4

Empirical Properties of Limit Order Books

When my information changes, I alter my conclusions. What do you do, sir?

(John Maynard Keynes)

Many LOBs record comprehensive digital transcriptions of their participants' submissions and cancellations of orders. These event-by-event records describe the temporal evolution of visible liquidity at the microscopic level of detail. During the past two decades, a vast number of empirical studies have analysed LOB data to address a wide array of questions regarding the high-frequency activity and price formation in financial markets. This work has served to illustrate many important aspects of the complex interplay between order flow, transactions and price changes. In recent years, several survey articles and books (see the list of references at the end of this chapter) have established new microstructural stylised facts and have highlighted both similarities and differences between different assets and different market organisations.

Although the breadth of such empirical work is substantial, an important consideration when reviewing previous studies of LOBs is that the high-frequency actors in financial markets have evolved rapidly in recent years. Market-making strategies, trading algorithms, and even the rules governing trade have changed over time, so old empirical observations may not accurately describe current LOB activity. Therefore, maintaining a detailed and up-to-date understanding of modern financial markets requires empirical analysis of recent, high-quality LOB data.

In this chapter, we present some up-to-date statistical results regarding order flow and LOB state for a collection of stocks that we study throughout the book:[1] PCLN (Priceline Group Inc.), TSLA (Tesla Motors Inc.), CSCO (Cisco Systems Inc.) and INTC (Intel Corp.). The stocks CSCO and INTC are large-tick stocks (i.e. with a spread close to the minimum value, $s \approx \vartheta$) and the stocks PCLN and TSLA are small-tick stocks (i.e. with a spread much larger than the tick, $s \gg \vartheta$)

[1] For a detailed discussion of our data and sample, see Appendix A.1.

(see Section 4.8 for more precise statements). As we will see at several points throughout this chapter, the statistical properties of large-tick stocks and small-tick stocks can be radically different.

4.1 Summary Statistics

Table 4.1 lists a range of summary statistics that describe the four stocks' aggregate activity and spread. A few notable features, common to all stocks, should be highlighted:

(i) The daily turnover (as a fraction of market capitalisation) is of the order of 0.5%. This result in fact holds across a wide selection of international stocks. This number has roughly doubled from 1995 to 2015.
(ii) The total volume displayed in the LOB within 1% of the mid-price (roughly half the daily volatility) is between 1% and 3% of the daily traded volume.
(iii) The activity at the best quotes takes place at a sub-second time scale (sometimes milliseconds), with much more activity for large-tick stocks than for small-tick stocks. This reflects the importance of queue position for large-tick stocks (see Section 21.4).
(iv) The number of trade-through market orders (i.e. orders that match at several different prices and therefore walk up the order book) is on the order of a few percent for small-tick stocks, and a few per thousand for large-tick stocks.

Point (ii) shows that the total outstanding volume in the LOB at any instant of time is only a small fraction of the total daily activity. This is a fundamental observation that has important consequences for trading, as we will discuss in Section 10.5.

4.2 Intra-day Patterns

Many properties of order flow and LOB state follow strong **intra-day patterns**, which we illustrate in this section. It is important to bear these patterns in mind when considering other order-flow and LOB statistics. For example, the distribution of the total volume at the best quotes will look very different when pooling together different times of day than when restricting observations to a specific time window. For the statistics that we calculate throughout the remainder of this chapter, we restrict our attention to market activity between 10:30 and 15:00 local time, where most quantities have an approximately flat average profile.

Table 4.1. *Summary statistics of aggregate activity for PCLN, TSLA, CSCO and INTC between 10:30 and 15:00 on all trading days during 2015. To remove effects stemming from the opening and closing auctions, we disregard the first and last hour of the regular trading time, except for variables marked with a dagger†. Cancellations include total order deletions and partial cancellations. We calculate the total market capitalisation, the daily traded volume on NASDAQ and the NASDAQ market share by using data from Compustat for the full trading days.*

		PCLN	TSLA	CSCO	INTC
Average share price (dollars)		1219.0	229.9	27.86	32.15
Average quoted spread† (dollars)		1.366	0.180	0.0105	0.0103
Average spread before transactions (dollars)		0.946	0.130	0.0109	0.0108
Fraction of MO that match at worse prices than the best quote		0.038	0.021	0.0017	0.0026
Fraction of MO that match a hidden order inside the spread		0.050	0.038	0.0311	0.0337
Mean total volume at best quotes (dollars $\times 10^3$)		129.6	38.35	248.4	163.6
Mean daily traded volume on NASDAQ† (dollars $\times 10^6$)		209.8	204.0	142.1	181.4
Mean fraction of daily traded volume that is hidden		0.340	0.241	0.086	0.074
Mean total volume of active orders within 1% of mid-price (dollars $\times 10^6$)		2.701	1.690	3.885	3.342
Average stock-specific market share of NASDAQ†		50.2%	36.0%	39.2%	39.2%
Average total market capitalisation† (dollars $\times 10^9$)		61.15	30.50	143.2	149.6
Average daily number (Best Quote Orders $\times 10^3$)	Limit Orders	8.959	15.83	84.90	116.5
	Market Orders	1.342	3.932	3.123	4.394
	Cancellations	5.873	10.44	75.584	101.2
Mean Inter-Arrival Time of Best Quote Orders (seconds)	Limit Orders	2.393	1.208	0.220	0.158
	Market Orders	14.264	4.831	5.895	4.230
	Cancellations	3.620	1.812	0.247	0.183
Median Inter-Arrival Time of Best Quote Orders (seconds)	Limit Orders	0.136	0.063	0.00013	0.00005
	Market Orders	0.081	0.094	0.22063	0.15983
	Cancellations	0.110	0.051	0.00038	0.00018
Mean Size of Best Quote Orders (dollars $\times 10^3$)	Limit Orders	82.12	23.63	15.23	10.44
	Market Orders	75.88	25.95	22.23	20.37
	Cancellations	80.26	21.54	14.18	9.769

4.2.1 Market Activity

In Figure 4.1, we plot the mean total volume of executed market orders (MO) and submitted limit orders (LO) during two-minute intervals throughout the trading day. In both cases, the activity exhibits a **U-shape profile**, or better a "J-shape profile" with asymmetric peaks at the beginning and end of the day and a minimum at around midday. Activity during the busiest periods is about four times greater than activity during the quietest periods.

