and $Q(p^*) = 0$. In this case, there does not exist a price p^* at which any pair of traders is willing to trade, so no transactions can take place.
(ii) One side is unreasonably greedy while the other side is not. For example, the sellers might be unreasonably greedy while the buyers are not, which would result in a sudden increase in p^* and a very small transacted volume $Q(p^*)$. A similar situation can also arise if some external event temporarily causes the number of buyers to far exceed the number of sellers, or vice-versa.

Before the advent of electronic trading, the solution that most markets adopted to remedy this problem was to replace the Walrasian auctioneer (who seeks only to connect buyers and sellers at a reasonable price) with a special category of market participants called **market-makers** (or *specialists*). These special market participants were legally obliged to maintain a fair and orderly market, in exchange for some special privileges. To achieve their goals in the above auction setting, market-makers perform two tasks: quoting and clearing.

- **Quoting**: At all times, a market-maker must provide a *bid-price* b, with a corresponding *bid-volume* V_b, and an *ask-price* a, with a corresponding *ask-volume* V_a. As long as they are not modified by the market-maker, these quotes are binding, in the sense that the market-maker must execute any incoming sell order with volume less than or equal to V_b at the price b, and any incoming buy order with volume less than or equal to V_a at the price a.
- **Clearing**: Once buyers and sellers have submitted orders that specify a price and a volume, the market-maker decides on a price p^* that makes the number of unsatisfied orders as small as possible. Satisfied orders are cleared at price p^*.

A market-maker's quotes play the role of signalling a sensible price to the whole market, and thereby help other market participants to coordinate around a reasonable price. This point is illustrated by the retail foreign exchange (FX) market, in which retailers publicly display their buy and sell prices for a range of different currencies. At any time, if someone is uncertain about the value of the US dollar, then a simple way to gain a reasonably clear picture is to look at the buy and sell prices posted by a currency retailer. In this way, the retailer provides public information about their perception of the price of this asset.

Similarly to a Walrasian auction, buyers and sellers submit orders that specify a price and a volume and that accumulate in a public buy order book $V_+(p)$ and a public sell order book $V_-(p)$. Because market-makers post their quotes publicly, these new orders typically scatter around b (for buys) and a (for sells). At some time chosen by the market-maker, he or she computes the solution (or, in a discrete setting, the closest permissible discrete value) of the equation

$$\sum_{p' \geq p^*} V_+(p') = \sum_{p' \leq p^*} V_-(p'), \tag{1.6}$$

where $V_\pm$ also includes the market-maker's initial quotes at the bid and ask plus any additional buy or sell volume that he or she chooses to add just before the auction (e.g. to manage risk, to manage inventory, or to prevent large excursions of the price if demand temporarily outstrips supply, or vice-versa). If the market-maker is satisfied with p^* and the corresponding volume $Q(p^*)$, then the relevant transactions occur at this price. If $p^* > a$, this signals that the buy pressure was strong and that the market-maker chose not to add more sell volume to compensate. Note that $p^* > a$ can only happen if one or more buyers posted a limit price above the ask, and therefore have a strong urge to buy. In this case, it is likely that the market-maker will increase the quoted prices for the next auction, in an attempt to rebalance supply and demand. By symmetry, similar arguments follow for the case

$p^* < b$. If $b \leq p^* \leq a$, then the market is relatively balanced, in the sense that every transaction has occurred at a price better than or equal to the posted quotes.

In a market where all trade is facilitated by designated market-makers, participants can be partitioned into one of two categories: the market-makers, who offer trading opportunities to the rest of the market, and the other traders, who have the opportunity to accept them. Traditionally, these categories are named to reflect their members' contribution to the flow of liquidity: market-makers are called **liquidity providers**; the other traders are called **liquidity takers**.

1.1.3 Electronic Markets and Continuous-Time Double Auctions

Over time, markets have evolved away from appointing designated market-makers. Today, most liquid markets – including stocks, futures, and foreign exchange – are electronic, and adopt a **continuous-time double auction** mechanism using a *limit order book* (LOB), in which a transaction occurs whenever a buyer and a seller agree on a price.