The J-shaped profile is similar across all stocks, independently of the tick size and market capitalisation. Why should this be so? One possible explanation is that the intense spike shortly after the market opens is caused by company news revealed during the previous overnight period, when markets are closed. The spike at the close is possibly due to traders who previously hoped to get a better deal having to speed up to finish their trades for the day. Another possible explanation is that individual traders implement non-uniform execution patterns to minimise their price impact (see Section 21.2).

Figure 4.2 shows how the volume at the best quotes varies according to the time of day. Here, we see a striking difference between large-tick and small-tick stocks: whereas small-tick stocks exhibit a J-shaped profile similar to those in Figure 4.1, large-tick stocks show a lack of liquidity in the early minutes of trading, followed by a slow increase throughout the day, then a final, steep rise shortly before the end of the trading day. The volumes available just before close are about 20 times larger than those just after open.

In each of Figures 4.1 and 4.2, the statistics are (within statistical errors) symmetric between buy orders and sell orders, so we only present the average of the two. Of course, this does not imply that this symmetry holds for a given time on a given day, where order imbalance can be locally strong, but rather that the symmetry emerges when averaging market activity over long times.

4.2.2 Bid-Ask Spread

Figure 4.2 also shows the intra-day average values of the bid–ask spread $s(t) = a(t) - b(t)$. The spread narrows throughout the trading day, quite markedly for small-tick stocks and much more mildly for large-tick stocks. As we will discuss in Chapter 16, the narrowing of the bid–ask spread is often interpreted as a reduction of the adverse selection faced by liquidity providers, as overnight news gets progressively digested by market participants.

In summary, Figure 4.2 paints an interesting picture that suggests that liquidity is much more scarce at the open than it is at the close. Compared to submitting a limit order, the average cost of immediacy (i.e. of submitting a market order) is much larger in the morning than it is later in the day: not only is the spread wider,

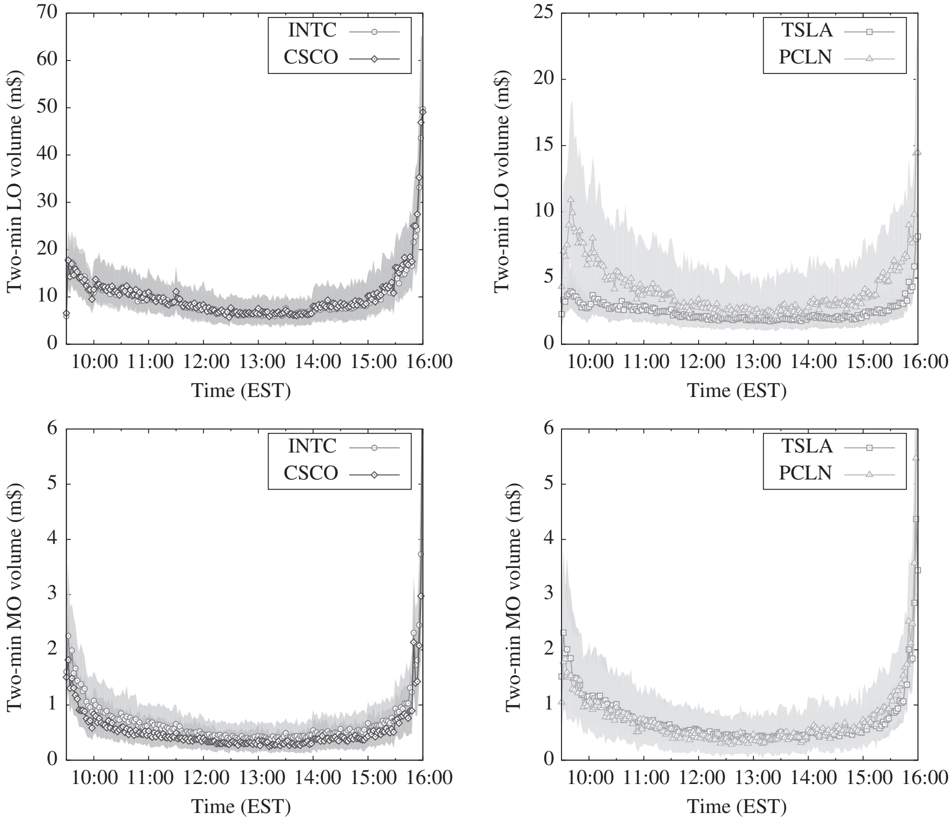

Figure 4.1. Total volume of (top panels) limit order and (bottom panels) market order arrivals during each two-minute interval throughout the trading day, for (left panels) the large-tick stocks INTC and CSCO, and (right panels) the small-tick stocks TSLA and PCLN. The markers show the average over all trading days and the shaded regions show the corresponding lower and upper quartiles.

but market orders need to penetrate deeper into the LOB to find their requested volume.

4.3 The Spread Distribution

Figure 4.3 shows the **spread distribution** for a selection of different stocks. As the figure illustrates, the distributions are very different for small-tick and large-tick stocks. For large-tick stocks, the distribution is sharply peaked at the minimum spread (which is equal to one tick, ϑ), with rare moments when the spread opens to 2ϑ or, in extreme cases, 3ϑ (or larger). For small-tick stocks, by contrast, the distribution is much wider, and the value of $s(t)$ can vary between one tick and several tens of ticks. The upper tail of the distribution decays approximately like an exponential.

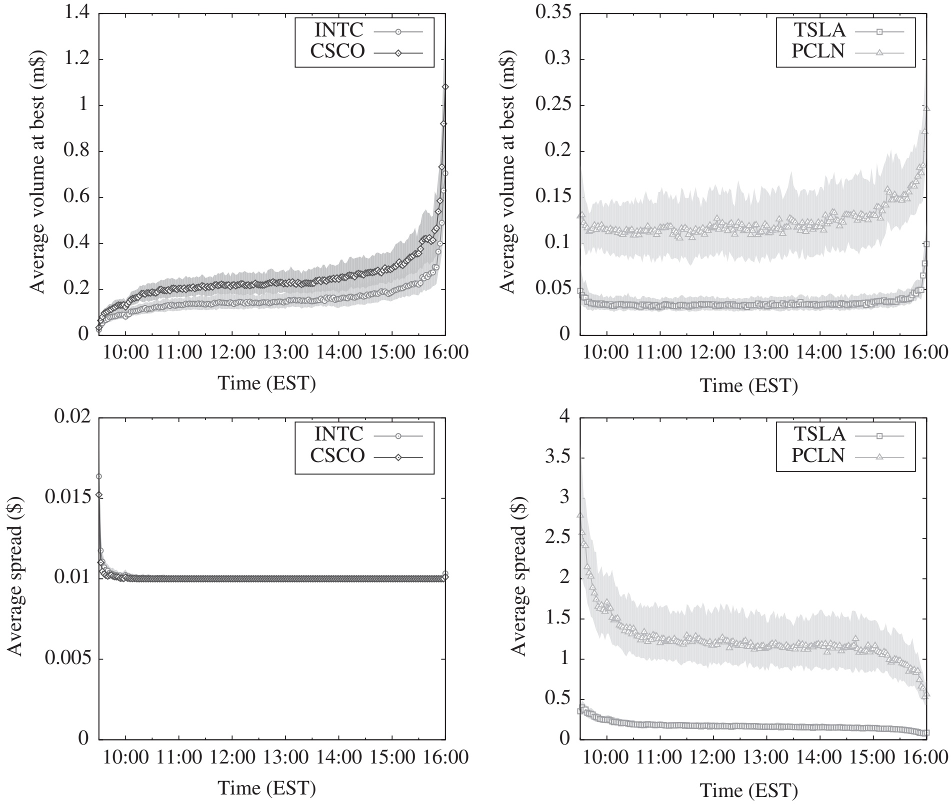

Figure 4.2. (Top panels) Average total volume of active orders at the best quotes and (bottom panels) average bid–ask spread, during each two-minute interval throughout the trading day for (left panels) the large-tick stocks INTC and CSCO, and (right panels) the small-tick stocks TSLA and PCLN. The markers show the average over all trading days and the shaded regions show the corresponding lower and upper quartiles.

Note that the bid–ask spread distribution can be measured in several different ways, such as at random instants in calendar time, at random instants in event-time, immediately before a transaction, etc. These different ways of measuring the spread do not yield identical results (see Figure 4.4). In a nutshell, the spread distribution is narrower when measured before transactions than at random instants. This makes sense, since liquidity takers carefully select the submission times of their market orders to benefit from relatively tight spreads.

4.4 Order Arrivals and Cancellations

Provided that their sizes do not exceed the volume available at the best quote, market orders can only be executed at the best bid or at the best ask. Limit orders,