LOBs require neither a Walrasian auctioneer nor designated market-makers to facilitate trade.[2] LOBs are updated in real time, and are observable by all traders. They therefore form an important part of the information sets used by traders when deciding how to act. As we discuss at many points throughout the book, the intertwined, interacting sequences of order-flow events generate transactions and price changes in an LOB. Analysing these sequences of events provides a direct route to quantifying and understanding price dynamics "from the bottom up", and therefore lies at the very heart of modern market microstructure.

1.2 The Ecology of Financial Markets

For a trade to take place, a buyer and a seller must agree on a price. Assuming that both parties do so willingly, it seems reasonable to insist that any trader conducting a trade must do so without feeling regrets. However, since assets are quoted every day (and nowadays are even quoted continuously during the day), at least one of the counterparties to a trade will have regrets as soon as the price moves, because he or she could have obtained a better price by waiting. So why do traders trade at all? Answering this puzzling question first requires a detailed understanding of the **ecology** of financial markets.

1.2.1 Trades and Information

Most attempts at explaining why market participants (other than market-makers) trade distinguish between two different types of activity:

[2] Interestingly, however, a category of market participants, who are still called market-makers, perform the useful function of providing some of the liquidity in an LOB. We return to this discussion in Section 1.2.4.

- **Informed trades** are attributed to sophisticated traders with information about the future price of an asset, which these traders buy or sell to eke out a profit.
- **Uninformed trades** are attributed either to unsophisticated traders with no access to (or the inability to correctly process) information, or to liquidity trades (e.g. trades triggered by a need for immediate cash, a need to reduce portfolio risk or a need to offload an inventory imbalance). These trades are often called *noise trades,* because from an outside perspective they seem to occur at random: they do not correlate with future price changes and they are not profitable on average.

In some cases, classifying a trade as informed is relatively straightforward. One clear example is **insider trading**. For example, consider a company with a stock price of \$95 at some time t. If an insider hears that a large corporation seeks to make a takeover bid by offering \$100 per share at some future time $t+T$, and that this information will not become public until time $t+T$, then the insider could buy some shares at the current value of \$95 and realise a near-certain profit of \$5 per share at time $t+T$. Deterministic arbitrage, such as exploiting the mispricing of derivative products, could also fall into the category of informed trading.

In most cases, however, this seemingly intuitive partitioning of trades as informed or uninformed suffers from a problem: information is difficult to measure – and even to define. For example, is an observation of another trade itself information? If so, how much? And how strongly might this impact subsequent market activity? For most large-cap US stocks, about 0.5% of the market capitalisation changes hands every day. Given that insider trading is prohibited by law, can a significant fraction of this vast market activity really be attributed to informed trades of the type described above?

1.2.2 Statistical Information

A less extreme (and perhaps more realistic) view of information can be framed in statistical terms: an informed trade can be defined as a trade whose ex-ante expected profit over some time horizon T is strictly positive, even after including all costs. In other words, information is tantamount to some ability to predict future price changes, *whatever the reasons for these changes*.

When speaking about statistical information, it is customary to decompose trading strategies into two categories:

- **Fundamental analysis** attempts to decide whether an asset is over-priced or under-priced. These strategies seek to use quantitative metrics, such as price-to-earnings ratios, dividend yields, macroeconomic indicators, information on the health and growth of a specific company, and even more qualitative indicators such as the charisma of the CEO.

- **Quantitative analysis** (or technical analysis) attempts to predict price movements by identifying price patterns, some of which are based on solid statistical evidence (such as price trends or mean-reversion), while others are less so (such as chartists' "head and shoulders" distributions, or price "support" and "resistance" levels).

Empirical data suggests that some of these signals are indeed correlated with future price changes, but that this correlation is very weak, in the sense that the dispersion of future price changes is much larger than the mean predicted price change. The time scale of these strategies is also extremely heterogeneous, and spans from months (or even years) for traditional long-only pension funds to just a few minutes (or even seconds) for some intra-day strategies.