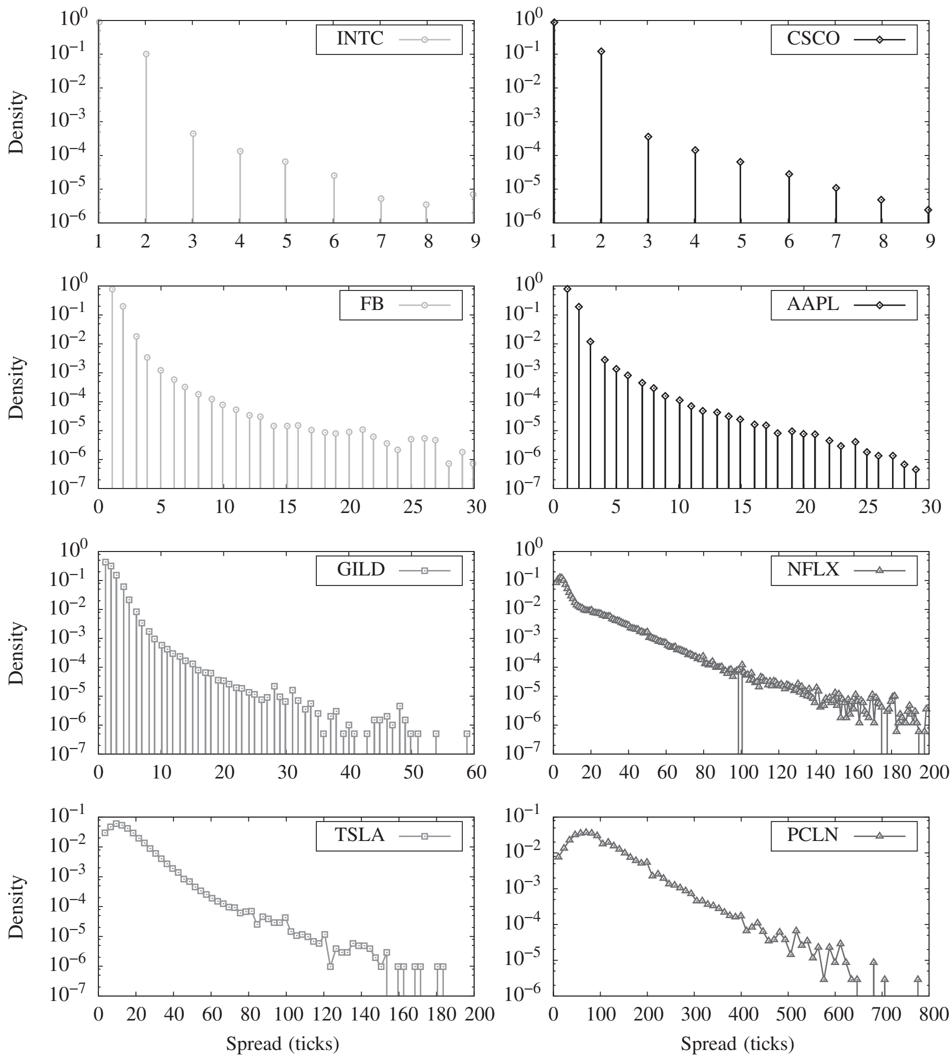

Figure 4.3. Empirical distributions of the bid–ask spread $s(t) = a(t) - b(t)$, measured at the times of events at the best quotes, for the stocks INTC, CSCO, FB, AAPL, GILD, NFLX, TSLA and PCLN. The stocks are ordered (from top left to bottom right) by their relative tick sizes.

on the other hand, can arrive and be cancelled at any of a wide range of prices. It is thus interesting to estimate the distributions of relative prices for arriving and cancelled limit orders, because (together with the arrivals of market orders) they play an important role in determining the temporal evolution of the state $\mathscr{L}(t)$ of the LOB (see for example Chapter 8).

Figure 4.5 shows the distribution of d, the same-side quote-relative prices for arriving limit orders, again for our selection of large-tick and small-tick

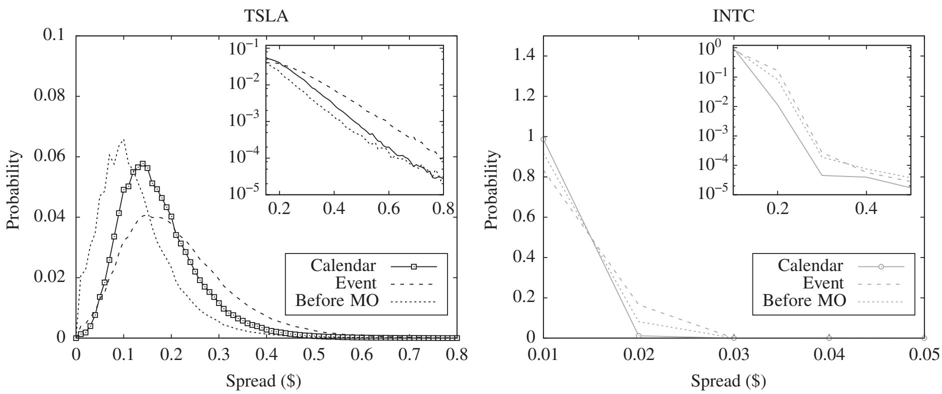

Figure 4.4. Empirical distribution of the bid–ask spread $s(t) = a(t) - b(t)$ for (left panel) TSLA and (right panel) INTC. The solid curves with markers show the results when measuring the spread at a random calendar time, the dashed curves show the results when measuring the spread at the times of events at the best quotes, and the dotted curves show the results when measuring the spread immediately before transactions.

stocks. In both cases, the **deposition probability** peaks at the best quotes, but there is also significant activity deeper in the LOB. Indeed, when plotting the distributions in semi-logarithmic coordinates, it becomes clear that the deposition probabilities decay rather slowly[2] with increasing relative price d. For large-tick stocks, the decay of the distribution is approximately monotonic. For such stocks, the probability of observing a spread $s(t) > \vartheta$ is in fact very small (see Figure 4.3), and when this happens, the overwhelming probability is that the next limit order will arrive within the spread. For small-tick stocks, by contrast, the LOB is usually sparse, with many empty price levels. There is also a secondary peak in the distribution at a distance comparable to the bid–ask spread. This suggests that it is the value of the spread itself, rather than the tick size, that sets the typical scale of the gaps between non-empty price levels in the LOB. Note that there is also substantial intra-spread activity for small-tick stocks (i.e. for $d < 0$), which makes sense because there are typically many empty price levels within the spread.

One can also study the corresponding distributions for limit order cancellations. For all non-negative relative prices, these plots (not shown) are remarkably similar to the distributions of limit order arrivals in Figure 4.5. This suggests that, to a first approximation, the cancellation rate is simply proportional to the arrival rate. One possible explanation for this result is that much LOB activity is associated with many limit orders being placed then rapidly cancelled.

[2] Several older empirical studies have reported this decay to follow a power law.

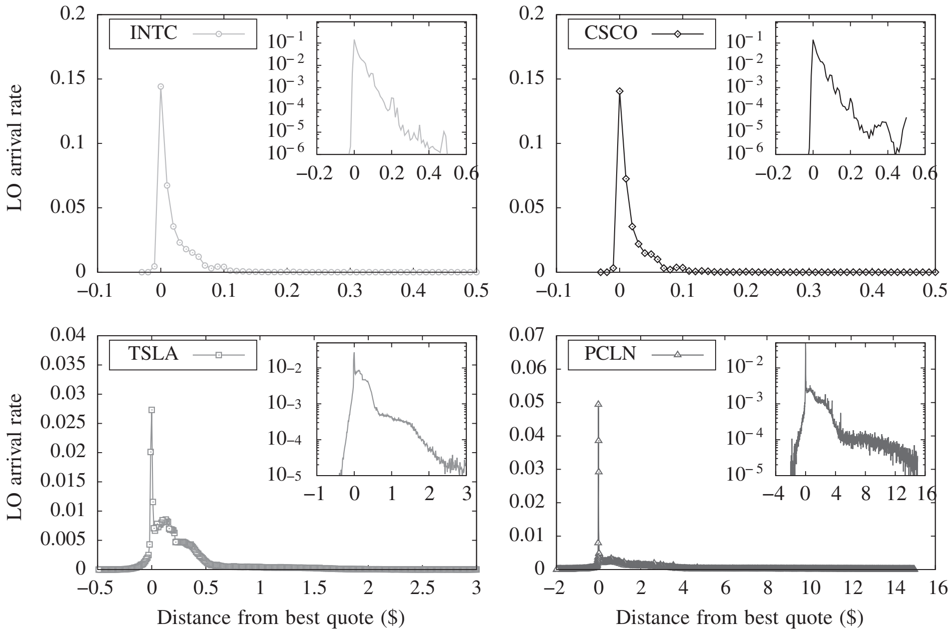

Figure 4.5. Empirical distributions of same-side relative prices d for arriving limit orders for (top panels) the large-tick stocks INTC and CSCO, and (bottom panels) the small-tick stocks TSLA and PCLN. The main plots show the body of the distributions and the inset plots show the upper tails in semi-logarithmic coordinates.

4.5 Order Size Distributions

Figure 4.6 shows the empirical cumulative density functions (ECDFs) for the sizes of limit orders and market orders, expressed in both lots and US dollars. The plot is in log-log coordinates, to emphasise that these distributions are extremely broad. For both limit and market orders, the most common order sizes are relatively small (a few thousands of dollars), but some orders are much larger. The upper tail of the ECDF appears to decay approximately according to a power law, with an exponent scattered around $-5/2$ (as represented by the dotted line). Although this power-law behaviour has been reported by empirical studies of many different markets, the value of the tail exponent varies quite significantly across these studies. Despite their quantitative differences, the same key message applies: order sizes are not clustered around some average size, but rather are distributed over a very broad range of values.

4.6 Volume at the Best Quotes

Figure 4.7 shows the ECDFs for the total volumes V_b and V_a available at the bid- and ask-prices, respectively. For both small-tick and large-tick stocks, the

 https://doi.org/10.1017/9781316659335.007

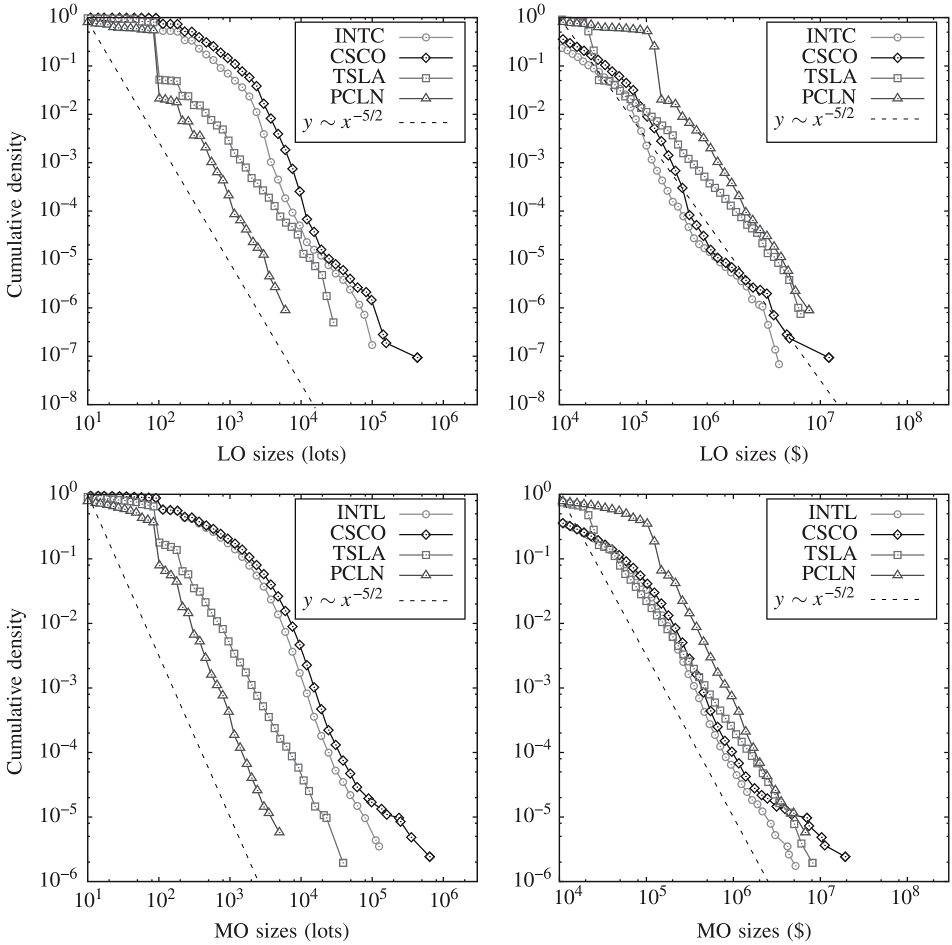

Figure 4.6. Empirical cumulative density functions (ECDFs) for the sizes of (top panels) market orders and (bottom panels) limit orders, expressed in (left column) number of shares and (right column) US dollars, for INTC, CSCO, TSLA and PCLN. The plots are in doubly logarithmic coordinates, such that power-law distributions appear as straight lines. The thin dotted line has slope $-5/2$.

distributions are approximately symmetric for buy and sell orders, and the tails of the distributions are similar to those of the distributions of order sizes in Figure 4.6. This makes sense because a single, large limit order arriving at $b(t)$ (respectively, $a(t)$) contributes a large volume to V_b (respectively, V_a). For small volumes, however, the distributions are quite different from those in Figure 4.6, because while small limit order sizes are common, observing a small value of V_b or V_a is much rarer. This is because the total volume available at $b(t)$ and $a(t)$ is typically comprised of several different limit orders.