In summary, we should expect that informed trades are either very rare and very successful (but unlawful!), like the insider example, or are more common but with weak information content and a low degree of individual success, like statistical arbitrage. Given that very successful trades occur rarely, if many traders really are informed, then the vast majority of such trades must be of the latter type. These trades can be based on information from any combination of a large number of diverse sources, each of which typically provides weak insights into future prices. In practice, this diversity of information signals is reflected in the diversity of market participants, who have different trading strategies, motives and time horizons – from long-term pension funds to day traders, hedgers, and even high-frequency trading (HFT).

1.2.3 To Trade, or Not To Trade?

Despite the prominence of arbitrage strategies, prices are notoriously hard to predict. In fact, as we will illustrate in Chapter 2, prices are close to being martingales, in the sense that the best estimate of the future price is simply the current price. This implies that the real information contained in the signals used by traders or investors is extremely weak. Given that trading entails considerable costs (such as brokerage fees, transaction fees and price impact, which we will discuss in detail throughout the book), the overall nagging feeling is that speculative traders trade too much, probably as a result of overestimating the predictive power of their signals and underestimating the costs of doing so.[3]

Why do traders behave in this way? One possible explanation is that they are blinded by the prospect of large gains and fail to recognise the true costs of their actions. Another is that it is extremely difficult to separate skill from luck in trading performance: even when trading with no information, a lucky trade can lead to a substantial profit. For example, if asset prices followed a simple symmetric random

[3] See, e.g., Odean, T. (1999). Do investors trade too much? *American Economic Review*, 89, 1279–1298.

walk, then in the absence of trading costs, any trade initiated at time t would have a 50% chance of being profitable at time $t+T$, whatever T!

Using numbers from real markets illustrates that evaluating real trading strategies is very difficult. Given a stock with an annual volatility of 15% and a reported annual return of 5%, it would take almost $T = 9$ years to test whether the actual return was statistically significantly different from zero at the one-sigma level.[4] Therefore, it can take a very long time to notice that a seemingly lucrative trading strategy is actually flawed, or vice-versa.

In summary, classifying trades is much less straightforward than the classical "informed-versus-uninformed" dichotomy might suggest. Although truly informed trades (such as insider trades) and truly uninformed trades (such as hedging or portfolio-balancing trades) likely both exist, between these two extremes lies a broad spectrum of other trades based on some sort of information that is difficult to define and even harder to measure. This lack of high-quality information also provides a possible explanation for why the ecology of modern financial markets is so complex, and contains many different types of traders seeking to earn profits on many different time horizons, despite the "no-trade" situation that we described earlier in this section. This important observation will be an overarching theme throughout the book.

1.2.4 Liquidity Providers: The Modern-Day Market-Makers

Most strategies and trading techniques attempt to earn a profit by forecasting the future price of an asset, then buying or selling it accordingly. But much as in the old days, financial markets can only function if some participants commit to providing liquidity. In the ecology of financial markets, liquidity providers offer to *both* purchase and sell an asset, and seek to earn the difference between the buy and sell price. Traders who implement this strategy in modern markets are still called "market-makers". In contrast to those in older markets, however, modern market-makers are not specifically designated market participants with special privileges. Instead, modern markets enable anyone to act as a market-maker by offering liquidity to other market participants.

Market-makers typically aim to keep their net inventory as close to zero as possible, so as not to bear the risk of the asset's price going up or down. Their goal is to earn a profit by buying low (at their bid-price b) and selling high (at their ask-price a), and therefore earning the **bid–ask spread**

$$s := a - b \tag{1.7}$$

for each round-trip trade.

[4] The order of magnitude of the time required is given by the square of the ratio of these two numbers: $(15/5)^2 = 9$ years.

An important consequence of the widespread use of LOBs in modern financial markets is a blurring of the lines between liquidity providers and liquidity takers. For example, if one market-maker noticed another market-maker offering to trade at a very attractive price, it would be illogical not to transact against this price, and to therefore act as a liquidity taker. Indeed, many successful high-frequency market-makers also implement sophisticated short-term prediction tools and exploit profitable high-frequency signals. Similarly, market participants who usually act as liquidity takers might instead choose to provide liquidity if they notice the bid–ask spread to be particularly wide.