As for the spread distribution, the distribution of queue volumes can be measured in calendar-time, in event-time or immediately before a market order.

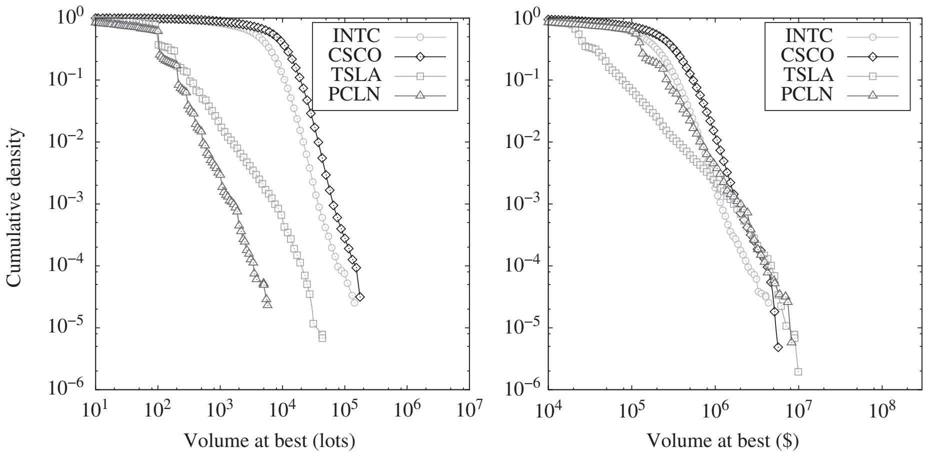

Figure 4.7. ECDFs for the total volumes available at the best quotes, for INTC, CSCO, TSLA and PCLN, expressed in (left panel) number of shares and (right panel) US dollars.

We again observe conditioning effects: for small-tick stocks, the volume at the best quote is higher immediately before being hit by a market order, indicating that liquidity takers choose to submit their orders when the opposite volume is relatively high. However, for large-tick stocks, the volume of the queue is *smaller* before transactions. In this case, as we discuss in Section 7.2, the queue volume itself conveys significant information about the direction of future price changes, because when liquidity is small at the ask (bid) and large at the bid (ask), the price typically moves up (down, respectively). When the queue is small, liquidity takers rush to take the remaining volume before it disappears.

4.7 Volume Profiles

Figure 4.8 shows the mean relative volume profiles, measured as a function of the distance to the opposite-side quote $d^\dagger$ (see Section 3.1.6). For this figure, and throughout this section, we use opposite-side distances because (by definition) there is always a non-zero volume available at the same-side best quotes. This causes a spurious sharp peak in the mean relative volume profile when measured as a function of the same-side distance d, which disappears for opposite-side distances.

The mean relative volume profiles first increase for small distances, then reach a maximum before decreasing very slowly for large distances. This shows that there is significant liquidity deep in the book even at very large distances from the best quotes. We also observe strong round-number effects: liquidity providers seem to prefer round distances (from the opposite best) for their limit orders, such as multiples of half dollars.

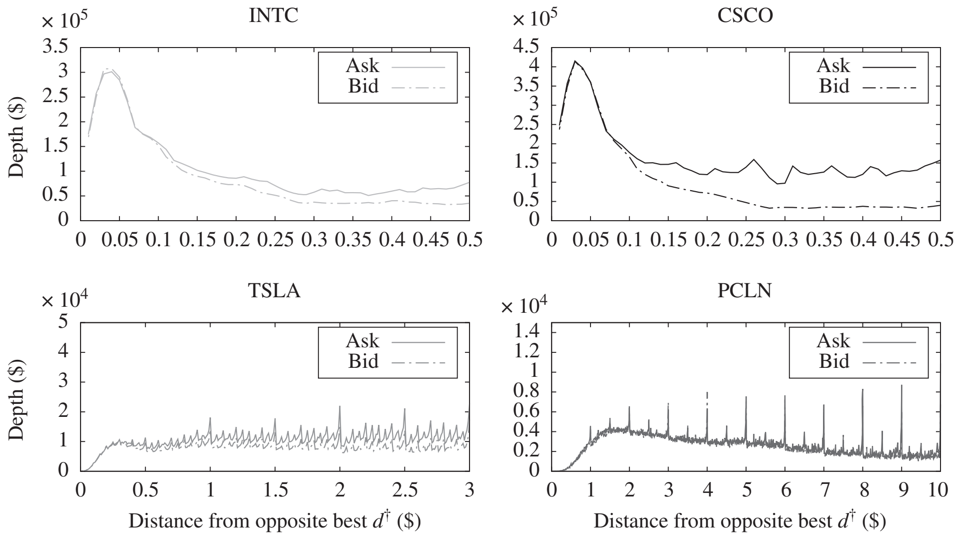

Figure 4.8. Mean relative volume profiles (measured as a function of the distance $d^\dagger$ to the opposite-side quote) of the (dash-dotted curves) bid side and the (solid curves) ask side of the LOBs of (top panels) the large-tick stocks INTC and CSCO, and (bottom panels) the small-tick stocks TSLA and PCLN.