Despite this emerging complexity of modern markets, the simple separation of market participants into two classes – speculators (or liquidity takers), who typically trade at medium-to-low frequencies, and market-makers, who typically trade at high frequencies – is a useful first step towards understanding the basic ecology of financial markets. As we will illustrate throughout the book, considering the interactions and tensions between these two groups provides useful insights into the origins of many interesting phenomena.

1.3 The Risks of Market-Making

Throughout this chapter, we have gradually built up a picture of the basic ecology of modern financial markets: market-makers offer opportunities to buy and/or sell, with the aim of profiting from round-trip trades, while speculators buy or sell assets, with the aim of profiting from subsequent price changes.

Based on this simple picture, it seems that market-makers have a much more favourable position than speculators. If speculators make incorrect predictions about future price moves, then they will experience losses, but they will still trade with market-makers, who will therefore still conduct round-trip trades. In this simplistic picture, speculators bear the risk of incorrect predictions, while market-makers seemingly always make a profit from the bid–ask spread. Is it really the case that market-makers can earn risk-free profits?

1.3.1 Adverse Selection

The simple answer to the last question is: no. Market-makers also experience several different types of risk. Perhaps the most important is **adverse selection** (also called the "winner's curse" effect). Adverse selection results from the fact that market-makers must post binding quotes, which can be "picked off" by more informed traders who see an opportunity to buy low or to sell high. This informational asymmetry is a fundamental concern for market-makers and is the final piece of the financial ecosystem that we consider in this chapter.

For a market-maker, the core question (to which we will return in Chapter 16) is how to choose the values of b and a. If the values of b and a remain constant at all times, then the market-maker always earns a profit of s for each round-trip (buy and sell) trade. All else being equal, the larger the value of s, the larger the profit a market-maker earns per trade. However, the larger the value of s, the less attractive a market-maker's buy and sell prices are to liquidity takers.

In situations where several different market-makers are competing, if one market-maker tries to charge too wide a spread, then another market-maker will simply undercut these prices by stepping in and offering better quotes. Most modern financial markets are indeed highly competitive, which prevents the spread from becoming too large. But why does this competition not simply drive s to zero?

Consider a market-maker trading stocks of a given company by offering a buy price of $b = \$54.50$ and a sell price of $a = \$55.50$. If the buy and sell order flow generated by liquidity takers was approximately balanced, then the market-maker would earn $s = \$1.00$ per round-trip trade. If, however, an insider knows that the given company is about to announce a drop in profits, they will revise their private valuations of the stock downwards, say to \$50.00. If the market-maker continues to offer the same quotes, then he or she would experience a huge influx of sell orders from insiders, who regard selling the stock at \$54.50 to be extremely attractive. The market-maker would therefore quickly accumulate a large net buy position by purchasing more and more stocks at the price \$54.50, which will likely be worth much less soon after, generating a huge loss. This is adverse selection.

Although the above example uses a strongly imbalanced order flow to provide a simple illustration, market-makers face the same problem even when facing more moderate imbalances in order flow. If a market-maker holds b and a constant in the face of an imbalanced order flow, then he or she will accumulate a large net position in a short time, with a high probability of being on the wrong side of the trade. As we will discuss at several times throughout the book, market-makers compensate for this potential loss by charging a non-zero spread – even in situations where market-making is fiercely competitive (see Chapter 17).

1.3.2 Price Impact

To mitigate the risk of being adversely selected, market-makers must update their values of b and a to respond to their observations of order-flow imbalance. If a market-maker receives many more buy orders than sell orders, then he or she can attempt to reduce this order-flow imbalance by increasing the ask-price a (to dissuade future buyers), increasing the bid-price b (to encourage future sellers), or both. Similarly, if a market-maker receives many more sell orders than buy orders, then he or she can attempt to reduce this order-flow imbalance by decreasing b, decreasing a, or both. An important consequence of this fact is that trades have **price impact**: on average, the arrival of a buy trade causes prices to rise and the arrival of a sell trade causes prices to fall. This is precisely what the spread s compensates for.

More formally, let

$$m := \frac{a+b}{2} \tag{1.8}$$

denote the **mid-price**, and let $\mathbb{E}[m_\infty]$ denote the expected future value of m. Since a (possibly small) fraction of trades are informed, then if we observe a buy trade, we should expect the future value of m to be greater than the current value of m. Similarly, if we observe a sell trade, then we should expect the future value of m to be less than the current value of m. To express this mathematically, let ε denote the **sign of the trade**, such that $\varepsilon = +1$ for a buy trade and $\varepsilon = -1$ for a sell trade. Let $\mathcal{R}_\infty$ denote the expected long-term impact of a trade,

$$\mathcal{R}_\infty := \mathbb{E}\left[\varepsilon \cdot (m_\infty - m)\right]. \tag{1.9}$$

Because on average the arrival of a buy trade causes m to increase and the arrival of a sell trade causes m to decrease, it follows from the definition of ε that $\mathcal{R}_\infty > 0$.

The concept of price impact is central to the understanding of financial markets, and will be discussed many times throughout this book. The above introduction is highly simplified, but it already provides a quantitative setting in which to consider the problem faced by market-makers. Recall that if b and a remain constant, then a market-maker earns $s/2$ for each of the two legs of a round-trip trade. Since on average the price moves in an adverse direction by an amount $\mathcal{R}_\infty > 0$, a market-maker's net average profit per trade (in this highly simplified framework) is given by $s/2 - \mathcal{R}_\infty$. We provide a much more detailed version of this argument in Chapter 17.

Market-making is thus only profitable if $s/2 > \mathcal{R}_\infty$, i.e. if the mean profit per trade is larger than the associated price impact. The larger the value of $\mathcal{R}_\infty$, the less profitable market-making becomes. The ideal situation for a market-maker is that $\mathcal{R}_\infty = 0$, which occurs when the sign of each trade is uncorrelated with future price moves – i.e. when trades have no price impact.

1.3.3 Skewness

Unfortunately for market-makers, the distribution of price changes after a trade is very broad, with a heavy tail in the direction of the trade. In other words, whereas most trades contain relatively little information and are therefore innocuous for market-makers, some rare trades are triggered by highly informed market participants, such as our insider trader from Section 1.2.1. These traders correctly

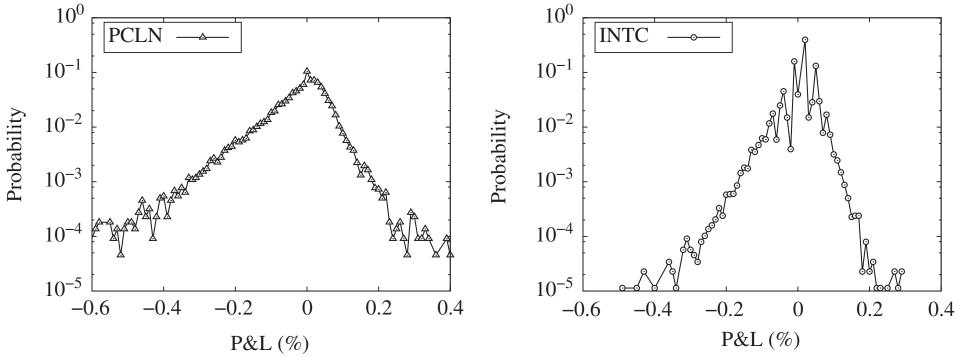

Figure 1.2. Empirical distributions of the P&L of a round-trip trade for (left panel) PCLN and (right panel) INTC. The negative skewness of the distributions is a specific feature of the risk borne by market-makers.

anticipate large price jumps before market-makers are able to update their quotes, and may therefore cause market-makers huge losses (see Figure 1.2).