To what extent are these average relative volume profiles representative of a typical volume profile snapshot? To address this question, we also plot snapshots of the volume profiles, taken at 10:30:00 on 3 August 2015, in Figure 4.9. In contrast to the mean relative volume profiles, for which liquidity is (on average) available over a wide range of consecutive prices, LOB snapshots tend to be sparse, in the sense that they contain many prices with no limit orders. This effect is particularly apparent for small-tick stocks, but less so for large-tick stocks (see Figure 4.5).

4.8 Tick-Size Effects

As we discussed in Section 3.1.5, the relative tick size ϑ_r varies considerably across different assets. In this section, we illustrate several ways in which LOBs with different relative tick sizes behave very differently. To illustrate our findings on a wide range of different stocks, throughout this section we consider a sample of 120 different US stocks traded on NASDAQ, for which the (non-relative) tick size is fixed to $\vartheta = \$0.01$ but for which the relative tick size ϑ_r varies considerably, because prices themselves can vary between a few cents (so called "penny stocks") and thousands of dollars (see, e.g., PCLN in Table 4.1).

4.8.1 Tick and Spread

First, we address how ϑ_r impacts the mean bid–ask spread $\langle s \rangle$ (see Figure 4.10). As the figure illustrates, when ϑ_r is small (i.e. when the stock price is large), the

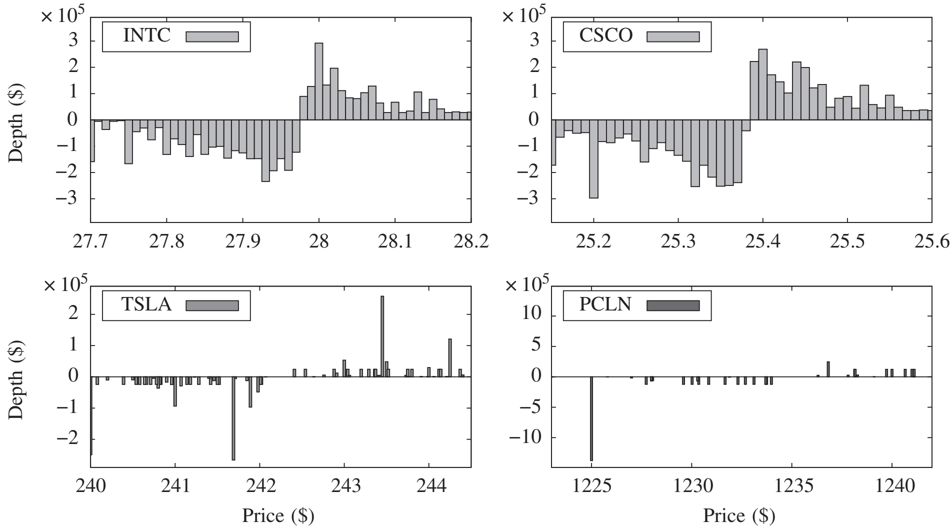

Figure 4.9. Typical snapshots of the volume profiles for INTC, CSCO, TSLA and PCLN, taken at 10:30:00 (EST) on 3 August 2015.

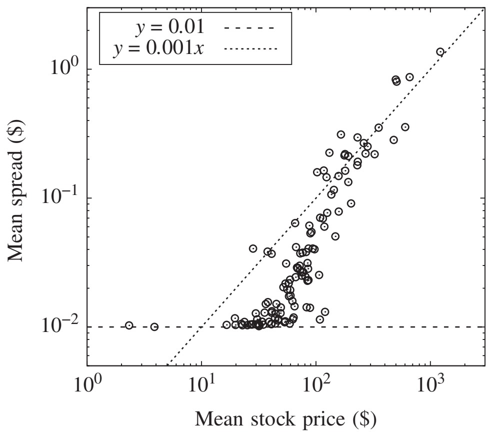

Figure 4.10. Mean bid–ask spread versus the mean share price, for a sample of 120 different US stocks traded on NASDAQ. The tick size on NASDAQ constrains the spread from below (dashed black line).

average spread is roughly proportional to the price of the stock itself. Empirically, the relationship $\langle s\rangle \cong 0.001 \times m(t)$ appears to hold approximately. When the price is smaller than \$10, this proportionality breaks down, because the spread cannot be smaller than $\vartheta = \$0.01$. The mean relative spread is thus much larger for large-tick stocks.

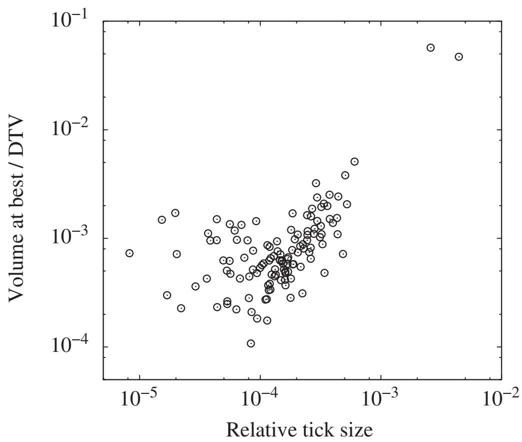

Figure 4.11. Average ratio between the daily mean volume at the best quotes and the daily traded volume versus the relative tick size ϑ_r, for a sample of 120 different US stocks traded on NASDAQ.

4.8.2 Tick and Volume at Best

Second, we address how ϑ_r impacts the mean depths at the best quotes $\langle V_b \rangle$ and $\langle V_a \rangle$, expressed as a fraction of the daily traded volume (see Figure 4.11). For small-tick stocks, $\langle V_b \rangle$ and $\langle V_a \rangle$ are about $\cong 1\%$ of the daily traded volume. For very large-tick stocks, by contrast, $\langle V_b \rangle$ and $\langle V_a \rangle$ can reach $\cong 10\%$ of the daily traded volume. Note that very small relative tick sizes correspond to very large stock prices, since the absolute tick is fixed to $\vartheta = \$0.01$. This can explain why the ratio in Figure 4.11 appears to saturate for the very small-tick stocks, as a single limit order containing 100 shares of PCLN (our smallest-tick stock) already corresponds to a volume of $120,000!