Let us consider a simple example. Assume that with probability ϕ, an arriving trade is informed, and correctly predicts a price jump of size $\pm J$. Otherwise (with probability $1-\phi$), the arriving trade is uninformed, and does not predict a future price change. Under the assumption that all trades are independent, it follows that $\mathcal{R}_\infty = \phi J$. To break even on average, a market-maker facing this arriving order flow would need to set $s = 2\phi J$.

The variance σ^2_{MM} of the market-maker's gain per trade is then given by

$$\sigma^2_{\mathrm{MM}} = (1-\phi)\times 0 + \phi \times J^2. \tag{1.10}$$

Similarly, its **skewness** ς_{MM} (which measures the asymmetry of the distribution of the gains) is[5]

$$\varsigma_{\mathrm{MM}} = \frac{(1-\phi)\times 0 - \phi \times J^3}{\sigma^3_{\mathrm{MM}}} = -\phi^{-\frac{1}{2}}. \tag{1.11}$$

Hence, as the probability ϕ that an arriving trade is informed decreases to zero, the relative precision on the average impact decreases as $\sqrt{\phi}$, but the skewness diverges (negatively) as $-1/\sqrt{\phi}$.

This simple calculation illustrates that market-making is akin to selling insurance. Although profitable on average, this strategy may generate enormous losses in the presence of informed trades.

This very simple argument illustrates why liquidity is fragile and can disappear quickly during times of market turbulence: the possible down-side risks of large losses are huge, so liquidity providers are quick to reduce the amount of liquidity that they offer for purchase or sale if they perceive this risk to be too high. Therefore, even in the presence of liquidity providers, organising markets to ensure fair and orderly trading remains a difficult task.

[5] The skewness of a random variable with zero mean is usually defined by the ratio of its third moment to its variance, raised to the power 3/2. Other skewness definitions, less sensitive to outliers, are often used for financial data.

1.4 The Liquidity Game

Understanding the delicate dance between buyers and sellers in financial markets is clearly difficult, but at least one thing is clear: whether informed, uninformed, or even misinformed, all traders want to get the best possible price for what they buy or sell.

An important empirical fact that is crucial to understanding how markets operate is that even "highly liquid" markets are in fact not that liquid after all. Take, for example, a US large-cap stock, such as PCLN (see the right panel of Figure 1.3). Trading for this stock is extremely frequent, with each day containing several thousands of trades that collectively add up to a daily traded volume of roughly 0.5% of the stock's total market capitalisation. At any given time, however, the volume of the stock available for purchase or sale at the best quotes is quite small, and is typically only of the order of about 10^{-4} of the stock's market cap. Liquidity is slightly more plentiful for large-tick stocks, such as CSCO (see the left panel of Figure 1.3), but it is still small compared to the daily traded volume. This phenomenon is also apparent in other markets, such as foreign exchange (FX) and futures, in which trading is even more frantic.

Why is there so little total volume offered for purchase or sale at any point in time? The reason is precisely what we discussed when summarising the difficulties of market-making: adverse selection. Liquidity providers bear the risk of being picked off by an informed trader. To minimise this risk, and perhaps even to bait informed traders and to out-guess their intentions, liquidity providers only offer relatively small quantities for trade. This creates a kind of hide-and-seek game in financial markets: buyers and sellers want to trade, but both avoid showing their hands and revealing their true intentions. As a result, markets operate in a regime of small **revealed liquidity** but large **latent liquidity**. This observation leads to many empirical regularities that we will discuss in this book.

The scarcity of available liquidity has an immediate and important consequence: *large trades must be fragmented*. More precisely, market participants who wish to buy or sell large volumes of a given asset must chop up their orders into smaller pieces, and execute them incrementally over time. For example, it is not uncommon for an investment fund to want to buy 1% (or more) of a company's total market capitalisation. Using the numbers from earlier in this chapter, buying 1% of PCLN would require of the order of 100 or even 1000 individual trades and would correspond to trading twice the typical volume for a whole day. To avoid strongly impacting the market and thus paying a higher price, such a transaction would have to be executed gradually over several days. Therefore, even an inside trader with clear information about the likely future price of an asset cannot use all of this information immediately, lest he or she scares the market and gives away the private information (see Chapter 15). Instead, traders must optimise a trading

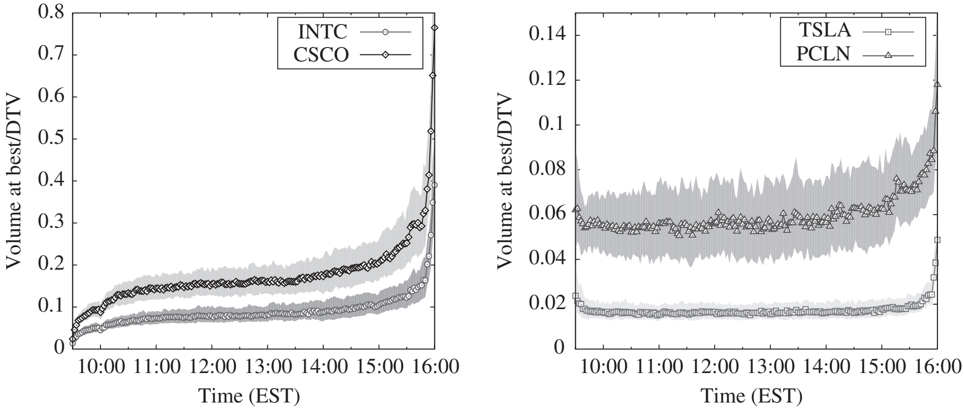

Figure 1.3. Total volume of active orders at the best quotes, normalised by the daily traded volume (during a full trading day), for (left panel) INTC and CSCO, two typical large-tick stocks, and (right panel) TSLA and PCLN, two typical small-tick stocks. The markers show the average over all trading days and the shaded regions show the corresponding lower and upper quartiles.

schedule that seeks to attain the best possible execution price – a topic that we will address in Chapter 21.

From a conceptual viewpoint, the most important conclusion from our discussions in this chapter is that *prices cannot be in equilibrium*, in the traditional sense that supply and demand are matched at some instant in time. Since transactions must be fragmented, the instantaneous traded volume is much smaller than the underlying "true" supply and demand waiting to be matched. Part of the imbalance is necessarily latent, and can only be *slowly* digested by markets. At best, the notion of equilibrium prices can only make sense on longer time scales, but on such time scales, the information set evolves further. In summary, the equilibrium price is an elusive concept – and an ever-moving target.

Take-Home Messages

(i) Organising markets to facilitate fair and orderly trade is a very difficult task.
(ii) Historically, market stability was ensured by designated market-makers, who quoted bid- and ask-prices and cleared matched orders (possibly including their own) in one large auction. Market-making was a risky but potentially lucrative business.
(iii) Most of today's markets implement a continuous-time double-auction mechanism, removing the need for a designated market-maker. In

practice, this mechanism is often implemented using a limit order book (LOB).

(iv) LOBs allow traders to either provide firm trading opportunities to the rest of the market (liquidity provision) or to accept such trading opportunities (liquidity taking).

(v) Today's liquidity providers play a similar role to the market-makers of yesterday. In modern markets, however, any market participant can choose to act as a liquidity provider or a liquidity taker, which makes these activities highly competitive.

(vi) Liquidity providers run the risk of being "picked-off" by better-informed traders who correctly anticipate future price moves. This is known as adverse selection.

(vii) Because of adverse selection, investors typically do not display their full intentions publicly. The revealed liquidity in an LOB is therefore only a small fraction of all trading intentions.

1.5 Further Reading

General

Lyons, R. (2001). *The microstructure approach to foreign exchange rate*. MIT Press.
Biais, B., Glosten, L., & Spatt, C. (2005). Market microstructure: A survey of microfoundations, empirical results, and policy implications. *Journal of Financial Markets*, 8(2), 217–264.
Hasbrouck, J. (2007). *Empirical market microstructure: The institutions, economics, and econometrics of securities trading*. Oxford University Press.
Amihud, Y., Mendelson, H., & Pedersen, L. H. (2012). *Market liquidity: Asset pricing, risk, and crises*. Cambridge University Press. https://www.gov.uk/government/collections/future-of-computer-trading.
Foucault, T., Pagano, M., & Röell, A. (2013). *Market liquidity: Theory, evidence, and policy*. Oxford University Press.
Lehalle, C. A., & Laruelle, S. (2013). *Market microstructure in practice*. World Scientific.
Easley, D., Prado, M. L. D., & O'Hara, M. (2014). *High-frequency trading: New realities for traders, markets and regulators*. Risk Books.
Laumonier, A. "*6*" and "*5*", Zones Sensibles Edt.