4.8.3 Tick and Volume of Trades

Finally, we consider how ϑ_r impacts the ratio of the average volume of a market order to the average volume at the best quotes (see Figure 4.12). For large-tick stocks, market orders are typically small ($< 10\%$) compared to the mean volume at the best quotes. For small-tick stocks, by contrast, arriving market orders typically consume more than half of the outstanding volume (which is itself quite small in this case).

4.9 Conclusion

In this chapter, we have presented a selection of empirical properties of order flow and LOB state. Some of these statistical properties vary considerably across

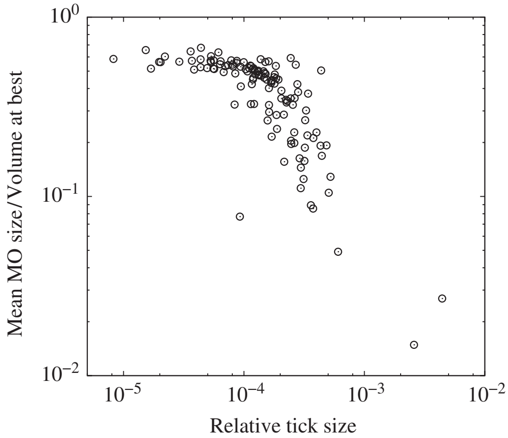

Figure 4.12. Average ratio between the daily mean market order size and the daily mean volume at the best quotes versus the relative tick size ϑ_r, for a sample of 120 different US stocks traded on NASDAQ.

different stocks, whereas others appear to be more universal. The most important feature distinguishing different stocks is the relative tick size ϑ_r, which itself varies considerably across different stocks – even among stocks with the same (absolute) tick size ϑ.

Although the statistics in this chapter provide an interesting glimpse into the behaviour of real LOBs, they are just the first small step towards a more comprehensive understanding of real market activity. Developing such an understanding requires addressing not only specific LOB properties in isolation, but also the interactions between them. In the subsequent chapters, we will consider some of the most relevant questions to help quantify both these interactions and the complex phenomena that emerge from them.

Take-Home Messages

(i) At any instant of time, the total volume in an LOB is a small fraction of the corresponding daily traded volume.

(ii) When measured over a suitably long time horizon, all market statistics are approximately symmetric between buys and sells.

(iii) Many market quantities present a strong average daily profile: volumes and activity exhibit a J-shaped pattern, with most activity happening close to the open and close of the market, whereas the spread undergoes a steep decrease after the open, then decreases more gradually throughout the remainder of the day.

(iv) For large-tick stocks, the spread is almost always equal to one tick. Small-tick stocks show a broader distribution of spread values.
(v) Most activity takes place close to or inside the spread. This is partly due to the fact that traders have little incentive to display publicly their trading intentions long in advance (except to gain queue priority in the case of large-tick stocks).
(vi) The size distributions of limit orders, market orders and best queue volumes have heavy tails.

4.10 Further Reading

General

Madhavan, A. (2000). Market microstructure: A survey. *Journal of Financial Markets*, 3(3), 205–258.
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.
Chakraborti, A., Toke, I. M., Patriarca, M., & Abergel, F. (2011). Econophysics review: II. Agent-based models. *Quantitative Finance*, 11(7), 1013–1041.
Cont, R. (2011). Statistical modeling of high-frequency financial data. *IEEE Signal Processing Magazine*, 28(5), 16–25.
Gould, M. D., Porter, M. A., Williams, S., McDonald, M., Fenn, D. J., & Howison, S. D. (2013). Limit order books. *Quantitative Finance*, 13(11), 1709–1742.
Abergel, F., Chakraborti, A., Anane, M., Jedidi, A., & Toke, I. M. (2016). *Limit order books*. Cambridge University Press.

Empirical Studies of LOBs

Biais, B., Hillion, P., & Spatt, C. (1995). An empirical analysis of the limit order book and the order flow in the Paris Bourse. *The Journal of Finance*, 50(5), 1655–1689.
Bollerslev, T., Domowitz, I., & Wang, J. (1997). Order flow and the bid–ask spread: An empirical probability model of screen-based trading. *Journal of Economic Dynamics and Control*, 21(8), 1471–1491.
Gourieroux, C., Jasiak, J., & Le Fol, G. (1999). Intra-day market activity. *Journal of Financial Markets*, 2(3), 193–226.
Kempf, A., & Korn, O. (1999). Market depth and order size. *Journal of Financial Markets*, 2(1), 29–48.
Challet, D., & Stinchcombe, R. (2001). Analyzing and modeling 1+ 1d markets. *Physica A: Statistical Mechanics and Its Applications*, 300(1), 285–299.
Bouchaud, J. P., Mézard, M., & Potters, M. (2002). Statistical properties of stock order books: Empirical results and models. *Quantitative Finance*, 2(4), 251–256.
Zovko, I., & Farmer, J. D. (2002). The power of patience: A behavioural regularity in limit-order placement. *Quantitative Finance*, 2(5), 387–392.
Farmer, J. D., Gillemot, L., Lillo, F., Mike, S., & Sen, A. (2004). What really causes large price changes? *Quantitative Finance*, 4(4), 383–397.

Hollifield, B., Miller, R. A., & Sandas, P. (2004). Empirical analysis of limit order markets. *The Review of Economic Studies*, 71(4), 1027–1063.

Huang, W., Lehalle, C. A., & Rosenbaum, M. (2015). Simulating and analysing order book data: The queue-reactive model. *Journal of the American Statistical Association*, 110(509), 107–122.