Organisation of Markets, Market-Making and Bid-Ask Spreads

Bagehot, W. (1971). The only game in town. *Financial Analysts Journal*, 27(2), 12–14.
Cohen, K. J., Conroy, R. M., & Maier, S. F. (1985). Order flow and the quality of the market. *Market-making and the changing structure of the securities industry* (pp. 93–110). Lexington Books.
Glosten, L. R., & Milgrom, P. R. (1985). Bid, ask and transaction prices in a specialist market with heterogeneously informed traders. *Journal of Financial Economics*, 14(1), 71–100.

Domowitz, I., & Wang, J. (1994). Auctions as algorithms: Computerized trade execution and price discovery. *Journal of Economic Dynamics and Control*, 18(1), 29–60.

Handa, P., & Schwartz, R. A. (1996). Limit order trading. *The Journal of Finance*, 51(5), 1835–1861.

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

Jones, C. M. (2002). A century of stock market liquidity and trading costs. https://ssrn.com/abstract=313681.

See also Section 16.5.

The Ecology of Financial Markets

Handa, P., Schwartz, R. A., & Tiwari, A. (1998). The ecology of an order-driven market. *The Journal of Portfolio Management*, 24(2), 47–55.

Lux, T., & Marchesi, M. (1999). Scaling and criticality in a stochastic multi-agent model of a financial market. *Nature*, 397(6719), 498–500.

Farmer, J. D. (2002). Market force, ecology and evolution. *Industrial and Corporate Change*, 11(5), 895–953.

May, R. M., Levin, S. A., & Sugihara, G. (2008). Complex systems: Ecology for bankers. *Nature*, 451(7181), 893–895.

Bouchaud, J. P., Farmer, J. D., & Lillo, F. (2009). How markets slowly digest changes in supply and demand. In Hens, T. & Schenke-Hoppe, K. R. (Eds.), *Handbook of financial markets: Dynamics and evolution*. North-Holland, Elsevier.

Farmer, J. D., & Skouras, S. (2013). An ecological perspective on the future of computer trading. *Quantitative Finance*, 13(3), 325–346.

Jones, C. M. (2013). What do we know about high-frequency trading? https://ssrn.com/abstract=2236201.

Biais, B., & Foucault, T. (2014). HFT and market quality. *Bankers, Markets & Investors*, 128, 5–19.

Bouchaud, J. P., & Challet, D. (2017). Why have asset price properties changed so little in 200 years. In Abergel, F. et al. (Eds.), *Econophysics and sociophysics: Recent progress and future directions* (pp. 3–17). Springer.

https://en.wikipedia.org/wiki/Trading_strategy.

Noise Trading, Overconfidence and Poor Performance

Black, F. (1986). Noise. *The Journal of Finance*, 41(3), 528–543.

Shleifer, A., & Summers, L. H. (1990). The noise trader approach to finance. *The Journal of Economic Perspectives*, 4(2), 19–33.

Barber, B. M., & Odean, T. (1999). Do investors trade too much? *American Economic Review*, 89(5), 1279–1298.

Barber, B. M., & Odean, T. (2000). Trading is hazardous to your wealth: The common stock investment performance of individual investors. *The Journal of Finance*, 55(2), 773–806.

Barber, B. M., & Odean, T. (2001). Boys will be boys: Gender, overconfidence, and common stock investment. *The Quarterly Journal of Economics*, 116(1), 261–292.

The Case for Index-Fund Investing: https://personal.vanguard.com/pdf/ISGIDX.pdf.

See also Section 20.5.